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)

Publication of the journal is supported financially by
the Ministry of Science and Education of the Republic of Croatia

TABLE OF CONTENTS

European Journal of Analytic Philosophy EuJAP | Vol. 18 | No. 1 | 2022
https://doi.org/10.31820/ejap.18.1.1

EXTREME PERMISSIVISM REVISITED

Tamaz Tokhadze[1]

[1] University of Sussex

Original scientific article – Received: 23/09/2021 Accepted: 04/12/2021

ABSTRACT

Extreme Permissivism is the view that a body of evidence could rationally permit both the attitude of belief and disbelief towards a proposition. This paper puts forward a new argument against Extreme Permissivism, which improves on a similar style of argument due to Roger White (2005, 2014). White's argument is built around the principle that the support relation between evidence and a hypothesis is objective: so that if evidence E makes it rational for an agent to believe a hypothesis H, then E makes it rational to believe H, for all agents. In this paper, I construct a new argument against Extreme Permissivism that appeals to a logically weaker, less demanding view about evidential support, Relational Objectivity: whether a body of evidence E is more likely if H is true than if H is false is an objective matter and does not depend on how any agent interprets the relationship between E and H. Relational Objectivity is solely concerned with the conditional probabilities called likelihoods and does not put substantive constraints on an agent's prior and posterior credences. For this reason, the presented argument avoids the standard permissivist criticism levelled against White's argument.

Keywords: *permissivism; uniqueness; Roger White; objectivity of evidential support; relational objectivity; epistemic standards; likelihoods*

Correspondence: t.tokhadze@sussex.ac.uk

1. Introduction

Can a body of evidence equally justify both believing a proposition and its negation? According to the view called *Extreme Permissivism*, the answer is yes. More fully:

> Extreme Permissivism (EP): There are some bodies of evidence E, such that E rationally permits believing that H and believing that $\neg H$.

Roger White (2005, 2014), who has coined the term "Extreme Permissivism", has put forward several arguments against EP. At the heart of White's central arguments is the idea I call the *Objectivity of Evidential Support* (Objectivity, for short): the view that the support relation between evidence and a proposition (or hypothesis) is absolute and does not change from agent to agent. So, according to Objectivity, if the evidence supports a hypothesis, e.g. anthropogenic climate change, then the evidence supports the hypothesis for all agents. And given some plausible assumptions, Objectivity seems to entail that rational individuals cannot respond differently to the same evidence.[1]

Many permissivists have argued that White presupposes "a superseded view of evidential support" (Douven 2009, 347). According to this standard permissivist response, it is a mistake to view the support relation as a two-place relation between evidence and a hypothesis (or a proposition); instead, evidential support can be sensitive to various third, agent-relative factors, such as epistemic standards, personal credence functions, epistemic goals or cognitive abilities.[2] For instance, according to this line of thought, equally informed jurors may come to different but equally justified conclusions about whether a defendant is guilty because they have different epistemic standards on what counts as sufficient and relevant evidence for the defendant's guilt.

In this paper, I offer a novel argument against EP that captures the plausible thought behind Objectivity and is immune to the popular permissivist objection that I've outlined above. My argument substitutes Objectivity with a less demanding thesis I call *Relational Objectivity*: whether a body

[1] A similar argument can be found in Feldman (2007) and Matheson (2011). All the other criticisms of EP that I'm aware of are committed to Objectivity: e.g. Hedden (2015), Dogramaci and Horowitz (2016), Greco and Hedden (2016), Stapleford (2019).

[2] The view that evidential support, at least in some cases, is agent-relative (or requires some agent-relative factor) has been defended by Douven (2009), Titelbaum (2010), Decker (2012), Kelly (2014), Meacham (2014), Peels and Booth (2014), Schoenfield (2014), Kopec and Titelbaum (2016, 2019), Podgorski (2016), Simpson (2017), Jackson (2019), Jackson and Turnbull (forthcoming).

of evidence E is more likely if a hypothesis, H, is true than if H is false depends on E and H themselves and not on how any agent interprets the relationship between E and H. As its name suggests, Relational Objectivity is a relational or contrastive principle. It is solely concerned with the conditional probabilities called *likelihoods*, and does not put substantive constraints on prior and posterior credences of an agent. For this reason, Relational Objectivity is wholly compatible with the view that there are important agent-relative factors that influence what an agent ought to believe and to what degree.

The paper proceeds as follows. After discussing some preliminaries in section 2, I introduce and analyse White's argument against EP in section 3. Discussing White's argument will allow us to see in what respects the novel argument against EP, introduced in section 4, departs from White's original argument. I conclude in section 5 that it is possible to reject EP and still retain an important permissivist idea that subjective, agent-relative factors rationally influence an agent's doxastic states. Hence, I submit that the proposed novel argument against EP improves on a similar style of argument due to White.

2. Varieties of Permissivism

There are many versions of Permissivism. In this paper, we are solely concerned with Permissivism about (categorical, coarse-grained) belief, which states that a body of evidence can equally justify both the attitude of belief and disbelief towards a proposition.

The arguments against EP discussed in this paper are compatible with weaker versions of Permissivism, such as *Moderate Permissivism* and *Credal Permissivism*:[3]

> Moderate Permissivism: There are some bodies of evidence E, such that E rationally permits two belief-attitudes towards a proposition, where suspension of judgment about the proposition is among the permitted attitudes.
>
> Credal Permissivism: There are some bodies of evidence E, such that E rationally permits more than one credence towards a proposition.

[3] See also Jackson (2019) for a discussion of various versions of Permissivism.

Moderate Permissivism allows that in some cases, a body of evidence permits, say, belief that H and suspending judgement about H. One may endorse Moderate Permissivism without endorsing EP. And while, as stated, EP does not logically imply Moderate Permissivism, it is plausible to assume that if EP is true, then Moderate Permissivism is also true.[4]

Regarding Credal Permissivism: it is a thesis about *fine-grained* doxastic attitudes and does not imply any similar thesis about the coarse-grained attitude of belief. While EP is a thesis about (coarse-grained) belief, it implies Credal Permissivism on a widely accepted assumption that one should believe a proposition *only if* one assigns a *high credence* to that proposition.

Credal Permissivism is an extremely popular position within contemporary epistemology;[5] and for that reason alone, it should be considered vastly more plausible than EP.

The negation of Permissivism is called Uniqueness. As with Permissivism, Uniqueness comes in many forms. Because this paper criticises EP, it advocates the view I call *Moderate Uniqueness* (the negation of EP):

> Moderate Uniqueness: For any hypothesis H and evidence E, it is not the case that E justifies/rationally permits belief that H and belief that $\neg H$.[6]

Now that all the key terms have been defined, we are ready to state and analyse White's argument against EP.

3. White's Argument against EP

As noted in the introduction, the key premise in White's argument is Objectivity: the view that the support relation between evidence and a proposition (or hypothesis) is absolute and does not change from agent to agent. White (2014) has specified Objectivity in modal terms, as the thesis

[4] We could have stated EP as the thesis that some bodies of evidence equally justify all *three belief-attitudes* towards a proposition: belief, disbelief and suspension. But as we are solely focused on the permissibility of adopting opposing beliefs towards a proposition, it is unnecessary to strengthen EP.

[5] To quote Douven (2009, 348) "(…) to the best of my knowledge no one calling him- or herself a Bayesian thinks that we could reasonably impose additional constraints that would fix a unique degrees-of-belief function to be adopted by any rational person". Douven is completely right. Even contemporary *Objective Bayesians*, such as Williamson (2010), grant that some evidential situations permit more than one credence towards a hypothesis.

[6] I will use the terms "justified" and "rational" interchangeably. While in other contexts, the two notions could be distinguished, such a distinction would serve no useful purpose in this paper.

that evidential support relations hold *necessarily*: that is, if E supports H then *necessarily* E supports H.

The claim that the evidential support relation holds necessarily may sound unobvious, and even trivially false to some. To take White's (ibid., 313-314) example that illustrates the worry about the necessity claim:

> That the gas gauge reads Full supports the conclusion that the tank is full. But it need not. Suppose we know that the gauge is stuck on Full, or even that the wiring is switched so that it tends to read Full only when the tank is empty. In these cases the gauge's reading Full seems to support no conclusion or the opposite conclusion.

So, according to the above example, the evidence g: "The gas gauge reads Full" may support different conclusions, depending on what else we know about the gas gauge. To this example, White responds that g, in itself, does not support any conclusion about the tank. It is only when we combine g with our background evidence that we can meaningfully talk about what the evidence supports. For instance, if our background evidence is that the gauge is typically reliable, then g unequivocally supports the conclusion that the tank is full. As White (ibid., 314) concludes, when our background evidence is sufficiently specified, "it is hard to make sense of the idea that all of that information might have supported a different conclusion".

Now, using some additional premises, White gives the following argument against EP from Objectivity (2014, 314):[7]

The Evidential Support Argument

> (P1) If E supports H then necessarily E supports H.
>
> (P2) It cannot be that E supports H and E supports $\neg H$.
>
> (P3) Necessarily, it is rational for S to believe that H iff S's total evidence supports H.
>
> Therefore:
>
> (C1) If an agent whose total evidence is E is rational in believing H, then it is impossible for an agent with total evidence E to rationally believe $\neg H$.

[7] The argument is quoted verbatim, but the order of premises and the original formalism is changed for the uniformity of reading.

The first premise is White's version of Objectivity. Regarding the two other premises: P2 is what might be called the *Univocity Principle* (*Univocity*, for short), the view that "evidence speaks univocally, not equivocally" (Weisberg 2020, 2). So, according to Univocity, if evidence points to H it cannot also point to $\neg H$. The last premise, P3, is a bridge principle connecting evidential support with justified/rational belief. And the conclusion of White's argument, C1, is equivalent to Moderate Uniqueness (the negation of EP).

Permissivists have found the argument unconvincing. The most popular criticism of the argument is centred around White's account of evidential support. Several authors have argued that the relation of support is always relative to a third relatum. To quote Kopec and Titelbaum:

> (…) support facts obtain only relative to a third relatum; absent the specification of that third relatum, there simply is no matter of fact about whether the evidence justifies the hypothesis. (Kopec and Titelbaum 2019, 208)

Permissivists have developed a couple of different interpretations of this "third relatum" (see Kopec and Titelbaum 2016, 194); the most common interpretation is in terms of epistemic standards (Schoenfield 2014). Epistemic standards are the norms of evaluating and reasoning about evidence deemed reliable or truth conducive.

A popular and elegant way of representing epistemic standards is in terms of Bayesian *credence functions*. According to the standard Bayesian position, the degree to which an agent ought to believe a hypothesis, H, depends on (at least) two factors: (i) her (total body of) evidence and (ii) her prior probability in H. Prior probabilities (or priors) encode an agent's degree of belief in H before receiving evidence E. An agent's priors may reflect her epistemic standards: say, how much an agent values the simplicity of a hypothesis compared to its explanatory power. So, equally rational agents may adopt non-trivially different priors, depending on how much weight they give to the simplicity considerations over the explanatory considerations; and different priors may lead to non-trivially different posteriors.

Hence, permissivists contend that two individuals can rationally respond to the same body of evidence differently if they endorse different epistemic standards.

In the next section, I'll state a novel argument against EP which avoids the standard criticism of White's argument. I will substitute White's

Objectivity with a relatively undemanding principle about evidential support, which I call *Relational Objectivity*: the view that whether a body of evidence is more likely if a hypothesis is true than if the hypothesis is false is an objective matter. On White's argument, the objective support relation has the belief-guiding role (see the third premise, P3; more on this in the next section). By contrast, the presented argument won't assume that relational facts about support (fully) determine what an agent should believe and to what degree. For this reason, Relational Objectivity won't be susceptible to the standard permissivist objections; or so I will argue.

4. A New Argument against EP

My argument against EP consists of three premises and a theorem of the probability theory. The first premise is a conditional that states that EP implies the existence of a certain type of permissive cases, and the other two premises are epistemic principles which I call the *Moderate Principle*, and *Relational Objectivity*. In what follows, first, I'll state the argument in a premise-conclusion form and then discuss these premises one at a time.

The Relational Objectivity Argument

(1) If EP is true, then two equally informed agents who rationally suspend judgment about H can rationally come to adopt opposing doxastic attitudes about H upon learning some new evidence E: one agent may believe H and the other agent $\neg H$.

(2) The Moderate Principle: If evidence E justifies you in believing that H and prior to learning that E, you were not justified in believing H, then E makes it rational to increase your probability in H; i.e., $P(H \mid E) > P(H)$, where P represents your credence function and P is a rational credence function for you to have.

(3) Relational Objectivity: Whether evidence E is more likely on H than on $\neg H$, depends on the evidence and hypotheses themselves and not on how any agent interprets the relationship between the evidence and hypotheses; i.e., for any two equally informed agents with rational credence functions P and P^*, it cannot be the case that $P(E|H) > P(E|\neg H)$ and $P^*(E|H) \leq P^*(E|\neg H)$.

(4) Theorem: For any H, E and credence function P, E confirms H iff $P(E|H) > P(E|\neg H)$.[8]

Therefore:

(5) Moderate Uniqueness: For any hypothesis H and evidence E, it is not the case that E justifies belief that H and belief that $\neg H$.

The argument is valid. To see this, assume for reductio that EP is true. Given the first premise and the Moderate Principle, EP entails that a body of evidence E could confirm H for one agent and $\neg H$ for some other agent. Now, it is a theorem of the probability theory, that, for any H and probability function P, E confirms H iff $P(E|H) > P(E|\neg H)$. And by Relational Objectivity, if the inequality $P(E|H) > P(E|\neg H)$ is true for some agent, then it is true for all equally informed agents. Therefore, it cannot be the case that E confirms H for one agent and $\neg H$ for some other agent; contrary to our assumption.

Now that we have established the argument's validity, let us proceed to discuss each of its premises, one at a time.

The first premise does not follow from the definition of EP, but it articulates the key idea behind EP; that some bodies of evidence, in themselves, are radically permissive: so the reason why two individuals can adopt opposing doxastic attitudes towards H in light of their shared evidence is not because of their prior convictions about H but because of their different evaluation of the same evidence.[9] Let me provide an example: consider two open-minded agents who, at some time, share the same (background) evidence K about the existence of God of traditional theism, G, and these agents rationally suspend judgment on G, in light of K. Now, according to premise (1), if EP is true, it may be possible that upon learning some new evidence E these agents rationally come to opposing conclusions about God's existence; so that one agent rationally believes G, while the other rationally believes $\neg G$.

I should note that premise (1) does not imply any substantive constraints on an agent's degrees of belief or credences. If two agents are agnostic

[8] I always assume that for all hypotheses x, $0 < P(x) < 1$.

[9] Many permissivists like Kelly (2014), Schoenfield (2014), and Kopec and Titelbaum (2019) explicitly argue that Permissivism is true because rational individuals can evaluate the same evidence in different ways, and not because they already have opposing attitudes towards *H* without any evidence. See, for instance, Kopec and Titelbaum's (2019., Sect. 4) *Reasoning Room* example. So given the published defences of EP, a version of EP that is incompatible with premise (1) does not seem to be an appealing view even for permissivists.

about God's existence, this does not imply that their credences in God are the same. For instance, two agents may agree that a necessary condition for believing G is that it has a high probability of being true, say higher than 0.7. So, on this assumption, both agents may suspend judgment on G, even if one is, say, 0.6 confident in G and the other is 0.4 confident in G. As premise (1) is highly plausible, I expect that it won't be a controversial step in my argument.

The second premise, the Moderate Principle, as its name suggests, is a moderate, uncontroversial thesis. It does not say that you are rational in believing H based on E whenever E rationally increases your probability or rationally *confirms* H. It only states a *necessary* (and not sufficient) condition on when it is rational to *start* believing H based on E; and this necessary condition is that E rationally confirms H. For instance, let H be the hypothesis that Smith did the crime, let E be some new body of evidence; say, the evidence that Smith's fingerprints were found in the crime scene. Now, if prior to receiving evidence E you were not rational in believing H, and if E makes you rational in believing that H, then E must, at least, make it rational to be more confident in H than before.[10]

What makes an agent's credence function rational (or rationally permissible)? Subjective Bayesians hold that the only rationality requirement on an agent's credence function P is that P is a probability function (that is, P satisfies the standard axioms of probability). This requirement is called *coherence*. As we will see, via premise (3), I'll defend an additional constraint on P that goes beyond (probabilistic) coherence;[11] but I won't appeal to any set of conditions that are jointly sufficient for P to be rational (for an agent).

It is important to note that whether E confirms H for an agent depends on what else the agent knows or takes for granted. So it is useful to divide an

[10] One may object to the Moderate Principle for a reason related to the debate about the permissibility to form a belief in light of "*mere statistical evidence*". For instance, suppose that new evidence reduces your probability for a hypothesis but gives you non-statistical evidence for it, where you previously had only statistical evidence for it. Now, if we think that mere statistical evidence cannot suffice for rational belief, then we'll get cases where gaining evidence can justify moving from suspension of judgment to belief, despite reducing the probability of the believed proposition. Such alleged counterexamples against the Moderate Principle are irrelevant to the argument of this paper. Even if mere statistical evidence is insufficient for rational belief (which is a controversial assumption), we can restrict the argument against EP to the cases that do not involve a transfer from statistical to non-statistical evidence. After all, there is no reason whatsoever to think that EP is only true when an agent's evidence changes from statistical to non-statistical evidence.

[11] Of course, it is not surprising that any argument against EP should go beyond a purely subjective Bayesian account of confirmation. I should also note that, while subjective Bayesianism is a popular view, many (e.g. see Maher 1996; Hawthorne 2005) have argued that any purely subjective account of confirmation faces some serious problems, most notably the so-called problem(s) of old evidence. See Maher (1996) for a detailed argument that a purely subjective account of confirmation is unattainable.

agent's total body of evidence into two parts: new evidence E and a body of *background evidence*, denoted by K. What counts as new evidence E and what counts as background evidence K is largely an arbitrary matter and depends on an agent in question and her context of reasoning. For instance, suppose you are particularly interested in how a piece of evidence E bears on the hypothesis, H, that Smith did the crime. E may be the evidence that Smith's fingerprints were found at the crime scene. In evaluating evidence E, your background evidence will include every relevant proposition that you take for granted at that time: such as common-sense propositions about how the world works (e.g. people leave fingerprints and that the fingerprint matching technology is highly accurate) and the assumption that the evidence has not been planted, etc. So given your background evidence K, it is clear that E confirms H.

If we make an agent's background evidence explicit,[12] the Moderate Principle can be stated more fully as follows (for simplicity, I assume that both E and K are sets of propositions):

> The Moderate Principle: Suppose your total body of evidence is $E \cup K$. If evidence E justifies you in believing that H and prior to learning that E, K alone did not justify you in believing H, then $P(H \mid E \wedge K) > P(H \mid K)$, where P represents your credence function and P is a rational credence function for you to have.

The Moderate Principle is as plausible as an abstract epistemic principle can be. It is neutral between permissivist and impermissivist epistemologies. For instance, within the subjective Bayesian framework, the Moderate Principle is obviously right: after all, if relative to your credence function P, a new piece of evidence E does not increase your credence in H, then why start believing H on E? If E does not add to your credence in H, then E cannot be a part of your reason for believing H.

Therefore, I also do not expect the Moderate Principle to be a controversial premise in my argument.

By contrast, the third premise, Relational Objectivity, is a controversial premise from a permissivist perspective. So, it requires a more detailed discussion and motivation, compared to the previous two premises.

[12] Sometimes, for the sake of readability, I won't explicitly mention an agent's background evidence. But it should be remembered that the talk of confirmation only makes sense relative to a given background evidence. I will make background evidence explicit only when necessary.

Relational Objectivity is a comparative principle: it is solely concerned with how likely evidence E is if a hypothesis H is true *than* if H is false. Relational Objectivity is *not* concerned with either a *prior probability* of a hypothesis H, denoted by $P(H)$, or a *posterior probability* of H, $P(H|E)$. Instead, Relational Objectivity is about the conditional probabilities of the following type—*P*(*evidence* | *hypothesis*), called *likelihoods*. A likelihood encodes what a hypothesis, H, says about evidence E: that is, how likely E is on the supposition that H is true.

As with the confirmation relation, whether $P(E|H) > P(E|\neg H)$ depends on a body of background evidence K. So, taking background evidence into account, Relational Objectivity can be stated more precisely as follows:

> Let P and P^* be rational credence functions of two agents who share the same background evidence K; then for any evidence E it cannot be the case that $P(E|H \wedge K) > P(E|\neg H \wedge K)$ and $P^*(E|H \wedge K) \leq P^*(E|\neg H \wedge K)$.

The relevance of Relational Objectivity for our argument is made explicit by the theorem of probability theory:

> Theorem: For any H, E, K, and probability function P, E confirms H relative to K iff $P(E|H \wedge K) > P(E|\neg H \wedge K)$.

So, given this theorem, Relational Objectivity is equivalent to the thesis that if a piece of evidence E (relative to the fixed background evidence) rationally confirms a hypothesis, then it rationally confirms the hypothesis for all (equally informed) agents.

It may be useful to note that, instead of Relational Objectivity, I could have used a similar principle that has been defended by Maher (1996, 163). Maher has argued that the following, more objectivist analysis of confirmation should substitute the subjective Bayesian analysis of confirmation:

> Let $R(K)$ denote the set of all probability functions that are rationally permissible on background evidence K; then E confirms H relative to K, iff for all $P \in R(K)$, $P(H|E \wedge K) > P(H|K)$.

The gist of Relational Objectivity and Maher's principle is the same: on both principles, whether evidence E rationally confirms H does not depend on how an agent subjectively evaluates the relationship between E

and H (relative to K). Clearly, my argument would remain valid if we substitute Relational Objectivity with Maher's principle.

But, unlike Maher's principle, Relational Objectivity makes explicit that the objectivity of confirmation is due to, what Hawthorne (2005, 278) has called "the objectivity or "publicness" of likelihoods that occur in Bayes' theorem" (More on this in section 4.1). So, Relational Objectivity is stated in a way that emphasises this publicness or objectivity of likelihoods.

Why accept Relational Objectivity? Firstly, Relational Objectivity is logically weaker than White's Objectivity: the former is entailed by the latter but not the other way around. So, any reason for accepting Objectivity is also a reason for accepting Relational Objectivity. Let me elaborate on this.

Objectivity is concerned with the traditional notion of evidential support which is closely related to the notion of rational belief. This is made explicit by the third premise of White's argument:

> (P3) Necessarily, it is rational for S to believe that H iff S's total evidence supports H.

By contrast, as Theorem makes explicit, Relational Objectivity is concerned with the notion of *confirmation*. And the confirmatory relation between E and H is necessary but often insufficient for an agent to rationally believe H on E, even if S's total body of evidence is E (I give an example shortly). An alternative way of explaining the difference between White's Objectivity and Relational Objectivity is by invoking Carnap's (1962, Preface to the Second Edition) well-known distinction between "concepts of firmness" and "concepts of increase in firmness". White's Objectivity concerns the firmness of a hypothesis; it says that whether a hypothesis is sufficiently firm or probable (for belief) is an objective matter. By contrast, Relational Objective concerns whether the evidence increases the firmness or confirms the hypothesis. And, as it is well-known, the evidence may increase the firmness of a hypothesis without making the hypothesis firm (or sufficiently firm). For instance, consider a detective who received reliable testimony that a suspect, John, was seen near the crime scene. Suppose that this piece of evidence, T, is the detective's total body of evidence that Jonn committed the crime (denoted by J). Now, even if T rationally confirms or increases the probability of J, it is clearly irrational to believe that J solely on the basis of T. In Carnap's terms, T increases the firmness of J but does not make J firm enough (for the detective). Hence, while confirmation is necessary for rational belief, it is often insufficient.

To sum up: Relational Objectivity is motivated by the same core idea as Objectivity, that the evidential support relation is objective, at least to some extent. But, unlike White's Objectivity, Relational Objectivity only commits us to a moderate view about the extent to which the support relation is objective.

Certainly, permissivists may call Relational Objectivity into question. But, as we will see, to call Relational Objectivity into question requires more than the appeal to the standard permissivists claims: that subjective, agent-relative factors such as epistemic standards, goals, and personal credence functions have a rational influence on an agent's doxastic states. So, I'll be happy to concede to permissivists that there is no objective support relation in White's sense: where the objective support relation fully determines what an agent ought to believe. However, as I argue next, Relational Objectivity won't commit us to such a demanding view about objective support.

4.1 Relational Objectivity

As I've already explained, Relational Objectivity is solely concerned with the type of conditional probabilities called likelihoods. Unlike prior and posterior probabilities, likelihoods are widely considered to be the most objective part of Bayesian inference. To illustrate this, let us consider one of the most common forms of Bayes' theorem:

$$(1)\quad P(H|E) = \frac{P(E|H) * P(H)}{P(E|H) * P(H) + P(E|\neg H) * P(\neg H)}$$

Equation (1) enables us to calculate posterior probability, $P(H|E)$, in terms of the prior probability of H, $P(H)$, and two likelihoods: $P(E|H)$ and $P(E|\neg H)$.[13] Notice that (by the law of total probability) the denominator in Bayes' theorem—$P(E|H) * P(H) + P(E|\neg H) * P(\neg H)$—equals to the *expectedness of evidence*, $P(E)$. Hence, $P(E)$ is equivalent to the *probability-weighted average of likelihoods*. So, (1) can be simplified to:

$$(2)\quad P(H|E) = \frac{P(E|H) * P(H)}{P(E)}$$

And by using equation (2), if we take the *ratio* of $P(H|E)$ and $P(\neg H|E)$ we get the *ratio form of Bayes' theorem*:

[13] Sometimes, likelihoods written as $P(E|\neg H)$ denote the likelihoods of a special kind known as *catchall* likelihoods. Catchall likelihoods are discussed at the end of this section.

$$(3)\quad \frac{P(H|E)}{P(\neg H|E)} = \frac{P(E|H) * P(H)/P(E)}{P(E|\neg H) * P(\neg H)/P(E)}$$

By simplifying, we get:

$$(4)\quad \frac{P(H|E)}{P(\neg H|E)} = \frac{P(E|H)}{P(E|\neg H)} * \frac{P(H)}{P(\neg H)}$$

Finally, let R_{Post} be the ratio of *posteriors*, R_L the ratio of *likelihoods*, and R_{Prior} the ratio of *priors*, then, the ratio form of Bayes' theorem can be summarised succinctly as:

$$(5)\quad R_{Post} = R_L * R_{Prior}$$

As equation (5) makes explicit, the *impact* of evidence on any pair of priors is completely exhausted by R_L, the ratio of likelihoods.[14]

Relational Objectivity is solely concerned with the value of R_L and not at all concerned with R_{Prior} and R_{Post}. This is an important selling point of Relational Objectivity as prior probabilities are unanimously acknowledged as the most subjective and problematic part of Bayesian inference. And what makes Relational Objectivity more appealing is that it is not a *quantitative* but a *comparative* principle: Relational Objectivity is not concerned with the precise numerical values of likelihoods, but only with their comparative probabilities. As $R_L = P(E|H)/P(E|\neg H)$, it follows that $P(E|H) > P(E|\neg H)$ iff $R_L > 1$. So, on Relational Objectivity, the exact value of R_L is unimportant; what is important is whether R_L is greater than 1.

Now, even the so-called subjective Bayesians—that is, Bayesians who allow the multitude of *coherent* prior distributions as rationally permissible—accept that R_L is the most objective part of Bayesian inference (Hawthorne 2005, 283). The objective status of R_L is due to the fact that, in many cases, an agent's evidence defines an objective (or inter-subjectively justified) probability distribution over a set of competing

[14] The claim that the ratio of likelihoods is the only factor that impacts how the evidence changes the ratio of priors (which, as equation (5) illustrates, is a fact of probability theory) should not be conflated with a different claim that the ratio of likelihoods provides the adequate *measure* of the degree to which the evidence confirms a hypothesis. While some have argued that likelihoods are sufficient to adequately measure confirmation, not everyone accepts this. See Festa and Cevolani (2017) for a relevant discussion and references.

hypotheses *without presupposing any prior probability distribution over these hypotheses*.[15]

To illustrate the independence of likelihoods from prior probabilities, consider the following diagnostic example (a more philosophical example is considered shortly):

> You are a physician who assesses a patient on whether she has some skin disease D. Based on the extensive medical records, you know that the symptoms S, a peculiar rash on her hands, is 90% likely if she has D and only 10% likely if she does not have D. So, you know that $P(S|D) = 0.9 > P(S|\neg D) = 0.1$.

Based on this information, you already know that evidence S confirms D: $P(D \mid S) > P(D)$. And as the likelihood ratio is quite high, $P(S|D)/P(S|\neg D) = 0.9/0.1 = 9$, we know that the evidence S makes the posterior ratio, $P(D|S)/P(\neg D|S)$, nine times greater than the prior ratio, $P(D)/P(\neg D)$. So it is clear that S provides quite good evidence for D. But, this being said, the new evidence, S, is *insufficient* to conclude that the posterior probability of D is high (say, higher than 0.5). This is so, because the prior of D may be quite low. So, suppose that D is a rare disease and only 1 in 1000 have it. And if your prior in D is 1/1000, then, unintuitively, simple calculations show that your posterior probability in D should be less than 1%: $P(D|S) \approx 0.009$. And in general, even if you do not have sufficient information to provide an objective, uncontentious estimate of the prior of D, you can still rationally conclude that S is more likely on D than on $\neg D$: hence you can rationally conclude that S confirms D.

As this diagnostic example illustrates, likelihoods may be independent of priors and in many contexts have "objective or inter-subjectively agreed values" (Hawthorne 2005, 283). For this reason, even subjective Bayesians accept the special status of likelihoods; for instance, Edwards et al. (1963, 199) called likelihoods *public* because "[i]n many applications practically all concerned find themselves in substantial agreement with regard to [likelihoods]".

Certainly, fixing the precise numerical values of likelihoods is not always as easy and objective as in the above diagnostic example. However, Relational Objectivity is *not* concerned with precise numerical values of likelihoods but merely with their comparative plausibilities. So, on Relational Objectivity, whether $P(E|H)$ is greater than $P(E|\neg H)$ is an

[15] See also Bandyopadhyay et al. (2016) for a detailed discussion about the special status and role of likelihoods in Bayesian inference.

objective matter, even if $P(E|H)$ and $P(E|\neg H)$ do not always have objective numerical values. This makes Relational Objectivity a modest and appealing thesis even from a subjectivist perspective; as there are many cases where the exact *numerical values* of likelihoods are highly debatable, but we may still be in a position to know comparative claims about these likelihoods. To illustrate this, consider the following, more philosophically interesting example. Suppose two agents agree that the existence of evil constitutes evidence against God's existence, in the sense that the existence of evil is less likely if God exists than if God does not exist: $P(evil\ exists|God) < P(evil\ exists|no\ God)$. And the agreement about the comparative plausibilities of these likelihoods requires neither the agreement about the priors, nor the agreement about the precise numerical values of these likelihoods.

While in many scientific and philosophical settings comparative claims about likelihoods are objective (or intersubjectively justified), it is unrealistic to suppose that this is always the case. Essentially, the problem is that sometimes it is not possible to approximate in a non-subjective manner the values of the so-called *catchall* likelihoods: the likelihoods that contain catchall (or composite) hypotheses. A catchall hypothesis is a disjunction of simple (or non-composite) hypotheses. To take an easy example: the hypothesis H_1: "the coin is fair" is simple while the hypothesis $\neg H_1$: "the coin is not fair" is a catchall, as $\neg H_1$ is the disjunction of all the specific alternatives to H_1 (there are many specific ways in which the coin fails to be fair, if not assumed otherwise). Now, suppose the coin is tossed ten times and eight heads are obtained (denote this observation as "e"). The likelihood of e on the supposition that the coin is fair, H_1, is completely objective and does not require the specification of a prior distribution over the competing hypotheses (no matter what the prior distribution is, $P(e|H_1) = 45 * 0.5^{10} \approx 0.04$). By contrast, the corresponding catchall likelihood, $P(e|\neg H_1)$ is sensitive to the prior distribution. To calculate the value of $P(e|\neg H_1)$ we must know the values of all *ordinary likelihoods* of e on each specific (mutually exclusive) alternative to H_1 and the *prior distribution* over these alternatives. In symbols:

$$P(e|\neg H_1) = \frac{\Sigma_{i\neq 1} P(e|H_i) * P(H_i)}{P(\neg H_1)}$$

So, mathematically, catchall likelihoods are reducible to priors and ordinary likelihoods: if we know the values of priors and likelihoods, then we can calculate the value of any catchall likelihood. Thus, in many important settings, the values of catchall likelihoods cannot be neatly

separated from the values of priors. The reader may worry that this feature of catchall likelihoods calls the argument of this paper into question, as the argument relies on the independence of likelihoods and priors.

But the sensitivity of catchall likelihoods on priors is wholly consistent with my argument. For a start, there is an important *asymmetry* between catchall likelihoods and priors (Fitelson 2007, Section 5). If one knows the values of priors and ordinary likelihoods, then one can calculate the value of the corresponding catchall likelihoods; but not the other way around. Knowing the value of catchall likelihoods and ordinary likelihoods does not determine the prior distribution.[16] So, less information is required to determine the values of catchalls than to determine the values of priors. Because of this, the subjectivity of priors does not necessarily translate to the subjectivity of catchall likelihoods: we could have an objective approximation of the value of a catchall likelihood, but not the value of the corresponding priors.

I expect the following objection at this point: "But what if two equally rational agents have different estimates of priors and, due to this, they disagree about comparative claims involving catchall likelihoods? Does not this show that Relational Objectivity is false?" The answer is "No". Relational Objectivity does not entail a strong and unobvious claim that the inequalities between likelihoods are always objectively well-defined. It is compatible with Relational Objectivity that in some cases, the available evidence does not (objectively) justify even comparative claims about likelihoods (or confirmation). And if such cases obtain, i.e., if the available evidence does not justify comparative claims about likelihoods, then, according to my argument against EP, it is irrational to believe (or disbelieve) the relevant hypothesis, H, on that evidence.[17] Instead, we should conclude that there is no fact of the matter whether E is more likely on H than on $\neg H$. This conclusion is entirely compatible with Relational Objectivity, as it does not require that the inequalities between likelihoods are objectively well-defined on any evidence.

Hopefully, this discussion convinces even those sympathetic with permissivist epistemologies that Relational Objectivity is a plausible,

[16] To take the coin example again, we could be justified to think that the likelihood of getting eight heads out of ten tosses is higher on the supposition that coin is *not* fair than on the supposition that the coin is fair, even if it is not possible to estimate the value of priors in a non-subjective manner.

[17] For instance, Sober (2008, 26-30) has argued that when we deal with "deep and general" theories, such as the general theory of relativity, then some comparative claims about likelihoods cannot be objectively justified. See Sober (Chapter 1, Sections 1.2, 1.3) for a detailed discussion. Sober's overall view agrees with our conclusion that if evidence and hypotheses do not justify the comparative claims about likelihoods, then an agent should abstain from forming beliefs on such evidence.

moderate principle, especially compared to such principles as White's Objectivity that imposes very strong constraints on rational belief.

In the next and final section, I summarise the key points of the paper and conclude that the presented argument against EP improves upon White's similar style of argument.

5. Conclusion

I've presented a novel argument against Extreme Permissivism (EP). According to this argument, EP is false because two equally informed agents who suspend judgement about a proposition H cannot adopt opposing attitudes towards H upon receiving the same new evidence; as adopting opposing attitudes towards H requires that E confirms H for one agent, but $\neg H$ for the other agent; but given a plausible and relatively weak principle about likelihoods—which I've called Relational Objectivity—this cannot be the case.

Since Relational Objectivity is only concerned with relational probabilities of likelihoods, the presented argument against EP is wholly compatible with a plausible permissivist idea that non-evidential factors—such as epistemic standards, goals, and credence functions—have (some) rational influence on what an agent ought to believe (and to what degree).

To illustrate this, suppose that Credal Permissivism is true (that is, equally informed agents can adopt different credences toward a proposition). Now, consider two equally informed detectives, Salome and Naomi, who adopt non-identical credence functions P_S and P_N and suspend judgment on whether John committed the crime (denoted by J). Salome may be more sceptical about John's guilt and attach lower prior in J; for simplicity, assume that relative to their shared background evidence K, $P_S(J) = 0.25$ and $P_N(J) = 0.4$. Now, further suppose that they learn a new piece of evidence E that rationally confirms J: $P(E|J) > P(E|\neg J)$. Since Salome had a lower prior in J, her posterior in J, $P_S(J|E)$, may not be high enough for her to believe that J. By contrast, from Naomi's point of view, $P_N(J|E)$ could be sufficiently high to believe that J.

This example is wholly consistent with the presented argument against EP. So, even if Relational Objectivity is true, two agents may have non-trivially differing credences towards a proposition and rationally disagree about whether to believe or suspend judgment on that proposition.

Certainly, supporters of EP may find Relational Objectivity too demanding and advance some novel objections against it. But, as I've argued, these objections must go beyond the standard permissivist claims that subjective, non-evidential factors have a rational influence on what an agent ought to believe and to what degree.

Acknowledgments

I am very grateful to Tony Booth, Corine Besson, Elizabeth Jackson, and JiMin Kwon for helpful comments and discussion on different parts and versions of this paper. Special thanks to Nina Abesadze for helpful discussions and feedback that shaped the argument of this paper. I'm also very grateful to two anonymous referees for this journal for challenging me to clarify and improve some important aspects of the paper.

REFERENCES

Bandyopadhyay, Prasanta S., Gordon Jr. Brittan, and Mark L. Taper. 2016. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. Springer International Publishing.

Carnap, Rudolf. 1962. *Logical Foundations of Probability. 2nd.* Chicago: University of Chicago Press.

Decker, Jason. 2012. "Disagreement, Evidence, and Agnosticism." *Synthese* 187: 753–783. https://doi.org/10.1007/s11229-011-0010-x.

Dogramaci, Sinan, and Sophie Horowitz. 2016. "An Argument for Uniqueness about Evidential Support." *Philosophical Issues* 26 (1): 130-147. https://doi.org/10.1111/phis.12078.

Douven, Igor. 2009. "Uniqueness Revisited." *American Philosophical Quarterly* 46: 347–61.

Edwards, Ward, Harold Lindman, and Leonard J. Savage. 1963, 70. "Bayesian Statistical Inference for Psychological Research." *Psychological Review*: 193-242.

Feldman, Richard. 2007. "Reasonable Religious Disagreement." In *Philosophers without Gods: Meditations on Atheism and the Secular Life*, edited by Louise Antony, 194–214.

Festa, Roberto, and Gustavo Cevolani. 2017. "Unfolding the Grammar of Bayesian Confirmation: Likelihood and Antilikelihood Principles." *Philosophy of Science* 84 (1): 56–81. https://doi.org/10.1086/688935.

Fitelson, Branden. 2007. "Likelihoodism, Bayesianism, and Relational Confirmation." *Synthese* 156: 473–489.

https://doi.org/10.1007/s11229-006-9134-9.

Greco, Daniel, and Brian Hedden. 2016. "Uniqueness and Metaepistemology." *Journal of Philosophy* 113 (8): 365–395.

Hawthorne, James. 2005. "Degree-of-Belief and Degree-of-Support: Why Bayesians need Both Notions." *Mind* 114 (454): 277-320. https://doi.org/10.1093/mind/fzi277.

Hedden, Brian. 2015. "A Defense of Objectivism about Evidential Support." *Canadian Journal of Philosophy* 45 (5-6): 716-743. https://doi.org/10.1080/00455091.2015.1123038.

Jackson, Elizabeth. 2019. "A Defense of Intrapersonal Belief Permissivism." *Episteme* 1-15. https://doi.org/10.1017/epi.2019.19.

Jackson, Elizabeth, and Margaret Greta Turnbull. Forthcoming. "Permissivism, Underdetermination, and Evidence." In *The Routledge Handbook of the Philosophy of Evidence*, edited by Clayton Littlejohn and Maria Lasonen-Aarnio.

Kelly, Thomas. 2014. "Evidence can be Permissive." In *Contemporary Debates in Epistemology*, edited by Matthias Steup, John Turri and Ernest Sosa, 298-312. Wiley Blackwell.

Kopec, Matthew, and Michael G. Titelbaum. 2016. "The Uniqueness Thesis." *Philosophy Compass* 11: 189-200. https://doi.org/10.1111/phc3.12318.

Kopec, Matthew, and Michael G. Titelbaum. 2019. "When Rational Reasoners Reason Differently." In *Reasoning: Essays on Theoretical and Practical Thinking*, edited by Magdalena Balcerak Jackson and Jackson Balcerak-Jackson, 205-231. Oxford: Oxford University Press.

Maher, Patrick. 1996. "Subjective and Objective Confirmation." *Philosophy of Science* 63 (2): 149-174.

Matheson, Jonathan. 2011. "The Case for Rational Uniqueness." *Logos & Episteme* 2 (3): 359–373. https://doi.org/10.5840/logos-episteme20112319.

Meacham, Christopher. 2014. "Impermissive Bayesianism." *Erkenntnis* 79: 1185–217.

Peels, Rik, and Anthony Booth. 2014. "Why Responsible Belief is Permissible Belief." *Analytic Philosophy* 55: 75–88. https://doi.org/10.1111/phib.12036.

Podgorski, Abelard. 2016. "Dynamic Permissivism." *Philosophical Studies* 173 (7): 1923-1939. https://doi.org/10.1007/s11098-015-0585-z.

Schoenfield, Miriam. 2014. "Permission to Believe: Why Permissivism is True and What it Tells us About Irrelevant Influences on Belief." *Noûs* 48: 193–218. https://doi.org/10.1111/nous.12006.

Simpson, Robert. 2017. “Permissivism and the Arbitrariness Objection.” *Episteme* 14 (4): 519-538. https://doi.org/10.1017/epi.2016.35.

Sober, Elliott. 2008. *Evidence and Evolution: The Logic behind the Science*. Cambridge University Press.

Stapleford, Scott. 2019. “Intraspecies Impermissivism.” *Episteme* 16 (3): 340-356. https://doi.org/10.1017/epi.2018.14.

Titelbaum, Michael G. 2010. “Not Enough There There: Evidence, Reasons, and Language Independence.” *Philosophical Perspectives* 24 (1): 477-528.

Weisberg, Jonathan. 2020. “Could’ve Thought Otherwise.” *Philosophers’ Imprint* 20 (12): 1-24. http://hdl.handle.net/2027/spo.3521354.0020.012.

White, Roger. 2005. “Epistemic Permissiveness.” *Philosophical Perspectives* 19: 445-459. https://doi.org/10.1111/j.1520-8583.2005.00069.x.

White, Roger. 2014. “Evidence cannot be Permissive.” In *Epistemology, Contemporary Debates*, edited by Matthias Steup, John Turri and Ernest Sosa, 312-323. Wiley-Blackwell.

Williamson, Jon. 2010. *In Defence of Objective Bayesianism.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199228003.001.0001.

European Journal of Analytic Philosophy EuJAP | Vol. 18 | No. 1 | 2022
https://doi.org/10.31820/ejap.18.1.2

GNOSTIC DISAGREEMENT NORMS

Domingos Faria[1]

[1] LanCog, Centre of Philosophy, University of Lisbon

Original scientific article – Received: 29/08/2021 Accepted: 26/01/2022

ABSTRACT

Our main question in this paper is as follow: (Q) What are the epistemic norms governing our responses in the face of disagreement? In order to answer it, we begin with some clarification. First, following McHugh (2012), if we employ a useful distinction in normativity theory between evaluative and prescriptive norms, there are two readings of (Q)—we explore such distinction in section 2. And secondly, we accept gnosticism, that is, the account that the fundamental epistemic good is knowledge. It is with this assumption that we want to answer (Q). So, if gnosticism is true, what is the plausible answer to (Q)? In section 3 we argue for gnostic disagreement norms as response to (Q) and in section 4 we apply such norms to particular cases of disagreement.

Keywords: *disagreement; evaluative norms; prescriptive norms; disagreement norms; gnostic disagreement norms; knowledge-first epistemology.*

Correspondence: domingosfaria@edu.ulisboa.pt

1. Introduction

Our main question in this paper is as follows:

(Q) What are the epistemic norms governing our responses in the face of disagreement?

With this question we want *disagreement norms* not only for cases of peer and idealized disagreement, but also for ordinary or daily cases of disagreement. Thus, we intend to propose norms that target all types of disagreement.[1] In order to answer question (Q), we begin with some clarification. First, following Conor McHugh (2012), if we employ a useful distinction in *normativity theory* between evaluative and prescriptive norms, there are two readings of (Q)—we explore such distinction in section 2. And secondly, we accept *gnosticism* (see Williamson 2000; Littlejohn and Dutant 2021), that is, the account that the fundamental epistemic good is knowledge. It is with this assumption that we want to answer (Q). Roughly, we are assuming gnosticism because (i) it gives a good account of the value of knowledge[2] and (ii) its consequences for the case of disagreement are hitherto unexplored.[3] So, if gnosticism is true, what is the plausible answer to (Q)? In section 3 we argue for gnostic disagreement norms as response to (Q) and in section 4 we apply such norms to particular cases of disagreement. As the main contribution of this paper, we highlight the application of the distinction between evaluative and prescriptive norms to key cases of disagreement, based on a gnostic epistemology.

2. Evaluative and Prescriptive Norms

In this section, we want to clarify the question (Q). This question will be approached by employing a distinction in normativity theory. Namely, following McHugh (2012), Simion, Kelp, and Ghijsen (2016), and Simion

[1] This methodology is also followed, for example, by Hawthorne and Srinivasan (2013), and Broncano-Berrocal and Simion (2021).

[2] For example, Williamson (2000, 79) holds that knowledge is more valuable than mere true belief because "present knowledge is less vulnerable than mere present true belief to rational undermining by future evidence".

[3] There are a few exceptions, such as Hawthorne and Srinivasan (2013), Miragoli and Simion (2020), and Broncano-Berrocal and Simion (2021). What are the advantages of our gnostic account compared to the previous ones? In this paper we focus more on source epistemic properties (i.e., properties of the method of belief formation) to deal with disagreement, whereas the previous gnostic accounts focus more on state epistemic properties (i.e., properties of the belief itself). But these approaches are compatible.

(2019), if we employ a useful distinction in *normativity theory* between evaluative and prescriptive norms, we can distinguish two readings of (Q).[4] On the one hand, evaluative norms are primarily about what is *good* or valuable;[5] they are concerned with an *ought-to-be*. Thus, in this sense such norms are "evaluations of (e.g.) properties, objects, events or states of affairs" (McHugh 2012, 10). For example, we can have an evaluative norm for schools (ENS) that says "a good school is a safe school (i.e., schools ought to be safe)". Or, more simple cases, such as "a good driving is safe driving", "a good knife is sharp", and so on. Since we are just attributing value to something and not prescribing a certain course of conduct for agents, "evaluations do not presuppose accountability or blameworthiness" (McHugh 2012, 10).

On the other hand, prescriptive norms are mainly about what one *ought to do*.[6] According to McHugh (2012, 9), "they require, permit or forbid certain pieces of conduct on the part of agents, and are apt to guide that conduct". For instance, we can have a prescriptive norm for schools (PNS) that says "the school board ought to hire school safety monitors". In this case, the norm PNS prescribes a certain conduct for the school board. And the school board is responsible if it fails to comply with PNS. For, we are accountable to prescriptive norms, "in the sense that violating them is liable to leave us open to blame" (McHugh 2012, 9). The same is true of other prescriptive norms, such as "Drive 50 km/h within city bounds" or "Don't steal".

Despite being distinct norms, there is a relation between evaluative norms and prescriptive norms. For, evaluative norms often have implications for prescriptive norms; namely, we can often *derive* a prescriptive norm from an evaluative norm. Precisely because *prescriptive norms* serve to make it likely that the *evaluative norms* will be met. For instance, a *prescriptive norm*, such as PNS, serves to make compliance with an *evaluative norm*, such as ENS, more likely. Thus, "evaluative norms often come first and prescriptive norms are in their service" (Simion, Kelp, and Ghijsen 2016, 386).

However, "these implications may not be straightforward" (McHugh 2012, 10); for, it is possible to violate an *evaluative norm* without violating a *prescriptive norm*. For instance, back to our example about school safety,

[4] Some people, like Wedgwood (2007), think that notions like "ought" have many different readings. Here we want to explore only the readings proposed by McHugh (2012), Simion, Kelp, and Ghijsen (2016), and Simion (2019).
[5] Following Geach (1956), goodness or value is used in this context in an attributive sense.
[6] We can hold that *prescriptive norms* in epistemology are equivalent to "decision procedure rules" in ethics.

the school board may have complied with PNS, while ENS may be violated, if the people working as school safety monitors are negligent, or incompetent, etc. And it is possible to violate a *prescriptive norm* without violating an *evaluative norm*. For example, it is possible that the school board fails with PNS because it does not hire security monitors, while ENS may be complied, if teachers and the school board decide to work overtime to monitor school safety.[7]

Based on this normative framework, our initial question (Q), about what are the epistemic norms for disagreements, can be understood as an evaluative question (EQ) and as a prescriptive question (PQ):

> (EQ) What is the good epistemic doxastic attitude in cases of disagreement?
> (PQ) What ought one to do, epistemically speaking, in cases of disagreement?

An *evaluative norm for disagreement* answers such questions by determining what counts as a good doxastic attitude in the face of disagreement. And a *prescriptive norm for disagreement* determines what one ought to do in the face of disagreement.[8]

3. Arguing for Gnostic Disagreement Norms

As emphasized in the introduction, we accept *gnosticism*, that is, the account that the fundamental epistemic good is knowledge. 'Gnosticism' is a term introduced by Clayton Littlejohn to refer to

[7] This framework, involving the distinction between evaluative norms and prescriptive norms, is not accepted by some philosophers, such as Hughes (2021b). Thanks to an anonymous referee for alerting us to these criticisms of the framework we are using. While our aim here is not to react to each of these criticisms, we can briefly summarize the objections and our responses. Hughes' first objection is that we can straightforwardly derive prescriptive norms with the same content as evaluative norms from evaluative norms. We grant that this can happen; but in the specific context of disagreement, as we develop and justify it in the next sections, it makes sense to understand the contents of the evaluative and prescriptive norms of disagreement as different. Hughes' second objection is that there are circumstances in which a prescriptive norm, such as the one we propose, does not do the job that prescriptive norms are supposed to do. But we argue in the next sections that this problem does not seem to affect our prescriptive norm of disagreement, given that such a norm is reasonably conducive to action and informative for beings like us. And Hughes' last objection is that, in cases in which being rational is hard, further norms (such as 'undergo anti-bias training') can emerge. But we can minimize this problem by developing our prescriptive norm of disagreement in terms of virtues, as we do in the next sections. A detailed analysis of all the objections presented by Hughes (2021b) will be addressed on another occasion.

[8] Typically, these questions are not answered in the literature on disagreement.

> the view that there is one and only one fundamental epistemic good and that that good is knowledge. On this view, the beliefs that realize the fundamental good are the ones that constitute knowledge. (Littlejohn 2017, 227)

It can be argued that gnosticism is the best view of epistemic value; see, for example, Williamson (2000), Hyman (2015), Littlejohn and Dutant (2021) on the value of knowledge as an argument for gnosticism. Timothy Williamson argues that knowledge is more valuable than mere true belief; for, knowledge is more stable, namely

> one's belief in a proposition p is more robust to evidence if one knows p than if one merely believes p truly; one is less likely to lose belief in p in the course of interacting with the environment by discovering new evidence which lowers the probability of p. (Williamson 2000, 8)

Moreover, we also want to draw attention to the hitherto neglected interaction between gnosticism and disagreement, aside from Simion's recent series of publications (see, for example, Miragoli and Simion 2020; and Broncano-Berrocal and Simion 2021). It is with this gnostic assumption that we want to answer (Q). Thus, if *gnosticism* is true, how to answer (EQ) and (PQ)? Our proposal is as follows:

> *Gnostic evaluative norm of disagreement* (END): In cases of disagreement about whether *p*, one holding steadfast *p* is a good doxastic attitude if and only if one *knows* that *p*.

> *Gnostic prescriptive norm of disagreement* (PND): In cases of disagreement about whether *p*, one must: hold steadfast *p* if and only if one has *good cognitive dispositions* in believing that *p* (that is, in forming or retaining the belief *p*, one exhibits dispositions that tend to manifest epistemic quality states—*knowledge*—in normal counterfactual cases).

In this account, the evaluative norm (END) comes first and determines what counts as a good doxastic attitude in the face of disagreement, namely it states that a good steadfast position qualifies as *knowledge*. From (END) we derive the prescriptive norm (PND), which determines what one ought to do in the face of disagreement, namely it prescribes that only *virtuous beliefs*, resulting from *good cognitive dispositions* (i.e. good dispositions

about belief formation and retention), should be held steadfast.[9] However, in the case of a bad cognitive disposition, such a gnostic approach prescribes changing the doxastic attitude in a way conducive to knowledge through *conciliation* with the interlocutor, if the latter has better cognitive dispositions with respect to the disputed belief about whether *p*.[10] If none of the subjects of the disagreement about whether *p* has knowledge or better cognitive dispositions than their interlocutor, and such cognitive dispositions of these subjects are equally bad, then it seems that the most reasonable thing to do is to suspend judgment or belief, such as Hawthorne and Srinivasan (2013) suggests for similar norms. In the last section we will apply our gnostic norms of disagreement to practical cases to analyse when it is appropriate to follow a steadfast view or, instead, a conciliatory view.

We want to underline that the proposed norms are applied to cases of disagreement with *outright belief*. However, disagreements can also involve *credences*. For instance, I may have a .8 credence in *p*, whereas my interlocutor has a .4 credence in *p*. Can the proposed framework explain how we ought to update our credences in cases of disagreement? If so, how? First of all, our gnostic norms of disagreement were initially intended only for cases of outright belief. Second, we can somehow adapt our norms for credence cases. Following a gnostic or knowledge-first epistemology, namely the strategy proposed by Williamson (2021c), credences can be understood operationally as species of belief. Specifically, we can hold that a credence that *p* reduces to a *belief* in a proposition about the probability of *p*. For instance, a credence of .9 that the ticket will lose reduces to a belief with the content that *the probability the ticket will lose is .9*. And such a belief can constitute knowledge. So, a person can know and believe that the probability that her ticket will lose is .9, even though she does not know and believe that her ticket will lose. Thirdly, based on this understanding of credences, we can easily apply our gnostic evaluative norm of disagreement to credence cases in the following way: In cases of disagreement about whether *p* (where *p* is a belief with the content about the probability of a proposition *q*), one holding steadfast *p* is a good doxastic attitude if and only if one knows that *p*. Something similar can be said in relation to the gnostic prescriptive norm (PND).[11]

[9] This is inspired by Williamson (2021a, 2021b), Lasonen-Aarnio (2021a, 2021b). We want to maintain that Williamson's strategy for the epistemic norm of belief can be replicated for the epistemology of disagreement. According to Williamson, the epistemic norm governing belief is knowledge: believe that *p* only if you know that *p*. And from such a norm, Williamson (2021a) derives a norm for evaluating agents according to which one ought to manifest knowledge-conducive cognitive dispositions. For other accounts of gnostic disagreement norms, see Hawthorne and Srinivasan (2013) and Miragoli and Simion (2020).

[10] We are grateful to an anonymous referee for alerting us to this point.

[11] Many thanks to Bruno Jacinto and an anonymous referee for pressing us on this.

Such (PND) gives a general *guide for action* that serves to make compliance with (END) likely. That is so because, in general, *good cognitive dispositions* typically generate knowledge and bad cognitive disposition typically do not generate knowledge. More specifically, cognitive dispositions are classified as "good" or "bad" according to the quality of the epistemic states they tend to manifest across normal counterfactual cases (in which *knowledge* is the best quality status). Here we are understanding *normal counterfactual cases* as non-deviant cases relative to the subject of evaluation, but where the type of situation and the disposition manifested are very similar (to the actual case).

In order to further clarify the previous point, we can give some examples. For instance, dispositions to form or retain beliefs based on "wishful thinking" or whims are bad cognitive dispositions, since beliefs formed through such dispositions do not tend to lead to quality epistemic states across a relevant range of normal counterfactual cases. The same goes, in general, for dispositions to ignore testimony from reliable sources or from experts. However, a disposition to form or retain perceptual beliefs as a result of ordinary perception is a good cognitive disposition, because beliefs formed through this disposition tend to lead to quality epistemic states across a relevant range of normal counterfactual cases.

Our gnostic prescriptive norm of disagreement (PND) could be formulated in other terms. For example, in a recent paper, Littlejohn and Dutant (2021) propose that it is rational for one to believe that *p* if and only if it is probable that one knows that *p*. This account could be adapted to serve the role of a prescriptive norm for disagreement according to which, in cases of disagreement about whether *p*, one must: hold steadfast *p* if and only if it is probable that one knows that *p*.[12] In normal cases (i.e., non-deviant cases, such as non-skeptical cases), there seems to be a correspondence between Littlejohn and Dutant's (2021) adapted norm and our prescriptive norm. For, if *S* has good cognitive dispositions in believing that *p* (namely, dispositions that tend to lead to knowledge in normal counterfactual cases), then it is probable that *S* knows that *p*. However, we prefer our formulation of the gnostic prescriptive norm in terms of *dispositions*, because even in bad scenarios (such as the Evil Demon world cases or Gettier-cases), in which it is not probable that *S* knows that *p*, *S* may still have good cognitive dispositions, given that dispositions are evaluated with respect to normal counterfactual cases. For example, a person who is deceived by an evil Cartesian demon does not know propositions about her surroundings based on her perception. Even so, the dispositions manifested by such a person

[12] Many thanks to an anonymous referee for pressing us on this.

in forming perceptual beliefs may be good, if in normal counterfactual cases such dispositions are conducive to knowledge.

But what are the arguments for our Gnostic Disagreement Norms? Starting first by presenting the main argument for (PND), we want to highlight that:

1. A good prescriptive norm satisfies the following desiderata: (i) it makes compliance with (END) probable, (ii) it is an adequate guide to action or at least is informative, and (iii) it preserves, in general, the tie between what agents ought to do and the praiseworthiness/blameworthiness of the agent's actions.
2. (PND) satisfies the proposed desiderata better than the rival accounts.
3. Therefore, (PND) is a good prescriptive norm.

We think premise 1 follows from the definition of prescriptive norms. It is worth remembering, as we presented in the initial part of this paper, that the function of the prescriptive norm is to reinforce behavior conducive to compliance with the evaluative norm (this is the first desideratum). Concomitantly, in order for us to perform such a function, the prescriptive norm needs to be informative, giving us some guide to action (this is the second desideratum). As such, we are accountable to prescriptive norms, in the sense that in typical circumstances we can be blamed of violating them or we can be praised for complying with them (this is the third desideratum). These desiderata are part of the description of the initial framework as proposed by McHugh (2012), Simion, Kelp, and Ghijsen (2016), and Simion (2019).

Concerning the second premise, the main rival account seems to be a prescriptive *knowledge norm*. But the prescriptive norm of disagreement cannot be understood as a knowledge norm *simpliciter*. For, such knowledge norm is not an operationalizable guide for action; namely, because knowledge is a non-luminous condition (see Hawthorne and Srinivasan 2013). Thus, given our limited perspective on the world, we often don't have access to facts about where the knowledge is, and so a prescriptive knowledge norm doesn't provide us with an adequate guidance about what to do, thereby violating the second desideratum. However, (PND) seems to be a more operationalizable guide for action.[13] At least, even if there is no fully operational epistemology (see Williamson 2014a), generally it seems feasible for us and accessible from the subject's

[13] Arguments for an epistemology with guidance see, for example, Gibbons (2013) and Fassio and Gao (2021). In contrast, for an epistemology without guidance, see Hughes (2021a).

standpoint to attempt to develop virtuous dispositional traits, such as "be open-minded", "be conscientious", and so on.

Even accepting the general lack of luminosity, and hence the idea that norms are not fully operational, we can take the view that norms can be more or less *informative* as a guide for action.[14] Namely, we can argue that a dispositional norm, such as (PND), is more informative as a guide for action than a knowledge norm *simpliciter*. For, a norm that prescribes the development of epistemic virtues (with concrete examples of "being open-minded", "following the evidence", "considering objections", etc.) seems more informative, as a guide for action towards a gnostic goal, than a norm that merely says to believe propositions we know. However, even holding that our gnostic prescriptive norm of disagreement is informative, we also concede that developing virtuous dispositions is not always *accessible from the subject's standpoint*. For example, consider the case of the Benighted Cognizer, adapted from Goldman by Simion, Kelp, and Ghijsen (2016, 382):

> Ben is a member of an isolated and benighted community. Many of his methods of belief formation have no connection to truth whatsoever, but they are common lore in Ben's community. Let's suppose that Ben wants to know the best time to sow his crops. According to the lore of his community, in order to achieve this, he will first have to sacrifice a goat and bury it in a sacred place. Then he must sit outside his house until it starts to rain and then return to the burial place. If the sun is shining again by the time he will have arrived, it is time to sow the crops. If not, he will have to return home and continue sitting outside his house until the next rainfall. Ben has flawlessly implemented this procedure and has thereby arrived at a belief that it is time to sow the crops.

In this case, it can be argued that our gnostic prescriptive norm has no force for Ben; for, he is not in a position to develop virtuous dispositions.[15] By way of response we can claim that, given his circumstances, Ben is blameless for not developing better virtuous dispositions and, thus, he is blameless for violating the gnostic prescriptive norm. Thus, we can specify that the proposed disagreement norms only apply in circumstances or communities conducive to the development of virtuous dispositions.

[14] Broncano-Berrocal and Simion (2021) also argue that a norm can be informative even if it is not fully operational.

[15] Many thanks to an anonymous referee for pressing us on this.

Outside of these favorable circumstances, the subject may be excused for violating the disagreement norms.

Moreover, and more importantly to support our prescriptive norm (PND), there are cases (such as the Evil Demon world cases or Gettier-cases) that do not constitute knowledge, but there is no problem in being steadfast (for example, because we are experts and our disagreeing interlocutor is a layman, or because we manifest good cognitive dispositions and our disagreeing interlocutor manifests bad cognitive dispositions) and the belief in question deserves positive evaluation and not just the attribution of "blameless", since its formation and retention manifests good cognitive dispositions.

For example, suppose we have two envatted subjects S_1 and S_2, being fed by deceptive experiences, who disagree about whether *p*. But imagine that while S_1 manifests good cognitive dispositions in forming and retaining the belief that *p*, S_2 manifests poor cognitive dispositions in forming and retaining the belief that ¬*p*. Given that we are faced with a sceptical scenario, neither of these subjects has knowledge; thus, if we were to follow a prescriptive knowledge norm *simpliciter*, we would have to conclude that both subjects should abandon their respective beliefs. However, this seems implausible, since S_1, unlike S_2, manifests good cognitive dispositions (i.e. dispositions that tend to lead to quality epistemic states in normal counterfactual cases) and therefore S_1 deserves positive evaluation. This point cannot be accommodated with a knowledge norm *simpliciter*, but it can be with (PND). Thus, with (PND), even victims of massive deceit deserve a kind of praise by taking a steadfast position if they have good cognitive dispositions.[16]

In contrast, a knowledge norm has more to do with an evaluative norm. For that reason, our main argument for (END) is as follow:

1. In cases of disagreement about whether *p*, one holding steadfast *p* is a good doxastic attitude if and only if one's belief that *p* is a *good* belief.
2. One's belief that *p* is a good belief if and only if one knows that *p* (see Williamson 2000).
3. Therefore, (END): In cases of disagreement about whether *p*, one holding steadfast *p* is a good doxastic attitude if and only if one knows that *p*.

[16] This point is inspired by Simion, Kelp, and Ghijsen (2016), Lasonen-Aarnio (2021a, 2021b).

As for the first premise, it is epistemically good to hold steadfast that *p* in the face of disagreement if and only if continuing to believe *p* is epistemically better than disbelieve or suspend judgment about whether *p*. And if so, it is because such a belief is epistemically good. Thus, at least from an *objective* point of view, if it is good to hold firm that *p*, then such a belief is epistemically good. And, on the second premise, we accept Williamson's argument that

> if believing *p* is treating *p* as if one knew *p*, then knowing is in that sense central to believing. (…) Knowing is in that sense the best kind of believing. Mere believing is a kind of botched knowing. In short, belief aims at knowledge. (Williamson 2000, 47)

Thus, we can conclude that (END). And we have (PND) in order to make compliance with (END) likely.

4. Applying the Gnostic Disagreement Norms

In this last section we want to apply our Gnostic Disagreement Norms to particular cases of disagreement in order to verify the behaviour of our norms. To fulfill this task, we begin with the following question: Does the higher-order evidence from disagreement give *S* a defeater for his belief?[17] In other words, does learning that a person disagrees with you about *p* give you a defeater for *p*?[18]

Given the centrality of gnosticism in this paper, we follow a gnostic epistemology of defeaters according to which *defeaters* are evidence that the subject is not in a position to know the target proposition (see Gibbons 2013; Baker-Hytch and Benton 2015; Littlejohn and Dutant 2021). So, defeaters are indicators of ignorance; evidence that if the subject were to believe, her belief would fail to constitute knowledge. For instance, consider the following case (see Lasonen-Aarnio 2010):

> At time *t*1 Suzy sees a wall that appears red to her. On this basis, Suzy believes that the wall is red—let's call this proposition *p*. There is nothing abnormal about her perceptual abilities. At a slightly later time *t*2 she learns a proposition *q*

[17] This formulation of the question is common in the disagreement literature, e.g., Frances and Matheson (2018), Matheson (2021). Higher-order evidence is evidence about our evidence. There is also good recent literature on defeaters, such as Brown and Simion (2021).

[18] As Matheson (2021) stresses, such questions are relevant in discussing the epistemology of disagreement.

> from a highly reliable and trustworthy authority, that there is a red light shining on the wall.[19]

In this case at *t*2 Suzy acquires evidence and a proposition *q* that she is not in position to know that *p*. Thus, it is not rational for Suzy to continue to believe *p*. However, is knowledge always defeated by higher-order evidence (for example, by evidence learned in face of disagreement)? We want to underline that knowledge is not always defeated. For, firstly, not all instances of higher-order evidence are indicators of ignorance (see Hawthorne and Srinivasan 2013, 13). For example, where only one of two parties in disagreement is an expert (or has more evidence than the other), it seems that one can continue to know in the face of disagreement.[20]

Secondly, even if higher-order evidence constitutes an indicator of ignorance, it is possible to have *improbable knowing*, which is an instance of KK failure[21] (see Williamson 2014b). In other words, it can be likely on *S*'s evidence that *S* doesn't know *p* (and, thus, *S* has an indicator of ignorance), but *S* knows *p*. To see this, consider a simple case (see Williamson 2014b, 972):

> [T]he unconfident examinee answers questions on English history under the impression that he is merely guessing. In fact, his answers are correct, and result from lessons on it that he has completely forgotten he ever had. The example can be so filled in that it is extremely improbable on the examinee's evidence that he had any such lessons (…); nevertheless, he does know the historical facts in question.[22]

Williamson (2014b, 979) also present a more complex case. He argues that, given the limits of our discriminatory capacities and margin for error, a subject *S* can know that *p* <the pointer on the dial is within a certain range Q>, although it's improbable on *S*'s evidence that *S* knows that *p*.

And, lastly, a subject can acquire a defeater for *p* and, thus, an indicator of ignorance, and such a subject may still have knowledge that *p*, or more properly, *unreasonable knowledge* as Maria Lasonen-Aarnio (2010) calls it. Following Lasonen-Aarnio (2010)'s reasoning, at *t*1 a subject *S* can have a belief *p* that is well-based and and safely true. And, although at a later

[19] This case is an example of *undercutting defeat*.
[20] It is worth remembering that, as we stressed at the beginning of this paper, our methodology aims to address all cases of disagreement.
[21] I.e., a case of knowing something even though one does not know at the time that one knows that thing.
[22] Such example is not intended to show that knowledge and belief come apart.

time *t*2 such *S* can acquire a defeater (for instance, misleading higher-order evidence against *p*), if *S* simply retains her belief *p* on the same basis as before, then it does not follow that her belief *p* is no longer well-based and and safely true. Thus, *S* can continue to know *p* at *t*2 (despite not having a reasonable belief, given that it manifests bad cognitive dispositions).

Thus, applying this reasoning to cases of disagreement, we can hold that it is possible to have knowledge in face of disagreement. For, (i) not all instance of disagreement gives us defeaters (for example, in cases where we are the only expert); and (ii) in cases in which we really acquire a defeater in a disagreement case, it is possible to have *improbable knowing* or *unreasonable knowledge*. So, it is possible to comply with (END) in the face of disagreement.

But, even if knowledge is not defeated in cases of disagreement, there are many cases in which ignoring defeaters *is not* a *good cognitive disposition*. For instance, in general, in cases of disagreement, ignoring high-order evidence from reliable sources does not manifest a good cognitive disposition, since beliefs formed through such a disposition do not tend to lead to quality epistemic states across a relevant range of normal counterfactual cases. In short, by ignoring a defeater one can continue to have knowledge, but in general, by doing so, one is not manifesting good cognitive dispositions.[23] Thus, in such cases we are not complying with (PND). In order to better articulate this point, let's look at some particular cases of disagreement:

> **Competent Mathematician Case** (adapted from Williamson 2021b): Mary is a competent mathematician and she has just proved a surprising new theorem. She shows her proof to several distinguished senior colleagues, who all tell her that it involves a subtle fallacy. She cannot quite follow their explanations of her mistake [and she ignores such objections]. In fact, the only mistake is in their objections, obscured by sophisticated bluster; her proof is perfectly valid.

In this case we can say that, even if Mary is complying with (END), she is not complying with (PND). For, the disposition to ignore the testimony of experts/seniors, in cases like theses (in which it is not feasible to discriminate between misleading counter-evidence and correct counter-

[23] As Lasonen-Aarnio (2021b) maintains, "even knowing is compatible with manifesting dispositions that lead one astray across a range of relevant counterfactual cases".

evidence), does not tend to lead to quality epistemic states across a relevant range of normal counterfactual cases.[24]

> **Restaurant Case** (adapted from Christensen 2007): We go to dinner, and we agreed to split the bill evenly, adding 20% tip. We each mentally calculate our share, and I become highly confident that our shares are 43€, whereas you become highly confident that we each owe 45€. How should I react to my colleague's belief?

According to (PND), it is not correct to follow a steadfast position in this case, since a disposition to ignore an epistemic peer in this case, in which one could easily make a mistake, does not tend to lead to quality epistemic states across a relevant range of normal counterfactual cases. Namely, if we remain inflexible and do nothing else, like a careful recalculation, there is no way available to us to choose between who is right that would track what is best across a range of counterfactual cases.[25]

> **Elementary Math Case** (adapted from Lackey 2010): Harry and I are equally competent colleagues. We were drinking coffee and trying to determine how many people from our department will be attending the upcoming Workshop. I say: "Mark and Mary are going on Thursday, and Sam and Stacey are going on Friday, and since 2+2=4, there will be four other members of our department at that Workshop". In response, Harry asserts, "But 2+2 does not equal 4". How should I react to my colleague's belief?

According to (PND), it is appropriate to follow a steadfast position in this case, because a disposition to ignore a peer in this case, about a very elementary math equation (a proposition on which I am strongly justified and that it is rooted in the ways I reason), tend to lead to quality epistemic states in normal counterfactual cases.

In short, what does our Gnostic Prescriptive Norm prescribe? According to (PND), in *Competent Mathematician Case* and in *Restaurant Case*, it is not correct to follow a steadfast position, since agents in such cases are not

[24] For a similar case and diagnosis, see Lasonen-Aarnio (2010, 2021b), and Williamson (2021b). However, following Sosa's (2021) epistemology, we may have a different diagnosis; since, in order to have "first-hand knowledge", an agent may ignore criticism from well-qualified sources (see Carter 2021). But it can be argued that the Competent Mathematical Case is not an instance of first-hand knowledge, but an instance of more social or testimonial knowledge; for, the agent wants to share information with her scientific community and, in this case, the defeaters are relevant (see Greco 2020).
[25] For a similar case and diagnosis, see Lasonen-Aarnio (2019, 175).

manifesting good cognitive dispositions. While in *Elementary Math Case*, it is appropriate to follow a steadfast position, since the agent in question is exercising a good cognitive disposition. But from an *evaluative* point of view, in all these cases it is possible that agents are complying with (END). However, it is worth remembering that (END) is primarily about *what is good* and not about what *one ought to do*. Ideally there would be a correspondence between (END) and (PND).[26] When there is no such correspondence, we have (PND) as a guide for action in face of disagreement, since (PND) has priority when it comes to prescription and agent action.[27]

Acknowledgments

We are grateful to Bruno Jacinto, Ricardo Santos, Mona Simion, and Christoph Kelp for helpful comments and discussion of an earlier version of this paper. Earlier versions of this paper have been presented at the *Agreement and Disagreement Conference* (University of Kent), the *14th Conference of the Italian Society for Analytic Philosophy* (Noto), the *ArgLab Open Seminar* (IFILNOVA), and the *LanCog Social Epistemology Workshop* (CFUL). We are grateful to the audience for helpful discussion. Special thanks to two anonymous referees for their generous comments that helped us improve this paper. Any errors or omissions are our responsibility. Work for this paper was supported by the post-doctoral project CEECIND/01066/2017 of the Portuguese Foundation for Science and Technology.

REFERENCES

Baker-Hytch, Max, and Matthew A. Benton. 2015. "Defeatism Defeated." *Philosophical Perspectives* 29 (1): 40–66.

[26] Williamson (2021a) accepts a similar point, arguing that a properly functioning system is one that meets a "local norm", such as (END), and a "global norm", such as (PND).

[27] We can consider that in these cases there is no normative conflict, because there is some kind of order of priority between the two norms (END) and (PND). Namely, when the point of evaluation concerns what should be done in practice and as a guide to action, (PND) has priority. For a similar point, see Simion (2021). However, as objection it can be point out that an alternative approach to the epistemology of disagreement is to have a two-pronged account similar to the one we propose, but to think of cases in which the knowledge norm and the dispositional rationality norm conflict as epistemic dilemmas (see, e.g., Hughes 2019 who advocates this approach to disagreement). By way of response, we can say that the dilemmic view isn't action-guiding and, furthermore, we can stress that our initial framework, on which we base our gnostic norms of disagreement, is not 'dilemmic' given the distinction we draw between evaluative and prescriptive norms. Hughes (2021b) himself acknowledges that if this framework is plausible, there is no 'dilemmic epistemology'. Many thanks to an anonymous referee for pressing us on this.

https://doi.org/10.1111/phpe.12056.

Broncano-Berrocal, Fernando, and Mona Simion. 2021. "Disagreement and Epistemic Improvement." *Synthese* 199: 14641–14665. https://doi.org/10.1007/s11229-021-03437-5.

Brown, Jessica, and Mona Simion, eds. 2021. *Reasons, Justification, and Defeat*. Oxford University Press.

Carter, Adam. 2021. "Collateral Conflicts and Epistemic Norms." In *Epistemic Dilemmas: New Arguments, New Angles*, edited by Kevin McCain, Scott Stapleford, and Matthias Steup. London and New York: Routledge.

Christensen, David. 2007. "Epistemology of Disagreement: The Good News." *Philosophical Review* 116 (2): 187–217. https://doi.org/10.1215/00318108-2006-035.

Fassio, Davide, and Jie Gao. 2021. "Do We Really Need a Knowledge-Based Decision Theory?" *Synthese* 199: 7031–7059. https://doi.org/10.1007/s11229-021-03104-9.

Frances, Bryan, and Jonathan Matheson. 2018. "Disagreement". The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta. https://plato.stanford.edu/entries/disagreement/

Geach, Peter. 1956. "Good and Evil." *Analysis* 17 (2): 33–42. https://doi.org/10.1093/analys/17.2.33.

Gibbons, John. 2013. *The Norm of Belief*. Oxford: Oxford University Press.

Greco, John. 2020. *The Transmission of Knowledge*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108560818.

Hawthorne, John, and Amia Srinivasan. 2013. "Disagreement Without Transparency." In *The Epistemology of Disagreement*, 9–30. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199698370.003.0002.

Hughes, Nick. 2019. "Dilemmic Epistemology." *Synthese* 196 (10): 4059–90. https://doi.org/10.1007/s11229-017-1639-x.

———. 2021a. "Epistemology Without Guidance." *Philosophical Studies*, 179: 163–196. https://doi.org/10.1007/s11098-021-01655-8.

———. 2021b. "Epistemic Dilemmas Defended." In *Epistemic Dilemmas*, edited by Nick Hughes. Oxford: Oxford University Press.

Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Lackey, Jennifer. 2010. "What Should We Do When We Disagree?" In *Oxford Studies in Epistemology*, edited by John Hawthorne and Tamar Szabó Gendler, 274–93. Oxford: Oxford University Press.

Lasonen-Aarnio, Maria. 2010. "Unreasonable Knowledge." *Philosophical Perspectives* 24 (1): 1–21. https://doi.org/10.1111/j.1520-8583.2010.00183.x.

———. 2019. “Guidance, Epistemic Filters, and Non-Accidental Ought-Doing.” *Philosophical Issues* 29 (1): 172–83. https://doi.org/10.1111/phis.12166.

———. 2021a. “Perspectives and Good Dispositions.” *Philosophy and Phenomenological Research*.

———. 2021b. “Dispositional Evaluations and Defeat.” In *Reasons, Justification, and Defeat*, edited by Jessica Brown and Mona Simion, 93–115. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198847205.003.0005.

Littlejohn, Clayton, and Julien Dutant. 2021. “Defeaters as Indicators of Ignorance.” In *Reasons, Justification, and Defeat*, edited by Mona Simion and Jessica Brown, 223-246. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198847205.003.0010.

Littlejohn, Clayton. 2017. “La Verdad En El Gnosticismo.” *Análisis Revista de Investigación Filosófica* 3 (2): 217. https://doi.org/10.26754/ojs_arif/a.rif.201621568.

Matheson, Jonathan. 2021. “Disagreement and Higher-Order Evidence.” In *The Routledge Handbook of the Philosophy of Evidence*, edited by Maria Lasonen-Aarnio and Clayton Littlejohn. London and New York: Routledge.

McHugh, Conor. 2012. “The Truth Norm of Belief.” *Pacific Philosophical Quarterly* 93 (1): 8–30. https://doi.org/10.1111/j.1468-0114.2011.01413.x.

Miragoli, Martin, and Mona Simion. 2020. “Gender, Race, and Group Disagreement.” In *The Epistemology of Group Disagreement*, edited by Fernando Broncano-Berrocal and J. Adam Carter, 125–38. London and New York: Routledge. https://doi.org/10.4324/9780429022500-7.

Simion, Mona, Christoph Kelp, and Harmen Ghijsen. 2016. “Norms of Belief.” *Philosophical Issues* 26 (1): 374–92. https://doi.org/10.1111/phis.12077.

Simion, Mona. 2019. “Knowledge-First Functionalism.” *Philosophical Issues* 29 (1): 254–67. https://doi.org/10.1111/phis.12152.

———. 2021. “Scepticism about Epistemic Dilemmas.” In *Epistemic Dilemmas: New Arguments, New Angles*, edited by Kevin McCain, Scott Stapleford, and Matthias Steup. London and New York: Routledge.

Sosa, Ernest. 2021. *Epistemic Explanations: A Theory of Telic Normativity, and What It Explains*. Oxford: Oxford University Press.

Wedgwood, Ralph. 2007. *The Nature of Normativity*. Oxford and New York: Oxford University Press.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press. https://doi.org/10.1093/019925656x.001.0001.

———. 2014a. "Knowledge First." In *Contemporary Debates in Epistemology*, edited by Matthias Steup, John Turri, and Ernest Sosa, 1-9. Second Edition. Wiley-Blackwell.

———. 2014b. "Very Improbable Knowing." *Erkenntnis* 79 (5): 971–99. https://doi.org/10.1007/s10670-013-9590-9.

———. 2021a. "$E = K$, but What about *R*?" In *The Routledge Handbook of Evidence*, edited by Maria Lasonen-Aarnio and Clayton Littlejohn. London and New York: Routledge.

———. 2021b. "Epistemological Ambivalence." In *Epistemic Dilemmas*, edited by Nick Hughes. Oxford: Oxford University Press.

———. 2021c. "Knowledge, Credence, and Strength of Belief". In *Expansive Epistemology: Norms, Action, and the Social World*, edited by A. K. Flowerree and Baron Reed. London and New York: Routledge.

European Journal of Analytic Philosophy EuJAP | Vol. 18 | No. 1 | 2022
https://doi.org/10.31820/ejap.18.1.3

DEBUNKING DOXASTIC TRANSPARENCY

Ema Sullivan-Bissett[1]

[1] University of Birmingham

Original scientific article – Received: 28/11/2022 Accepted: 9/2/2022

ABSTRACT

In this paper I consider the project of offering an evolutionary debunking explanation for transparency in doxastic deliberation. I examine Nicole Dular and Nikki Fortier's (2021) attempt at such a project. I suggest that their account faces a dilemma. On the one horn, their explanation of transparency involves casting our mechanisms for belief formation as solely concerned with truth. I argue that this is explanatorily inadequate when we take a wider view of our belief formation practices. I show that Dular and Fortier overstate the extent to which adaptive non-evidentially supported beliefs are rare, and the implausibility of disjunctive evolutionary systems. They should allow a role for the non-truth directed behaviour of our mechanisms of belief formation. On the other hand, we might restrict the explanation offered by Dular and Fortier to the deliberative context, that is, we might understand them as allowing for non-evidential belief formation outside of the deliberative context, but as identifying the key to explaining transparency in the truth-directed evolutionary mechanisms as they operate in the deliberative context. However, this would land them on the second horn of the dilemma: we would then have no different an explanation to one I have offered elsewhere (2018), an explanation which Dular and Fortier explicitly put aside as engaged in a project different from their own. I finish by briefly considering some broader implications relating to explaining transparency, the nature of belief, and the prospects for pragmatism. I conclude that Dular and Fortier's debunking explanation of transparency bestows an implausible role for truth in fixing our beliefs, or, if it doesn't, then we simply have the restatement of a view explicitly disavowed by the authors. We are left, then, with an explanation we ought not want, or an explanation we already had.

Keywords: *belief; transparency; doxastic deliberation; evolutionary debunking; reasons for belief.*

Correspondence: E.L.Sullivan-Bissett@bham.ac.uk

1. Transparency

Transparency characterizes deliberation over belief: "when asking oneself *whether to believe that p*" one must "immediately recognize that this question is settled by, and only by, answering the question *whether p is true*" (Shah 2003, 447). The structure of doxastic deliberation is such that non-truth related considerations cannot play a conscious role in our coming to a belief: the question *whether to believe that p* collapses into the question *whether p is true*.

It is important to note that the move from the question *whether to believe that p* to the question *whether p is true* is "immediate and not inferential" (Engel 2007, 198). There is no bridging step in doxastic deliberation moving one from the truth of *p* to belief that *p*. Rather, answering the question *whether p* just is to answer the question *whether to believe that p.* It is also worth noting that to accept that transparency characterizes doxastic deliberation is not to over-intellectualize that process, that is, for a subject's deliberation to exhibit transparency, she need not explicitly pose herself the question *whether to believe that p.* Rather, it is enough that her thinking "manifests some recognition that this is the question" in play (Shah 2003, 466).

Dular and Fortier offer an evolutionary debunking explanation of why transparency characterizes our doxastic deliberation. In what follows I will not argue *that* transparency characterizes our doxastic deliberation, but will assume that it does, and focus on how it might be explained.[1]

2. Extant Explanations

Two broad kinds of explanation have been offered of transparency in doxastic deliberation, these explanations work from the agential level; with the believer playing a role in the structure of doxastic deliberation, where that role falls out of the essential nature of belief.

Teleological accounts of belief explain transparency by appeal to the aim a believer adopts in posing the deliberative question. Since, according to this view, belief essentially aims at truth, when in the business of deliberating over *whether to believe that p*, truth considerations are the only ones of relevance, since they are the only ones conducive to achieving belief's aim. I have argued elsewhere that such an explanation places the

[1] For dissent see Conor McHugh (2012, 2013; cf. Archer 2017; Sullivan-Bissett 2017a), and Miriam McCormick (2015; cf. Sullivan-Bissett 2017b).

teleologist on the horns of a dilemma.[2] On the one hand, the so-called aim of belief is not appropriately characterized as an aim, since it does not behave in an aim-like way (it cannot be *weighed* against a believer's other aims (Owens 2003)).[3] On the other hand, if the aim could be weighed (such that my aim to believe truly could be weighed against my aim to form beliefs that will make me happy), it should not be the case that transparency characterizes our deliberation. If belief were governed by an aim, the question *whether to believe that p* would not *collapse into* the question *whether p is true.* So even if there were a weighable aim of truth essential to belief, it wouldn't be at the required strength to explain transparency (Sullivan-Bissett 2018, 3457-3460).

An alternative approach comes from normative accounts of belief, which have it that belief aims at truth only metaphorically. Instead of taking the idea of an aim of belief literally, normativists claim that belief is constitutively norm-governed. Many formulations of an explanatory norm for belief have been offered, but there is some agreement that it must be formulated in terms of one being *permitted to believe* that *p* when *p* is true ('ought' formulations would be too demanding (see Bykvist and Hattiangadi 2007)).

With a permissibility norm in place, one normative explanation for transparency has gone like this: when an agent deliberates on the question *whether to believe that p,* that deliberation is framed by the prescription to believe that *p* only if *p* is true. In doxastic deliberation, an agent expresses her commitment to this prescription, which activates two dispositions: one to be moved by considerations relevant to the truth of *p*, and one which blocks considerations which are irrelevant to the truth of *p* (Shah 2003, 467; Shah and Velleman 2005, 519).

However, as I have argued elsewhere, this explanation does not work. It being permissible to believe that *p* is a very different thing from it being *settled for you* to believe that *p*. If it were a permissibility norm which explained the structure of deliberation, it ought to be possible to nevertheless *not* believe that *p* upon ascertaining that *p* is true, after all, one is merely *permitted* and not *obligated* to believe it. But this non-adherence to the putative norm is not possible, and so the putative norm is not strong

[2] Not to be confused with Nishi Shah's (2003, 460-465) *Teleologist's Dilemma.* He argues that a descriptive account of the nature of belief (as being regulated for truth via some aim) will either be too strong to accommodate non-deliberative belief, or too weak to explain transparency. For more on this, see Asbjørn Steglich-Petersen (2006, 2017a) and Sullivan-Bissett and Lisa Bortolotti (2017).

[3] For discussion on whether the teleological account does indeed face this horn of the dilemma see the back and forth between Steglich-Petersen, and Sullivan-Bissett and Noordhof (Steglich-Petersen 2009; Sullivan-Bissett and Noordhof 2013; Steglich-Petersen 2017b; and Sullivan-Bissett and Noordhof 2017).

enough to generate transparency, and cannot explain it (see Sullivan-Bissett 2018, 3461-3462 and Sullivan-Bissett and Noordhof 2020).

As noted earlier, these accounts proceed by taking transparency to be a result of something the agent *does*, in virtue of some specified essential nature of belief (it being truth-aimed or norm-governed). In this way, they take the essential nature of belief to be relevant to the project of explaining transparency. This is problematic because in both cases, if the relevant aim or norm were doing the work of explaining transparency, there ought to be significantly more wiggle room upon discerning that *p* than there in fact is.[4] Let us turn then to Dular and Fortier's explanation, on which the believer is a much more passive player in the structure of doxastic deliberation.

3. Dular and Fortier's Evolutionary Debunking Explanation

Dular and Fortier aim to level the playing field between evidentialists (who have it that only reasons relating to evidence can be reasons to believe) and pragmatists (who have it that there can be reasons to believe not related to evidence). In particular, they aim to show that transparency does not help us adjudicate between these positions. To do so, they seek to provide an explanation of transparency consistent with pragmatism. At first blush, if we want transparency to be explainable by, or even consistent with, our account of epistemic reasons, evidentialism has a much easier time of it. For an evidentialist, reasons for believing that *p* are restricted to considerations relating to evidence supporting the truth of *p*, there are no non-evidential related reasons for belief. A natural thought then is that transparency characterizing our doxastic deliberation is simply evidentialism in action. Since evidential reasons are the only kind of reasons we can have for belief, of course only such reasons enter into deliberation over whether to believe a given proposition. There is no presence of pragmatic considerations in deliberation of this kind since no such considerations could represent *reasons for belief*. Of course, transparency ought not be understood as the claim that one "cannot so much as be struck by pragmatic considerations" (McHugh 2013, 449), the point is rather about which considerations can be motivationally efficacious. Evidentialism seems nicely aligned with doxastic deliberation so characterized, and might even find support in it (see e.g. Shah 2006).

[4] Not all explanations of transparency which draw on the essential nature of belief will be problematic in this way. For example, Kate Nolfi's (2015) action-oriented functional account of belief might naturally explain transparency by appeal to the essential functional role belief plays in a believer's mental economy. (To my knowledge Nolfi hasn't explicitly put her account to such work, but an explanation along these lines might not face the problems of the teleological and normative accounts.)

Again, at first blush, pragmatism looks to be in trouble. If there are ever non-evidential reasons for belief, why can we not take those into consideration when deliberating over *whether to believe that p?* It is not merely that they do not arise for us, rather, we can ostensibly recognize pragmatic considerations for believing (or not), but such considerations lack any motivational force when it comes to settling one way or another, they do not rise to the level of *reasons*. The pragmatist then owes us an explanation of why, from the first-person point of view, only one kind of reason (the evidential kind) gets to determine the outputs of doxastic deliberation.

Dular and Fortier's approach is to understand transparency as the result of evolutionary selection that, in its crudity, could not set us up more precisely in a way which reflected the truth about the nature of reasons to believe. Evolution has set us up to respond only to the class of reasons which more reliably produce adaptive beliefs (i.e. those that increase our inclusive fitness). That class is the one related to evidence, for obvious reasons. As W. V. O. Quine pointed out, "creatures inveterately wrong in their inductions have a pathetic but praiseworthy tendency to die before reproducing their kind" (Quine 1985, 39). So, Dular and Fortier argue, transparency results from our mechanisms of belief production being selected for being geared towards truth, such that non-evidential considerations cannot enter our deliberation *as* reasons to believe. Here is their explanation in full:

> In almost every case, only taking evidential considerations into account will be more fitness-conducive than if we instead formed beliefs according to how well they would advance our ends. For instance, if I am deliberating about whether to believe that there is a tiger over the hill, and the evidence available to me suggests that there is in fact a tiger over the hill, then it is hard to see how deliberating in a way that takes my ends into account (rather than strictly truth) would be more evolutionarily advantageous than simply deliberating in a way that respects the evidence. […] since beliefs based on evidentialist considerations are more likely to result in true beliefs than beliefs that are based on pragmatist ones, the pragmatist has an explanation for why people in fact collapse the question of whether to believe that *p* into the question whether *p.* It is not that, by doing so, they are incredibly adept at picking up on epistemic norms; rather, it's that evolution has selected for the cognitive system that would best promote our survival, and this cognitive system is one that is responsive to truth. (Dular and Fortier 2021, 1461-1462)

This is consistent with pragmatism understood as an account of how we *should* form beliefs since the explanation of transparency is concerned with the descriptive question of how we in fact do. Pragmatists then, unlike evidentialists, simply fail to find their account of reasons to believe reflected in the structure of doxastic deliberation. But that's a matter of contingency—evolution went with the route which would more often produce adaptive beliefs. True beliefs are usually the adaptive ones. The cost is our not being able to consciously respond to pragmatic reasons for belief, but that is not to say that there are no such reasons.

It might be thought that the kind of explanation on offer here misses something crucial about the explanatory project, namely, that the phenomenon to be explained is necessary to belief, and not some evolutionary contingency. I have argued elsewhere that the claim that transparency is a necessary phenomenon of belief is ill-motivated, and so explanations which do not honour necessity are not vulnerable to objection based on not so honouring (Sullivan-Bissett 2018, 3471-3472). Suffice to say for the sake of this paper that I am not in the business of objecting to the contingent nature of transparency. In discussing explanations which have contingency as an outcome, my argument is downstream of any debate concerning the reasonableness of weakening the metaphysical strength of transparency.

In the next two sections I argue that Dular and Fortier's approach finds itself on the horns of a dilemma. On the one horn, their evolutionary account of the relationship between belief and truth fails to properly accommodate the range of beliefs formed on non-evidential grounds (§4). One way to avoid this charge is to restrict the account to the context of doxastic deliberation. However, this places their account on the second horn—it is no different from an account put forward elsewhere which they explicitly distance themselves from as one engaged in a different explanatory project (§5). More broadly, a wider view of the nature of belief is needed to contextualise the explanation of transparency, and the details of this wider view will be relevant to whether evidentialism and pragmatism are in fact on a theoretical par (§6).

4. Horn One: Non-truth Directed Beliefs

Dular and Fortier often talk about our "cognitive system" responsible for beliefs as a whole. For example, they say "evolution has selected for the cognitive system that would best promote our survival, and this cognitive system is one that is responsive to truth" (2021, 1462), and that "we still ought to think that natural selection favors systems that generate true

beliefs" (2021, 1463). However, the claim that there's a single aim for mechanisms of belief formation (truth) leaves no room for the idea that in certain contexts, these mechanisms are not truth-aimed, but are nevertheless doing what they ought to be doing. In some cases, the beliefs it is adaptive to have are not aimed at truth. Dular and Fortier seem to disagree, they say, "notice how *rare* cases are where having a non-evidentially supported belief would be better for survival" (2021, 1461). I think they underestimate how many human beliefs are adaptive in spite of their not being aimed at truth.

There are many cases of false belief which seem to have been produced by mechanisms doing what they are *supposed to do*. That is, producing beliefs that are adaptive, but not in virtue of being true. Such cases reveal that our mechanisms for belief formation are not solely concerned with truth. Let us consider some examples.

We can think first in terms of biases, and distinguish hot biases (those driven by emotions and desires) and cold biases (those which are non-motivational and emerge from heuristics).[5] Cold biases include the availability heuristic (our tendency to rely on examples which easily come to mind as representative) and the confirmation bias (our tendency to search for or interpret information that confirms our existing beliefs) (for more see Mele 1997, 93–94). Such biases may be present in our cognitive systems because they tend to guide us towards truth, and so we may find no reason from these quarters to think our beliefs are aimed at anything but truth. Hot biases however are a different matter; these are biases arising from our desiring that *p* or desiring to *believe that p*. Examples include selective attention to evidence, and selective evidence gathering (for more see Mele 1997, 94–5, 2006, 110). It is biases such as these that might play a role in the belief-formation processes resulting in false beliefs which are nevertheless useful, for example, self-deceptive beliefs.[6]

It could be said in response to the foregoing that cases of useful false beliefs are not plausible candidates for adaptive beliefs produced by mechanisms doing what they are supposed to do, but are rather examples of our mechanisms for belief production going awry (and getting lucky in producing helpful beliefs). But there are good reasons to think this is not the case. For example, unrealistic beliefs about oneself lead to engagement

[5] This distinction has been traced back to R. P Abelson (1963) (see Elster 1983, 141).

[6] They may do so in at least two broad kinds of way. On an anti-intentionalist story, the hot biases influencing the formation of the self-deceptive belief are not intentional (see e.g. Mele 1997, 2006). More controversially perhaps, on an intentionalist story, a subject may intentionally bias her cognitive processes so as to favour a belief that *p* (see Talbott 1995). I favour an anti-intentionalist account, and some of what I say in what follows may more naturally fit such a framework.

in adaptive behaviours (Taylor and Brown 1994), and "optimal mental health is associated with *unrealistically positive* self-appraisals and beliefs" (see McKay and Dennett 2009, 505–508 for discussion). As Nolfi has pointed out in her discussion of the optimism bias, research

> supplies compelling evidence that we are, as a general rule, more successful in achieving our various ends when our beliefs about ourselves and about our relationship to the world around us are systematically distorted in particular ways. (Nolfi 2018, 192)

Self-deceptive beliefs may also be adaptive by (1) helping us to be better deceivers, (2) preventing parent-offspring conflict to maintain parental investment, or (3) for the same reasons as biases towards optimism, i.e. the various benefits of a positive stance (Trivers 2000, 2011, 2013, cf. Van Leeuwen 2007a). Other examples might include our moral beliefs, which are adaptive insofar as they facilitate cooperation and social cohesion (Ruse 1986; Joyce 2001; cf. Enoch 2011; Fitzpatrick 2014), and beliefs about epistemic normativity which are adaptive insofar as they make us better at responding to epistemic considerations (Street 2009; Sullivan-Bissett 2017c, 2020). Slightly less plausible perhaps are clinical delusions,[7] which might be adaptive insofar as they help to deceive others into social alliances (Hagen 2008; cf. Gold and Gold 2014), or insofar as they maintain behavioural interactions in the face of abnormal prediction-error signaling[8] (Fineberg and Corlett 2016; Mishara and Corlett 2009; cf. Lancellotta 2021). All of these cases are ones where the cognitive mechanisms responsible are not responding to truth, but are nonetheless producing adaptive beliefs.[9] Perhaps there are more besides (candidates might include beliefs arising from conspiratorial ideation or confabulation). It would be surprising if all of these cases were ones where things had gone *wrong*, especially given how widespread some of them

[7] Of course, any serious defence of the adaptiveness of delusions will need to attend to heterogeneity in this category (see Bortolotti 2009, 23-27 for an overview; see also Lancellotta and Bortolotti for discussion of adaptiveness informed by McKay and Dennett's 2009 shear pin hypothesis). Since I am not defending any of these views but referring to them only to illustrate the range of candidate adaptive beliefs, I won't discuss this further.

[8] Prediction-error theories have it that perceptual processing involves the generation of predictions about sensory input, from antecedently held perceptual hypotheses about the world, with the aim of updating the hypotheses to minimize the error of these predictions on the basis of comparison between the predictions and sensory input. Delusions are said to derive from the malfunctioning of this process, for example, faulty signals that a prediction isn't met leading to erroneous updating.

[9] I have been speaking as though these beliefs arise from the *same mechanisms* as those beliefs which seem to respond to evidence, or which are seeking to track truth. Of course, something rather different might be going on. It might be that the mechanisms producing adaptive beliefs (but not in virtue of their aiming at truth) are different from the mechanisms producing beliefs aimed at truth. If that were right, we would nevertheless have cases where the existence of false beliefs would be in virtue of design, not mistake.

are, and how they contribute positively to inclusive fitness. The existence of these beliefs looks, then, to be in virtue of design not mistake.[10] Of course, given that deliberative belief formation is characterized by transparency, our mechanisms of belief formation must be operating below the level of consciousness when they aim at something other than truth (cf. Talbott 1995). But their doing so gives us sufficient reason to be sceptical that such mechanisms are, in general, solely geared towards truth.

Dular and Fortier consider a nearby objection to their view which has it that evolution would in fact favour a cognitive system that had us believe in accordance with the evidence unless it would be more conducive to survival not to. They call this a "disjunctive" system. In response, they say that "this proposed system is simply too complex for evolution to have selected" (Dular and Fortier 2021, 1464). They go on:

> [I]t seems as though if evolution had favored the disjunctive system, we would also need *another* system that would help us determine when to believe in accordance with evidential considerations, and when to believe in accordance with non-evidential considerations. Given the rarity of cases in which non-evidential reasons for belief do a better job at promoting survival than evidential reasons for belief, coupled with the fact that the disjunctive system is so complex, we have good reason to think that natural selection would have favored the psychologies that we in fact have, rather than ones that employ the disjunctive system. (Dular and Fortier 2021, 1464)

Let us understand a *disjunctive evolutionary system* as one which performs multiple functions, depending on the circumstances of the organism. A single belief system which produced beliefs aimed at the truth in some circumstances, and produced beliefs aimed at usefulness unconnected to truth in others, would count as a disjunctive system in this sense. Dular and

[10] Of course, none of these views are without their critics (as indicated in the citations), but the sheer number of evolutionary adaptation accounts of various kinds of belief exert enough pressure on Dular and Fortier's claim that non-evidentially formed beliefs promoting survival represent a rarity (Dular and Fortier 2021, 1461). Indeed, as William J. Talbott notes, although knowing the truth can enhance our ability to fulfil our desires, there are many cases where it does not, and indeed may produce highly undesirable psychological effects, and so many adaptive beliefs are ones arising out of hot cognition (1995, 27-28). Another way of resisting the force of these putative examples of false but adaptive beliefs is to endorse non-doxasticism about them. If these attitudes are not beliefs, then they were not produced by mechanisms of belief production, and cannot be harnessed in any argument about the nature of those mechanisms. Dular and Fortier may well take this line, although it is unclear to me what the principled grounds would be, and usual ones which motivate non-doxasticism to do with relationships to evidence may not be natural ones for pragmatists to take. In any case, although far from settled, I refer readers to defences of doxasticism elsewhere (for delusion, see Bortolotti 2009 and Bayne and Pacherie 2005; for self-deception, see Van Leeuwen 2007b).

Fortier have it that the rarity of cases of false but useful beliefs gives us reason to think there is no such system operative in our mechanisms of belief formation.

However, if I am right that at least some of the beliefs surveyed above are examples of false but adaptive beliefs produced by design and not mistake, that right there is our evidence that a disjunctive system is not so complex after all. So how would it work? We might say something like this: at the coarsest level of description, our mechanisms for belief production have the function of producing *adaptive* beliefs. But they do that in one of two ways: by producing beliefs aimed at truth, or by producing beliefs which facilitate some other good but not as an approximation to truth. These two ways of producing adaptive beliefs might map onto two sets of historically appropriate conditions for the performance of the two respective functions. The historically appropriate conditions for the production of true beliefs plausibly diverge from the conditions appropriate for the production of beliefs not aimed at truth.

Take self-deception for example. In the case of wishful self-deception the subject desires that *p*, has ample evidence that not-*p*, and goes on to believe that *p*. Such cases are ones where believing that not-*p* might be psychologically damaging, and believing that *p* would prevent that damage and perhaps bring some additional psychological benefits. Plausibly a motivational condition is met in such situations, whereby mechanisms of belief production are *thus* geared not towards truth, but towards the effective functioning of the subject (see also Noordhof 2003 for related cases that don't involve self-deception). What is so implausible about this?

Dular and Fortier ask us to believe in a system geared solely towards truth because adaptive non-truth directed beliefs are rare, and a system which tracked truth sometimes and at other times did not would be too complicated to be favoured by evolution. We have seen though that there is a large swathe of candidates for false but adaptive beliefs, and, given that, we have reason to think that a disjunctive system is both possible and actual. This is an important point because Dular and Fortier ward off an objection to their view by appeal to the unlikelihood of disjunctive evolutionary systems. If such systems are not unlikely after all, Dular and Fortier owe an explanation of why pragmatism about reasons to believe is not better reflected in our practices of belief formation.

5. Horn Two: Restricting to the Deliberative Context

In reply to the foregoing Dular and Fortier may restrict their view to the context of deliberation where any non-evidential considerations cannot be consciously considered. That would leave it open that in the non-deliberative context, our mechanisms of belief formation could produce adaptive beliefs, but not by aiming at truth. This looks like a move they make in a footnote when they say:

> [O]ne might think that evolutionary pressures make us messy, such that sometimes we would form beliefs based on pragmatic considerations instead of, as we've claimed here, never doing so (given the truth of transparency). However, notice that transparency is a claim about explicit doxastic deliberation, and so our evolutionary explanation only concerns this context. Notice that this is compatible with people sometimes forming beliefs based on pragmatic considerations outside of the context of explicit doxastic deliberation. (Dular and Fortier 2021, 1464, fn. 38)

Here, Dular and Fortier seem not to be against our mechanisms for belief formation being a bit more pragmatic in some contexts, perhaps consistent with some of the ways suggested in the previous section. But in the absence of a disjunctive deliberation whereby we could appropriately regulate our responses to evidential and pragmatic considerations, from the first-person perspective, evolution had a choice: have us respond to only evidential considerations in deliberation, or only pragmatic ones. I think that Dular and Fortier are quite right when they say:

> [I]t is implausible that including non-evidential reasons in our doxastic deliberations would *generally* promote our survival better than only the evidentialist reasons, and, as it's widely known, evolution selects for general systems rather than more fine-grained ones. (Dular and Fortier 2021, 1462)

However, if the claim is that evolution selected for doxastic deliberation being restricted to epistemic considerations, and that's consistent with there being other mechanisms in play in other contexts where pragmatic considerations are involved, then the explanation here is no more than what is present elsewhere (notably Sullivan-Bissett 2018). Now of course, philosophers can happen upon the same view independently in ignorance of one another, such occurrences represent no good grounds for objection. However, this is not such a case. Dular and Fortier mention my argument in an aside, saying that

> [t]hough addressing the question of what distinguishes belief as a unique mental state, and not the topic of epistemic reasons, Sullivan-Bissett (2018) also gives an evolutionary explanation of transparency to defend her motivational account of the nature of belief. (Dular and Forter 2021, 1461, fn. 22)

It is on these grounds that they put the account aside.

But the account cannot so easily be put aside if Dular and Fortier are badging themselves as providing an evolutionary debunking account of transparency which does not generalize to belief formation in general, that is, an explanation which doesn't rule out non-evidential influences on belief outside of the deliberative context. And that is because I do *exactly the same thing.* It is false that my account of transparency is one put forward in the spirit of defending a motivational account of belief's nature, and so it cannot be put aside on the grounds that it is engaging in some different explanatory project. I adopt a motivational account of the essential nature of belief, taking up the characterization of the role suggested (but not endorsed) by Lucy O'Brien: all that is necessary for a state to be a belief is that it

> *by itself*, and relative to a fixed background of desires, disposes the subject to behave in ways that would promote the satisfaction of his desires if its content were true. (O'Brien 2005, 56)

This constitutive feature of belief simply sits in the background of my account of transparency. The work done by the adoption of the motivational view is that of picking out the essential nature of belief across worlds in a way which does not appeal to the connection between belief and truth. As we saw earlier, other explanations have taken transparency to fall out of what is essential to belief (its truth aim or norm). My approach is to divorce the essential nature of belief from truth, and to move away from seeing transparency as a consequence of the nature of belief. Instead, I ground the explanation of transparency in the particular (biological) circumstances of human belief formation. With respect to the motivational account being able to explain transparency, I am very clear that it cannot:

> [T]he account under consideration gives us a gap between *whether p is true* and *whether to believe that p.* But there is no such gap in doxastic deliberation, truth is not an optional end. A motivational account of belief does not have the resources to explain why this is the case. (Sullivan-Bissett 2018, 3465)

It is false, then, that my explanation of transparency is offered in the service of defending the motivational role component of my overall view. So what can explain transparency then? I explained it like this:

> Given that our mechanisms for belief-production have the relational proper function of producing devices with true contents, the story for their selection includes their Normally producing true beliefs. In the deliberative case our cognitive architecture is arranged such that when we deliberate as to *whether to believe that p*, we, at the agent-level, are only sensitive to the adaptor (our environment) because this makes the adapted device (the resulting belief) more likely to perform its derived proper function of being true. The neurological structures which secure transparency have been selected for their role in producing true beliefs. (Sullivan-Bissett 2018, 3470)

Now, notwithstanding some of the terms used to pick out what is going on, this strikes me as indistinguishable from the explanation offered by Dular and Fortier. Both explanations identify transparency as an evolutionary adaptation designed to guide our doxastic deliberation towards evidential considerations in the pursuit of true beliefs. Both explanations take beliefs guided as such to be more evolutionarily advantageous than beliefs which could be formed, deliberatively, on the basis of non-evidential reasons. So in their casting transparency as a product of evolution gearing our doxastic deliberation towards truth, Dolar and Fortier offer no more than is offered in the above explanation. Making it explicit that such an explanation is consistent with pragmatism is a point worth noting, but that ought not be confused with having offered "an alternative explanation of transparency" (Dular and Fortier 2021, 1457) as advertised. The point isn't merely that Dular and Fortier have offered an identical explanation as offered elsewhere, it is that their having done so is based in part on a misinterpretation of an account which is, in fact, identical to their own.

6. Broader Considerations

I want to close with some broader considerations on the nature of belief, transparency, and the prospects for pragmatism. On the face of it, transparency may seem to tell us something about the nature of belief (that it is constitutively linked to truth) and about the nature of reasons for belief (that they are linked to truth, or evidence suggestive of it). On its face, transparency thus looks friendly to fans of a constitutive belief-truth link, and fans of evidentialism.

We have seen that my approach is one that denies the *constitutive link* claim, while Dular and Fortier are interested in denying the *exclusivity of evidential reasons* claim. We have also seen that the strategy for doing so is—contra Dular and Fortier—exactly the same. The explanation given of transparency is one which has it as a phenomenon arising out of evolutionarily selected mechanisms of belief formation. For me, the lesson is that transparency does not follow as a result of something the believer does, given the essentially truth-related nature of belief, but is rather something to which the believer passively responds. For Dular and Fortier, explaining transparency as the consequence of evolutionarily selected mechanisms of belief formation allows them to divorce the nature of reasons to believe from the reasons evolution has set us up to recognize as reasons to believe in the deliberative context. However, as I have argued, this point need not rely on the claim that disjunctive evolutionary systems are implausible. Indeed, I take it that my account of transparency is perfectly friendly to pragmatism in precisely the way that Dular and Fortier take their account to be (even if, as it happens, I favour an error theoretic approach to reasons for belief, see my 2017c and 2020). It should thus be clear that the evolutionary debunking explanation of transparency is consistent with pragmatism, without entailing it.

Indeed, whether an explanation of transparency is consistent with a particular view of the nature of epistemic reasons will depend on broader considerations. That is, it will depend on the background notion of belief in operation. For the evidentialist this hardly needs stating: they will typically allow their preferred theory of epistemic reasons to carry the explanatory burden of transparency, set nicely against a background of belief as essentially connected to truth, either through a constitutive aim or norm. For the pragmatist, things are not as straightforward, since their preferred theory of reasons for belief does not so easily lend itself to an explanation of transparency (that explanation has to be located elsewhere). My conception of belief is ontically thin—it's simply an attitude which plays a given motivational role, all else equal. Given its thinness, the essential nature of belief is not apt to generate norms of good belief regulation, and so is silent on the debate between evidentialists and pragmatists. However, not all accounts of belief which de-prioritize truth and evidence will have this result.

Nolfi's work demonstrates this point. In particular, her approach teaches us that the rejection of a constitutive role for truth in belief does not amount to an endorsement of pragmatism, nor does it amount to a rejection of the idea that belief is constitutively normative. Rather, Nolfi's proper function account of belief has it that the relevant constitutive norm for belief is not one related to truth, but rather to the particular role that beliefs play in a

subject's mental economy (Nolfi 2015, 197). Unlike my view, her action-oriented function of belief is taken to be essential to it, and out of that falls the evidential constraint on epistemic reasons (Nolfi 2018, 188). The relationship between Nolfi's framework and what she calls *straightforward pragmatism* is not uncomplicated, but one of her examples helps illustrate the difference:

> Bella knows that she can secure substantial benefit or avoid substantial harm merely by believing that 2343 is a prime number, a false proposition, and one in support of which Bella has absolutely no evidence. Bella's future actions will be significantly more successful if she believes that 2343 is prime than if she fails to so believe. (Nolfi 2021, 11305)

As Nolfi notes, a straightforward pragmatist would say of this case that Bella ought to form the belief in question, given its practical advantages, and her doing so would not warrant negative epistemic evaluation. On Nolfi's action-oriented epistemology though, while we can recognize that the belief in question would facilitate successful action for Bella, we can also recognize that negative epistemic evaluation would be appropriate since the belief's action-related benefits do not survive contextual variation, and so the belief is not well-suited to fulfill the action-enabling function constitutive of belief (Nolfi 2021, 11305). The epistemic criticism that would be appropriate here comes from the nature of the pattern of belief-regulation responsible for the formation of Bella's belief—such a pattern would normally, as Nolfi puts it, "have gotten her into trouble" (Nolfi 2021, 11306). Similar things might be said about Dular and Fortier's own example of Jungkook, someone whose wrist will heal more quickly if they were to believe that it would heal faster than studies have shown (Dular and Fortier 2021, 1470). Dular and Fortier take it that in this case, despite the evidence suggesting Jungkook's wrist won't heal in fewer than three months "Jungkook should believe that they will heal in less than three months" (2021, 1470). But whether this intuition is shared will depend on the background notion of belief in play. De-prioritizing the role for truth or evidence in belief by making the relationship a contingent feature of evolutionary pressures doesn't automatically generate verdicts in favour of pragmatism.

So we see that breaking the constitutive link between belief on the one hand and truth or evidence on the other does not straightforwardly pave a way to pragmatism. What also matters is the essential nature of belief which forms part of the account. My explanation of transparency is silent on the nature of reasons for belief, but Nolfi's account which also does away with a constitutive belief-truth link is explicitly not pragmatism-

friendly. Her conception of belief is thicker than my own, identifying it as an attitude that is *meant* to play a particular role in a believer's mental economy (Nolfi 2015, 197), and of course there are better and worse ways for something to play a role.

Dular and Fortier are explicitly silent on the essential nature of belief (2021, 1457, fn. 8). Whether their overall approach might be harnessed in a rebalancing of the prospects of pragmatism in the light of transparency will in part depend on what they take the essential nature of belief to be.

7. Conclusions

In defending pragmatism from the *prima facie* problem of transparency, Dular and Fortier turn to evolutionary debunking to explain this phenomenon as an evolutionary fix not properly sensitive to the truth about reasons to believe. I argued that taken one way—as a claim about our mechanisms for belief formation *simpliciter*—Dular and Fortier's explanation took a whole host of our beliefs outside of the scope of an evolutionary account, unless they have it that such cases are malfunctions. I argued that that would not be wise. I suggested that their dismissal of the plausibility of disjunctive evolutionary systems was too quick, and so they owe an explanation of why our practices of belief formation do not reflect pragmatism. On the other hand, if Dular and Fortier are happy to have non-evidential mechanisms as part of the story too, then the quick explanation they give of transparency (as being evolution's fix for adaptive belief formation), offers no more than is offered elsewhere (in a paper they say they diverge from the aims of). Overall then, either the cost of debunking transparency results in an implausible role for truth in fixing our beliefs across contexts, in which case the alternative explanation offered is not one we should want. Or we have merely the restatement of a view explicitly disavowed by the authors, in which case we have no alternative explanation after all. Finally, I argued that to be consistent with pragmatism, it matters what background theory of belief informs the explanation given of transparency. In particular, whether the harnessing of evolutionary considerations to soften the link between beliefs on the one hand and truth and evidence on the other is consistent with pragmatism will depend in part on taking a wider view of the essential nature of belief, a view not taken by Dular and Fortier.

Acknowledgments

With thanks to the Arts and Humanities Research Council for funding the research of which this paper is a part (*Deluded by Experience*, grant no. AH/T013486/1). I am also grateful to two reviewers for this journal for helpful comments which improved the paper.

REFERENCES

Abelson, R. P. 1963. "Computer Simulation of Hot Cognition". In *Computer Simulation of Personality*, edited by Silvan S. Tomkins and Samuel Messick, 277–298. New York: Wiley.

Archer, Sophie. 2017. "Defending Exclusivity". *Philosophy and Phenomenological Research* 94 (2): 326–341. https://doi.org/10.1111/phpr.12268

Bayne, Tim, and Elisabeth Pacherie. 2005. "In Defence of the Doxastic Conception of Delusion". *Mind & Language* 20 (2): 163–188. https://doi.org/10.1111/j.0268-1064.2005.00281.x

Bortolotti, Lisa. 2009. *Delusions and Other Irrational Beliefs.* Oxford: Oxford University Press.

Bykvist, Krister, and Anandi Hattiangadi. 2007. "Does Thought Imply Ought?" *Analysis* 67 (4): 277–285. https://doi.org/10.1093/analys/67.4.277

Dular, Nicole, and Nikki Fortier. 2021. "Epistemic Reasons, Transparency, and Evolutionary Debunking". *Philosophia* 49: 1455–1473. https://doi.org/10.1007/s11406-020-00302-6

Elster, Jon. 1983. *Sour Grapes.* New York: Cambridge University Press.

Engel, Pascal 2007. "Belief and Normativity". *Disputatio* 2 (23): 179–204. https://doi.org/10.2478/disp-2007-0009.

Enoch, David. 2011. *Taking Morality Seriously: A Defense of Robust Realism.* New York: Oxford University Press.

Fineberg, Sarah K., and Philip R. Corlett. 2016. "The Doxastic Shear Pin: Delusions as Errors of Learning and Memory". *Cognitive Neuropsychiatry* 21 (1): 73–89. https://doi.org/10.1080/13546805.2015.1136206

Fitzpatrick, William J. 2014. "Why There is No Darwinian Dilemma for Ethical Realism". In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain, 237–255. New York: Oxford University Press.

Gold, Joel, and Ian Gold. 2014. *Suspicious Minds: How Culture Shapes Madness.* New York: Free Press.

Hagen, Edward H. 2008. "Non-bizarre Delusions as Strategic Deception". In *Medicine and Evolution*, edited by Sarah Elton and Paul O'Higgins, 181-216. Boca Raton: CRC.

Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.

Lancellotta, Eugenia. 2021. "Is the Biological Adaptiveness of Delusions Doomed?" *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-021000545-6

Lancellotta, Eugenia, and Lisa Bortolotti. 2020. "Delusions in the Two-Factor Theory: Pathological or Adaptive?" *European Journal of Analytic Philosophy* 16 (2): 37-57. https://doi.org/10.31820/ejap.16.2.2

McCormick, Miriam Schliefer. 2015. *Believing Against the Evidence: Agency and the Ethics of Belief*. Abingdon: Routledge.

McHugh, Conor. 2012. "Beliefs and Aims". *Philosophical Studies* 160 (3): 425–439. https://doi.org/10.1007/s11098-011-9728-z

———. 2013. "The Illusion of Exclusivity". *European Journal of Philosophy* 23 (4): 1117–1136. https://doi.org/10.1111/ejop.12032

McKay, Ryan, and Daniel Dennett. 2009. "The Evolution of Misbelief". *Behavioral and Brain Sciences* 32: 493–561. https://doi.org/10.1017/S0140525X09990975

Mele, Alfred R. 1997. "Real Self-deception". *Behavioral and Brain Sciences* 20: 91–136. https://doi.org/10.1017/S0140525X97000034

———. 2006. "Self-deception and Delusion". *European Journal of Analytic Philosophy* 2(1): 109–124.

Mishara, Aaron L., and Philip R. Corlett. 2009. "Are Delusions Biologically Adaptive? Salvaging the Doxastic Shear Pin". *Behavioral and Brain Sciences* 42 (6): 530–531. https://doi.org/10.1017/S0140525X09991464

O'Brien, Lucy. 2005. "Imagination and the Motivational Role of Belief". *Analysis* 65 (1): 55–62. https://doi.org/10.1093/analys/65.1.55

Nolfi, Kate. 2015. "How to be a Normativist about the Nature of Belief". *Pacific Philosophical Quarterly* 96: 181–204. https://doi.org/10.1111/papq.12071

———. 2018. "Why Only Evidential Considerations Can Justify Belief". In *Normativity: Epistemic and Practical*, edited by Conor McHugh, Jonathan Way, and Daniel Whiting, 179–199. Oxford: Oxford University Press.

———. 2021. "Epistemically Flawless False Beliefs". *Synthese* 198: 11291–11309. https://doi.org/10.1007/s11229-020-02787-w

Noordhof, Paul. 2003. "Self-Deception, Interpretation and Consciousness". *Philosophy and Phenomenological Research* 67 (1): 75–100.

https://doi.org/10.1111/j.1933-1592.2003.tb00026.x

Owens, David J. 2003. "Does Belief Have an Aim?" *Philosophical Studies* 115 (3): 283–305. https://doi.org/10.1023/A:1025157101089

Quine, W. V. O. 1985. "Natural Kinds". In *Naturalizing Epistemology*, edited by Hilary Kornblith, 57–76. Cambridge, MA: MIT Press.

Ruse, Michael. 1986. "Evolutionary Ethics: A Phoenix Arisen". *Zygon* 21 (1): 95–112. https://doi.org/10.1111/j.1467-9744.1986.tb00736.x

Shah, Nishi. 2003. "How Truth Governs Belief". *The Philosophical Review* 112 (4): 447–482. https://doi.org/10.1215/00318108-112-4-447

———. 2006. "A New Argument for Evidentialism". *The Philosophical Quarterly* 56 (225): 481–98. https://doi.org/10.1111/j.1467-9213.2006.454.x

Shah, Nishi and David J. Velleman. 2005. "Doxastic Deliberation". *The Philosophical Review* 11 (4): 497–534. https://doi.org/10.1215/00318108-114-4-497

Steglich-Petersen, Asbjørn. 2006. "No Norm Needed: On the Aim of Belief". *The Philosophical Quarterly* 56 (225): 499–516. https://doi.org/10.1111/j.1467-9213.2006.455.x

———. 2009. "Weighing the Aim of Belief". *Philosophical Studies* 145: 395–405. https://doi.org/10.1007/s11098-008-9239-8

———. 2017a. "Fictional Persuasion and the Aim of Belief". In *Art and Belief,* edited by Ema Sullivan-Bissett, Helen Bradley, and Paul Noordhof, 174–193. Oxford: Oxford University Press.

———. 2017b. "Weighing the Aim of Belief Again". *Logos and Episteme* 8 (1): 141–145. https://doi.org/10.5840/logos-episteme2017819

Street, Sharon. 2009. "Evolution and the Normativity of Epistemic Reasons". *Canadian Journal of Philosophy* 39: 213–248. https://doi.org/10.1080/00455091.2009.10717649

Sullivan-Bissett, Ema. 2017a. "Aims and Exclusivity". *European Journal of Philosophy* 25 (3): 721–731. https://doi.org/10.1111/ejop.12183

———. 2017b. "Transparency Defended". Symposium on Miriam McCormick's *Believing Against the Evidence: Agency and the Ethics of Belief. Syndicate Philosophy.*

———. 2017c. "Biological Function and Epistemic Normativity". *Philosophical Explorations* 20 (1): 94–110. https://doi.org/10.1080/13869795.2017.1287296

———. 2018. "Explaining Doxastic Transparency: Aim, Norm, or Function?" *Synthese* 195 (8): 3453–76. https://doi.org/10.1007/s11229-017-1377-0

———. 2020. “We Are Like American Robins”. In *Epistemic Duties: New Arguments, New Angles*, edited by Scott Stapleford and Kevin McCain, 954–110. Routledge.

Sullivan-Bissett, Ema, and Lisa Bortolotti. 2017. “Fictional Persuasion, Transparency, and the Aim of Belief”. In *Art and Belief*, edited by Ema Sullivan-Bissett, Helen Bradley, and Paul Noordhof, 153–173. Oxford: Oxford University Press.

Sullivan-Bissett, Ema, and Paul Noordhof. 2013. “A Defence of Owens’s Exclusivity Objection to Beliefs Having Aims”. *Philosophical Studies* 163 (2): 453–457. https://doi.org/10.1007/s11098-011-9824-0

———. 2017. “Another Defence of Owens’s Exclusivity Objection to Beliefs Having Aims”. *Logos and Episteme* 8 (1): 147–153. https://doi.org/10.5840/logos-episteme20178110

———. 2020. “The Transparent Failure of Norms to Keep Up Standards of Belief”. *Philosophical Studies* 177: 1213–1227. https://doi.org/10.1007/s11098-019-01242-y

Talbott, W. J. 1995. “Intentional Self-Deception in a Single Coherent Self”. *Philosophy and Phenomenological Research* 55 (1): 27–74. https://doi.org/10.2307/2108309

Taylor, Shelley. E., and Jonathan D. Brown. 1994. “Positive Illusions and Well-being Revisited: Separating Fact from Fiction”. *Psychological Bulletin* 116 (1): 21–27. https://doi.org/10.1037/0033-2909.116.1.21

Trivers, Robert. 2000. “The Elements of a Scientific Theory of Self-deception”. *Annals of the New York Academy of Sciences* 907: 114–31. https://doi.org/10.1111/j.1749-6632.2000.tb06619.x

———. 2011. *Deceit and Self-deception: Fooling Yourself the Better to Fool Others*. London: The Penguin Group.

———. 2013. *The Folly of Fools. The Logic of Deceit and Self-deception in Human Life*. New York: Basic Books.

Van Leeuwen, Neil. 2007a. “The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon”. *Philosophical Psychology* 20 (3): 329–48. https://doi.org/10.1080/09515080701197148

———. 2007b. “The Product of Self-Deception”. *Erkenntnis* 67 (3): 419–437. https://doi.org/10.1007/s10670-007-9058-x

European Journal of Analytic Philosophy EuJAP | Vol. 18 | No. 1 | 2022
https://doi.org/10.31820/ejap.18.1.4

A REPLY TO S SIDDHARTH'S 'AGAINST PHENOMENAL BONDING'

Gregory Miller[1]

[1] Atlanta, Georgia, USA

Original scientific article – Received: 15/10/2021 Accepted: 11/3/2022

ABSTRACT

In this journal S Siddharth has recently argued that the phenomenal bonding response to the subject summing argument for panpsychism is question begging, therefore we should reject constitutive forms of panpsychism. The argument specifically focuses on the proposals of Goff and Miller. In this reply, I show that the argument is unsound.

Keywords: *panpsychism; combination problem; phenomenal bonding; subject summing; combinationism; S Siddharth.*

Introduction

Philip Goff (2016; 2009a) has proposed a form of panpsychism called 'phenomenal bonding panpsychism':

> **Phenomenal bonding panpsychism:** non-fundamental consciousness is constituted by fundamental micro-consciousnesses and the phenomenal bonding relation that holds between them.

This view is supposed to avoid the subject-summing problem for panpsychism, which can be clearly stated in the following way: the existence of a group of *n* subjects does not give rise to a further, *n+1* subject. The phenomenal bonding strategy is to admit that the mere existence of *n* subjects does not give rise to further subjects but suggest that the existence of those subjects and the *appropriate relations between* them may do. The appropriate relation that would bundle a group of micro-

Correspondence: millergregory90@gmail.com

subjects into a larger subject is the *phenomenal bonding relation*, hereafter the PB relation.[1]

According to phenomenal bonding panpsychism, animal subjects are composite subjects. They are subjects of experience that have proper parts, as an animal or a brain does, but, importantly, their proper parts are also subjects of experience. For example, according to the phenomenal bonding panpsychist, human animals are conscious subjects, their micro-level parts are conscious subjects, and so too are large undetached proper parts of their brains.

In addition to this, animal subjects share overlapping sets of experiences with their proper parts. In other words, according to the phenomenal bonding panpsychist, if you had some set of experiences {e1, e2, e3 … e100}, then your proper parts also have subsets of those same token experiences (or the experiences which constitute them). Phenomenal bonding panpsychism is, therefore, a form of what Roelofs (2019) has called *combinationism*—the view that non-fundamental consciousness is grounded in, and explained by, combinations of fundamental consciousnesses and the real relations between them (Roelofs 2019, 6).

Whilst Goff holds that we must be mysterians about the PB relation and settle for a concept of phenomenal bonding that is merely role-playing, Gregory Miller (2017) has offered an improvement. Miller argues that we can have a positive not-merely-role-playing concept of the PB relation, i.e., a conception of the relation *over and above* the role it plays within the panpsychist theory. The positive conception that Miller proposes is the *co-consciousness* relation, i.e., the phenomenal unity relation in virtue of which sets of experiences have a conjoint phenomenology (Dainton 2000; Lockwood 1989).

In a recent article in this journal S Siddharth has argued against phenomenal bonding panpsychism. His argument is that Goff's version of phenomenal bonding panpsychism is question begging without a positive conception of the PB relation, i.e., without Miller's proposal, and that Miller's proposal fails.

Let's call his argument the *Begging the Question Argument*, and formalise it in the following way:

[1] Goff suggests that the PB relation should be the 'dual aspect' of some physical relation.

The Begging the Question Argument

(1) If phenomenal bonding panpsychism is to be a non-question begging solution to the subject summing argument, then it requires a positive not-merely-role-playing conception of the phenomenal bonding relation.
(2) We cannot form a positive not-merely-role-playing conception of the phenomenal bonding relation.
(3) Hence, phenomenal bonding panpsychism is a question begging response to the subject summing argument.

This argument is clearly valid, so the only way to respond is by questioning its two premises. Fortunately, both (1) and (2) are unjustified or false.

Below I will explain why this is the case. But we must first note two essential but unjustified assumptions underlying Siddharth's argument.

Two Assumptions: Not Shared and Not Justified

Let us clearly state two of the assumptions that Siddharth makes about conscious subjects and their experiences at the outset of his argument. This is important because these are two *very strong* assumptions about subjects which are essential to his justifications for both (1) and (2). Moreover, both are assumptions that directly conflict with phenomenal bonding panpsychism.

He states that we have "*intuitions* about the nature of subjects" (Siddharth 2021, 7) that inform us of two key features about them: the unity of subjects, and the privacy of their experiences. So, we have a 'unity intuition' and a 'privacy intuition'.

On the unity intuition, Siddharth claims that subjects seem to be a unity that cannot be divided into proper parts. He says, "subjects *seem* to be *ontological unities* (…) utterly indivisible" (Siddharth 2021, 7). Let's call the thesis which would correspond to the content of that intuition being veridical the *unity thesis*.

On the privacy intuition, Siddharth claims that subjects seem to have private experiences such that they cannot share token experiences. He says, "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 (…) two subjects cannot experience the same token experiential content" (Siddharth 2021, 7-8). As above, let's call the thesis

which would correspond to the content of that intuition being veridical the *privacy thesis*.

As you can see, the two theses are in direct conflict with phenomenal bonding panpsychism. Phenomenal bonding requires subjects to have proper parts, particularly proper parts which are other conscious subjects. The unity thesis rules this out: subjects do not have proper parts because they are utterly indivisible mereological simples. Phenomenal bonding also requires subjects share their experiences with their proper parts, which the privacy thesis rules out: two subjects cannot experience the same token experiential content.

As two theses about the essential features of subjects that will subsequently go on to support the rest of Siddharth's argument against phenomenal bonding, the unity and privacy theses must be well justified. The arguments Siddharth presents against phenomenal bonding panpsychism will only be as strong as the justifications for these theses.

Unfortunately, no justifications for the unity thesis or privacy thesis are given. Siddharth claims that we have an *intuition* about them but in his paper gives no reason for us—bystanders to the debate and phenomenal bonding panpsychists alike—to think that the intuitions are veridical.[2] He gives no reason to think that these intuitions in any way track the reality of what subjects—fundamental and non-fundamental—are like.[3]

With the lack of justification for these two essential theses noted, let's move on to look at the two premises of his argument: the claim that Goff's phenomenal bonding view is question begging without Miller's proposal, and the claim that Miller's proposal doesn't work. As I suggested, both of these claims are false or unjustified.

Justifying (1) – Without a Positive Concept, Phenomenal Bonding is Question Begging

How does Siddharth justify premise (1):

[2] Importantly, later in Siddharth's argument the two theses gain the status of *a priori truths*, equivalent in nature to truths in Euclidian geometry.

[3] Moreover, assuming the truth of these theses is not an innocuous assumption to make; both are very substantial claims. For instance, granting the unity thesis alone would rule out any view on which subjects are composite entities. We could not claim we are animals, brains, bundles of perceptions, unions of a mind and body, streams of consciousness across time, or sums of person-stages. Why give so much weight to an intuition that hasn't been fully justified?

> (1) If phenomenal bonding panpsychism is to be a non-question begging solution to the subject summing argument, then it requires a positive not-merely-role-playing conception of the phenomenal bonding relation.

He claims the following:

> The thrust of the subject summing argument is (…) that the notion of a composite subject itself seems incoherent, and thus impossible, on account of the ontological unity and privacy of subjects. 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. (Siddharth 2021, 9-10)

As we can see, the justification for (1) is that the subject summing argument is about the impossibility of composite subjects, which means it's about the impossibility of phenomenal bonding relations, so the phenomenal bonding response is question begging.

Let's call this the *Impossibility of Composite Subjects Argument* and formalise it in the following way:

> *Impossibility of Composite Subjects Argument*
>
> 1) The subject summing argument is about the impossibility of composite subjects.
> 2) If the subject summing argument is about the impossibility of composite subjects, then it is also about the impossibility of the phenomenal bonding relation.
> 3) Hence, the subject summing argument is about the impossibility of the phenomenal bonding relation.
> 4) If the phenomenal bonding relation is defined in a role-playing way and we haven't shown the relation is possible, or provided a positive concept of it, then as a response to the subject summing argument it is question begging.
> 5) Phenomenal bonding is defined in a role-playing way and we haven't shown the relation is possible.

6) Hence, without a positive conception of the phenomenal bonding relation phenomenal bonding is question begging as a response to the subject summing argument.

Unfortunately, this argument is unsound: both (1) and (4) are false or unjustified. Let's take them in that order.

Firstly, on (1). The subject summing argument is not about the incoherence of composite subjects. The subject summing argument is about the lack of a transparent, a priori explanatory relationship between the fundamental level conscious facts, and the non-fundamental conscious facts. In other words, it is about an explanatory gap between the fundamental conscious facts and the non-fundamental conscious facts. That the subject summing argument is about an explanatory gap is widely recognised within the literature (Chalmers 2016; Goff 2009b). Siddharth references the two popular formulations of this argument, one by Chalmers and the other by Goff, both of which are clearly stated in these terms. Goff claims that the mere existence of n subjects does not necessitate the existence of a further n+1 subject precisely because we can conceive of the group of n subjects existing without the further subject. Likewise, Chalmers claims that because we can conceive of *n* subjects existing in the absence of any subjects, the existence of those subjects cannot necessitate the existence of a further subject.

Premise (1) is therefore false. Nevertheless, there are arguments within the literature that purport to show that composite subjectivity is incoherent. These mereological problems include:

- The incompatible contexts argument (Basile 2008; 2010; James 1912; Roelofs 2016).
- The privacy argument (Roelofs 2019, 57).
- Boundary argument (Miller 2018b; Roelofs 2019, 59; Rosenberg 2004, chap. 4).
- The exclusivity of perspectives argument (Coleman 2012; 2014)
- Overwhelming experience argument (Albahari 2019).
- Maximality of consciousness argument (Roelofs 2019).

All these arguments have the form of identifying a feature, F, of subjects or their experiences, and showing why that feature precludes the possibility of subjects and their experiences being composite. For example, Roelofs' privacy argument identifies *privacy* as that feature, and shows that subjects

with private experiences can't be composites. Likewise, Coleman's perspectives argument identifies *having a perspective* as that feature, because perspectives, for Coleman, exclude other perspectives, which precludes perspectives being proper parts of other perspectives.

If one wants to equate the mereological arguments with the subject summing argument, then that is a claim that would need substantial justification. Why, in other words, should we think the subject summing argument is in fact the unity or privacy arguments? Why should we think it is the boundary argument? Siddarth gives no justification for this claim.

Moreover, there are responses to the mereological arguments within the literature, responses presented by phenomenal bonding panpsychists. For instance, Roelofs (2019) responds extensively to the mereological problems above. Likewise, Miller (2018a), and Roelofs and Goff (forthcoming) also address several of the mereological problems above.

Other than the first premise of the argument being false, why does this matter?

Because, if as Siddharth claims, the conclusion of the subject summing argument was "PB relations are impossible", then it would be argumentatively weak—to the point of begging the question—to respond by merely asserting, "PB relations are possible". But, given the impossibility of PB relations is *not* the conclusion of the subject summing argument, it isn't a weak response. The conclusion of that argument is that there is an explanatory gap; the proposal of a PB relation to bridge that gap is not a weak response to *that* argument.

Moreover, for Siddharth to construe the phenomenal bonding panpsychist as proposing a solution to the incoherence of composite subjects by employing the PB relations response, is to strawman the phenomenal bonding response and ignore the existing literature that does offer appropriate responses to the mereological problems.

Goff's initial phenomenal bonding proposal was a proposal to bridge the ostensible gap between micro-conscious facts and human-level facts about consciousness. The proposal was not intended to be a solution to the potential problem of composites, but a solution to the problem of the explanatory gap. Because phenomenal bonding was not *intended* to be a response to the mereological problems, and because phenomenal bonding panpsychists do not respond to the mereological problems with the phenomenal bonding proposal, to construe the view as a response to those arguments is to strawman it. Moreover, it is to strawman the view in such

a way that it *makes* the response question begging: it is being framed as a response to a problem that if it were a response to, then it would be question begging.

Again, why does this matter? Because it shows us where the burden of proof lies within the dialectic and the literature. Phenomenal bonding panpsychists respond to the subject summing argument by proposing the PB relation, they claim this closes the explanatory gap between the fundamental and non-fundamental conscious facts and bundles micro-subjects into large composite subjects. Presented with the incoherence of composite subjectivity objections of the sort mentioned above, they give (and should give) alternative responses.[4]

Lastly, Siddharth or their defender might respond by saying that it doesn't matter if the mereological problems are in fact the same problem as the subject summing problem, and it doesn't matter if phenomenal bonding has been presented as a response to those problems (making a strawman of it in the process). What matters is that those mereological problems rule out PB relations, and so the phenomenal bonding panpsychist needs to show that such relations are possible. In other words, the mereological problems show that composite subjects are incoherent, and phenomenal bonding panpsychists need to show that composite subjects aren't incoherent.[5]

The response to this point, however, is twofold. Firstly, to here show that Siddharth's argument is unsound, it is sufficient to show that (1) (and (4)) are unjustified or false. That I have already done, and the burden of proof now lies with Siddharth. Second, as already mentioned above, there are responses to the mereological problems within the literature, responses that phenomenal bonding panpsychists take to be sound. It is not necessary, therefore, to here respond to the mereological arguments. It's simply enough to note that phenomenal bonding panpsychists do try to show that composite subjects are possible.

Second, on (4). This is false for the same reason that (1) is; namely, that the subject summing argument is not about the incoherence and impossibility composite subjects. Yet, it remains unjustified regardless of that point. To defend (4), Siddharth claims the following:

[4] This is precisely the approach taken in Roelofs (2019), Miller (2018a), and Roelofs and Goff (forthcoming).
[5] I'd like to thank an anonymous referee for suggesting and clearly articulating this point.

> one can adopt such a method of defining relations in a brute manner to defend almost any unintelligible relation (…) 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. (Siddharth 2021, 10)

But this justification fails. Siddharth is not in the position to legitimately claim that the PB relation is an unintelligible one. He has not justified why, aside from claiming we have two intuitions about subjects, such a relation is unintelligible. As we noted in the previous section, why should these two intuitions be given so much weight? Why should we think their content is in any way veridical?

Justifying (2) – We can't form a positive concept of phenomenal bonding

What about for premise (2):

> (2) We cannot form a positive not-merely-role-playing conception of the phenomenal bonding relation.

To justify premise (2), Siddharth argues against the proposal made by Miller (2017): co-consciousness is the PB relation and we can form a positive not-merely-role-playing concept of it by, at the least, a process called *analogical extension*. Analogical extension is a process whereby you take a concept that was formed in one context (a prototype context) and extend it to a new, different context (the target context). Miller's proposal is that *co-consciousness* can be analogically extended from the prototype context in which it holds between one subject's experiences, to the target context in which it holds between two or more distinct subjects' experiences. In other words, we extend the concept of co-consciousness from being an *intra-subjective* relation to an *inter-subjective* relation.[6]

Miller's claim is that such a process allows us to see that co-consciousness can satisfy the three necessary and jointly sufficient conditions on a prospective PB relation (see Miller 2017, 542, 546):

- It must be a phenomenal relation.
- It must hold between subjects *qua* subjects of experience.
- It must necessitate further distinct subjects.

[6] See Roelofs (2016) for a defence of the notion of between subject phenomenal unity, i.e., of co-consciousness as an inter-subjective relation.

Because of this, the panpsychist no longer needs to be a mysertian about the PB relation (contra Goff (2016, 292-294)). Instead, they have a positive not-merely-role-playing conception (and Siddharth's argument would be unsound).

Let's call this the *anti-positive concept* argument and formulate it in the following way:

Anti-positive Concept Argument

1) Miller's analogical extension fails to give us a positive not-merely-role-playing conception of the phenomenal bonding relation.
2) Hence, we cannot form a positive not-merely-role-playing conception of the phenomenal bonding relation.

As before, I do not think this argument works.

Firstly, the argument is invalid. For (2) to follow from (1), i.e., to show that we cannot form a positive not-merely-role-playing conception of the phenomenal bonding relation, Siddharth would have to do more than simply show that analogical extension fails. He would have to show that analogical extension fails and that *it is the only method we have of forming a positive concept of the PB relation*. In other words, he'd have to rule out other potential alternatives or show there cannot be an alternative.

Are there alternatives? Yes. Introspection is an alternative.[7]

If we think that phenomenal bonding is true, then introspection is sufficient to form a concept of inter-subjective co-consciousness (Miller 2017, n. 21). Non-fundamental subjects, like humans and non-human animals, are composites with large proper parts that are also subjects. These proper parts undergo a subset of the experiences of the whole. Because of this, when a human subject introspects, it is thereby introspecting inter-subjective relations, viz. the relations that hold between the subjects that compose it. Hence, if phenomenal bonding is true, then we can form a concept of inter

[7] It may also be possible to form a concept of inter-subjective co-consciousness empirically, i.e., by discovering that co-consciousness holds between two or more subjects. As Miller (2017, 554–55) suggests, this may be what we're finding in the data regarding split-brain patients and the sharing of phenomenal content by hemispheres.

subjective co-consciousness via introspecting the phenomenal unity that holds between the experiences of our proper parts.[8]

In support of (1), Siddarth gives two justifications:

1) That the analogical extension of the co-consciousness relation involves a change in the type of relata, and therefore fails to meet Miller's second condition of a positive phenomenal bonding relation.
2) That analogical extension allows us to form concepts of relations we know are *a priori* incoherent.

First, on reason (1). Siddharth argues that when it comes to the analogical extension of co-consciousness, the relata in the prototype and target contexts changes type. In the prototype context the relata are experiences, whereas in the target context the relata are subjects. Because of this, Miller fails to meet the second condition of a prospective PB relation and, "analogical extension cannot help us form a positive conception of co-consciousness between subjects" (Siddharth 2021, 12).

The problem, however, is that the second condition on a prospective PB relation is not that it's relata *must* be subjects. The claim is that for a relation to be the PB relation "it must *hold between* subjects *qua* subjects of experience" (Miller 2017, 542, 546, emphasis added), not that subjects must be the relata. Because of this, there is no shift in the relata, and the second condition on a prospective PB relation is still met.

Siddharth's helpful criticism allows us to make clear a distinction, between being *directly* and *indirectly* related. Let's say that a relation *directly relates* two entities when those entities are its relata. For example, two colours are directly related by the similarity relation; two events are directly related by temporal relations; two species of animal are related by predation.

Let's say that a relation *indirectly relates* two entities when those entities have aspects (in the broadest sense of that term) which are the relata of the

[8] Moreover, without considerations of phenomenal bonding panpsychism, similar reasoning can be given to support the idea that introspection is sufficient to form a concept of inter-subjective co-consciousness. Subjects like animals are composite objects, and it seems like large proper parts of those animals are perfectly capable of being conscious. A human brain minus a few pieces of grey matter has everything required to be conscious, even if it's a proper part of a larger conscious being. Likewise, a dog missing one of its legs has everything required to be conscious even though it's a proper part of a larger conscious being. If so, then its highly plausible that just as there are many overlapping objects where my brain is roughly located, there are many overlapping conscious subjects. If so, then the unity between their experiences is an instance of inter-subjective co-consciousness.

relation, or are themselves an aspect of the relata of the relation. For example, two tins of paint are related by the similarity relation by having an aspect, i.e., their colours, which are the relata of the relation; two material objects are temporally related by being aspect of events, e.g., the sinking of the titanic and the building of the QE2, that are unfolding across time; two objects can be causally related by virtue of their properties, e.g., mass and velocity, being the relata of the relation.

Miller's condition should therefore be seen as the requirement that phenomenal bonding must, at the least, *indirectly* relate subjects. It must hold between subjects, but not necessarily have subjects as the relata. Inter-subjective co-consciousness does exactly that: it relates two or more subjects by, at the least, relating their experiences.

Second, on reason (2). Siddharth argues by *reductio* that analogical extension, at least in this case, isn't a good method for forming a positive concept of the phenomenal bonding relation. Following Miller (2017, 554) he claims that analogical extension is only permissible if it doesn't lead to *a priori* incoherence. If it leads to *a priori* incoherent notions, then it's obviously an unacceptable process of concept formation. So, he claims, "given [the] ontological unity and privacy of subjects, positing co-consciousness relation between two subjects leads to contradictions (…) privacy and ontological unity of subjects would be false' (Siddharth 2021, 13).

But (as may be expected by now) this should not be convincing, neither for the phenomenal bonding panpsychist nor the bystander. The phenomenal bonding panpsychist already rejects the unity and privacy theses, so they shouldn't find it convincing at all. Siddharth's argument relies on the truth of two theses to generate a *reductio* of his interlocutor's position, but those theses are already taken to be false by his interlocutor.[9] Moreover, as we noted earlier, Siddharth gives no justification for thinking the unity and privacy intuitions correspond to any true metaphysical theses, and the argument can only be as strong as those justifications.

Conclusion

The *begging the question argument* is unsound: both of its premises are false or unjustified. Because of this, we should still consider phenomenal bonding panpsychism to be a promising form of constitutive panpsychism.

[9] In order for the argument to be forceful, there must be good reasons to think the false theses must in fact be true.

There is room still for dissent and Siddharth's case against phenomenal bonding can be supported and further elaborated. However, that support must come in the form of justifying the key assumptions about the mereological simplicity of subjects and the privacy of their experiences, and developing those theses into arguments against constitutive forms of panpsychism. If strong defences of these theses can be given, then the phenomenal bonding panpsychist should be worried.

Acknowledgments

Thank you to the anonymous referees at EuJAP for taking the time to read and critique this paper. The quality of the paper would have suffered without their feedback. Thank you also to S Siddharth for engaging with phenomenal bonding panpsychism.

REFERENCES

Albahari, Miri. 2019. "Beyond Cosmopsychism and the Great I Am: How the World Might Be Grounded in Universal 'Advaitic' Consciousness." In *The Routledge Handbook of Panpsychism*, edited by William Seager, 119-130. New York: Routledge.

Basile, Pierfrancesco. 2008. "Is Mental Composition Impossible in Principle?" *Chromatikon: Annales de La Philosophie En Procès/Yearbook of Philosophy in Process* 4: 21–25.

———. 2010. "It Must Be True – But How Can It Be? Some Remarks on Panpsychism and Mental Composition." *Royal Institute of Philosophy Supplement* 67 (October): 93–112. https://doi.org/10.1017/S1358246110000044.

Chalmers, David J. 2016. "The Combination Problem for Panpsychism." In *Panpsychism: Contemporary Perspectives*, edited by Godehard Bruntrup and Ludwig Jaskolla, 179–214. 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 (1): 19–44. https://doi.org/10.1007/s10670-013-9431-x.

Dainton, Barry. 2000. *Stream of Consciousness: Unity and Continuity in Conscious Experience*. International Library of Philosophy. London and New York: Routledge.

Goff, Philip. 2009a. "Can the Panpsychist Get around the Combination Problem?" In *Mind That Abides: Panpsychism in the New Millennium.*, edited by David Skrbina and David (Ed) Skrbina, 129–35. Amsterdam, Netherlands: John Benjamins Publishing Company.

———. 2009b. "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*, by Godehard Bruntrup and Ludwig Jaskolla. Oxford: Oxford University Press.

James, William. 1912. *A Pluralistic Universe*. Hibbert Lectures. London: London, 1912.

Lockwood, Michael. 1989. *Mind, Brain, and the Quantum*. Oxford University Press.

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.

———. 2018a. *The Combination Problem for Panpsychism: A Constitutive Russellian Solution - An Investigation into Phenomenal Bonding Panpsychism and Composite Subjects of Experience*. Liverpool: University of Liverpool, PhD thesis.

———. 2018b. "Can Subjects Be Proper Parts of Subjects? The De-Combination Problem." *Ratio* 31 (2): 137–54. https://doi.org/10.1111/rati.12166.

Roelofs, Luke. 2016. "The Unity of Consciousness, within Subjects and between Subjects." *Philosophical Studies*, March, 1–23. https://doi.org/10.1007/s11098-016-0658-7.

———. 2019. *Combining Minds: How to Think about Composite Subjectivity*. Philosophy of Mind Series. New York: Oxford University Press.

Roelofs, Luke, and Philip Goff. forthcoming. "In Defence of Phenomenal Sharing." In *The Phenomenology of Self-Awareness and Conscious Subjects*, edited by Martine Nida-Rümelin, Donnchadh O'Conaill, and Julien Bugnon. Routledge.

Rosenberg, Gregg. 2004. *A Place for Consciousness: Probing the Deep Structure of the Natural World*. Philosophy of Mind Series. Oxford and New York: Oxford University Press.

Siddharth, S. 2021. "Against Phenomenal Bonding." *European Journal of Analytic Philosophy* 17 (1): (D1)5-12. https://doi.org/10.31820/ejap.17.1.3

BOOK REVIEW

Toril Moi, Rita Felski, and Amanda Anderson
CHARACTER: THREE INQUIRIES IN LITERARY STUDIES
University of Chicago Press, 2019
ISBN-13: 978-0-226-65866 (paper)
ISBN-13: 978-0-226-65883-4 (e-book)
Hardcover, $60.00
Paperback, $20.00
e-book, $14.01

David Grčki[1]

[1] University of Rijeka, Faculty of Humanities and Social Sciences

In their book *Character: Three Inquiries in Literary Studies*, Toril Moi, Rita Felski and Amanda Anderson explore new accounts of how we engage with and understand fictional characters. The book consists of an introduction and three separate essays; *Rethinking characters* (Moi), *Identifying with Characters* (Felski), and *Thinking with characters* (Anderson). In the introduction, they state that the goal of the book is not to build a normative theory of engagement with fictional characters but to "clear away some old restriction" and "open up new avenues of inquiry" (22). They argue that modernist-formalist theories that treat fictional characters as "character spaces" and "patterns in the text"—which have dominated literary studies for more than half a century—face serious problems. There is a value, they say, in understanding how laymen engage with fictional characters—by connecting with them on a personal and emotional level. Each author argues for these two points in a unique way.

In the first essay, Moi investigates "the origins of the taboo" of treating characters as real people, tracing it back to L. C. Knights's 1933 essay "How many children had Lady Macbeth?". Influenced by Russian and Czech formalists, L. C. Knights argues that fictional characters are "not our friends for life" and are not "as real as our familiar friends" (28). According to L. C. Knights, fictional characters "exist only as words on a printed page", "they have no consciousness" and "feeling that they are living people is an illusion" (28) and we should treat them as "patterns in the text" and abstractions. He attacks the idea that Shakespeare was the

Correspondence: dgrcki@gmail.com

creator of characters (which was influential at the time) and argues that Shakespeare was primarily a poet and that we should focus our analysis on Shakespeare's "rhythm and imagery" and "quality of the verse" (36). Moi argues that L. C. Knights has "professional and aesthetic agenda" (30, 40)––the defence of modernism. His agenda is to criticize "traditionalist, feminized, middlebrow sentimentality" and to replace it with "cool, modernist, impersonality" (43). Moi claims that, L. C. Knights's position fails as a coherent philosophical argument because there is no reason why we can't have character analysis and formal analysis when discussing a narrative work of art. The problem, according to Moi, is not just that L. C. Knights is wrong, but that his ideas are now dogma in literary criticism. Approach to character analysis inspired by L. C. Knights's work (modernist-formalist approach) is now the norm and treating characters as real people is something that only uneducated laymen do.

In the next section of her essay, Moi analyses John Frow's *Character and Person* (2014) and takes it as evidence that Knights's dogma is alive even today. Frow argues that a fictional character is a "complex concept of form" (49) and that we should analyse it as having two distinct ontological senses: "pieces of writing or imagining and person like entities" (51). He argues that there is something miraculous about engaging with pieces of writing or imagining as person like entities. Moi argues that there is nothing miraculous at play here. We often talk about fictive characters in much the same way we talk about real people, and we emotionally engage with them in a similar manner. According to Moi, Frow has created a pseudoproblem which stems from his post-Saussurean theoretical framework which is neither helpful nor useful for understanding fictional characters. Moi concludes by reiterating the dangers of succumbing to modernist-formalist framework of character engagement.

In the second essay, Felski explores the notion of identifying with fictional characters, stating that identification is "a sense of affinity or shared response" with a fictional character which can be "ironic as well as sentimental, ethical as well as emotional" and is "a default rather than an option; a feature, not a bug" (77). Like Moi, Felski frowns upon the idea that identification is something that only uneducated readers and viewers do. Identification, according to Felski, is a useful tool in explaining how we engage with fictional characters, and it offers us a way to account for our engagement with a wide array of fictional characters. From Emma Bovary with whom we identify based on an affinity and commonalities we share with her to characters like hobbits and rabbits with whom we identify more metaphorically than literally. Inspired by *Murray Smith* (1995), Felski argues that there are four strands of identification: alignment, allegiance, recognition, and empathy. Alignment is reader's or viewer's

access to the character—the character point of view (or perspective) that the audience gets from a particular character in a narrative. Allegiance explains in which way audience identifies with a fictional character through shared ethical and political values. Recognition, as the most basic form of identification, amounts to audience's visual perception of characters as human figures through perceptual cues—body, face and voice. The last part of identification is empathy—"sharing someone's feelings and responding with concern to those feelings" (105).

Felski offers two more ideas: "character as *Umwelt*" and "ironic identification". According to Felski, one of the reasons why the modernist-formalist approach to understanding fictional characters fails is because fictional characters are not constrained by the work they originated from. Fictional characters move from literature to movie and from movie to graphic novel via adaptation. They live in various fan fictions, they survive by "teleporting into new media" (88) through accessories, trademarks, and sayings. Felski calls this possibility of character transformation "character as *Umwelt*". Ironic identification is Felski's solution to the problem of antiheroes. The problem of antiheroes states that there is an explanatory gap in how we understand and engage with fictional characters that we do not empathize with, sympathize with or share ethical or moral values with. Felski uses Camus's Meursault and Dostoevsky's Underground Man as examples of antiheroes. She argues that we identify with those characters through irony. She states that, contrary to standard belief, irony and identification are not mutually exclusive. Ironic identification enables us to share a sense of estrangement and disassociation with those characters. According to Felski, "what is common is an experience of having nothing in common with others, of feeling at odds with the mainstream of social life" (113). Felski concludes by saying that there are various kinds of engagements with fictions and that identification with characters is the driving force behind them.

In the third essay, Anderson explores the importance of character's moral experience, moral thinking, and inner life. As Moi and Felski, Anderson argues that modernist-formalist approach to character engagement is deeply flawed and points to the works of *John Frow* (2014) and *Alex Woloch* (2003) as contemporary examples. The reason for this is because they fail to acknowledge the way in which "novels uniquely present forms of moral experience" and "interior moral reflection as it is extended across time" (130, 131). To make her point she uses the concept of "rumination"––a term borrowed from cognitive-behavioural psychology which describes person's obsessive and circular thinking, typically seen as pathological. It is a state of distress usually associated with worry, anxiety, and similar negative emotional states. Anderson argues that rumination can be seen in

a positive light, as a form of moral thinking and moral justification. In this sense, rumination is more akin to contemplation or reflection, but is not reducible to them. She acknowledges that rumination, in its productive form, "cannot be easily captured" (136). According to Anderson, we can learn a lot about moral clarification, moral epiphany and how we make moral decisions by understanding "extended periods of anguished reflection and elusive processes of grief and healing" (138). In order to do so, in literary context, we need to give centrality to the presentation of character. Anderson uses Eliot's reference to rumination in *Middlemarch* and analysis of rumination in Trollope's *The Last Chronicle of Barset* and Woolf's *Mrs. Dalloway* to make her point. Anderson concludes by saying that "analysis of rumination will allow us to better account for both moral dimension of fiction and the forms of thinking that characterize moral and political life more broadly" (166). To do this successfully, we need to give more attention to the analysis of fictional characters.

To conclude, this is a passionate and highly engaging book. Each author offers her unique perspective on the issue of understanding and engaging with fictional characters. Although their perspectives are different, their argumentations do not collide but supplement each other. Their accounts are relevant not only for literary studies but also for the philosophy of literature and aesthetics. This book will be interesting to experts investigating how we engage with fictional characters as well as to novices who are just curious about it.

Acknowledgments

This work has been supported in part by Croatian Science Foundation under the project number UIP-2020-02-1309.

BOOK REVIEW

Kengo Miyazono and Lisa Bortolotti
PHILOSOPHY OF PSYCHOLOGY: AN INTRODUCTION
Cambridge, Polity Press, 2021
ISBN-13: 978-1509515486 (paper)
ISBN-13: 509515488 (e-book)
Hardcover, $ 40.89
Paperback, $20.00
e-book, $24.95

David Grčki[1]

[1] University of Rijeka, Faculty of Humanities and Social Sciences

In their book, Philosophy of Psychology: An Introduction, Kengo Miyazono and Lisa Bortolotti offer an accessible overview of the various fields that are relevant to philosophy and psychology, from rationality and self-knowledge to autism and delusions. In the introduction authors state that "providing a clear definition of philosophy of psychology is challenging and perhaps not very helpful" (5) and add that in their view word 'psychology' is "somewhat similar to cognitive science" (5). Unlike other introductory books in philosophy of psychology that deal with theoretical foundations of scientific psychology and cognitive science, the distinguishing feature of the present book is its implicationst perspective. Myazono and Bortolotti are interested in providing an accessible discussion regarding the philosophical implications of various empirical studies for our conceptions as rational and self-knowing creatures. The underlying motif that cuts across the chapters of the book is that human cognition and agency are imperfect, i.e. they fail to meet certain ideal normative standards.

The book is meant to be used as a introduction to philosophy of psychology. The structure of its chapters fittingly reflects this aim. Each chapter begins with simple and clear definitions of main concepts and ends with questions about the content of the chapter and additional material in a form of articles, books, and online resources. The book contains eight chapters (excluding *Introduction* and *Conclusion*): *Rationality*, *Self-Knowledge*, *Duality*, *Moral Judgment*, *Moral Motivation and Behaviour*,

Correspondence: dgrcki@gmail.com

Free Will and Responsibility, *Delusions and Confabulation* and *Autism and Psychopathy*.

In the first chapter *Rationality*, Miyazono and Bortolotti are attempting to provide an answer to a difficult question; are human beings rational? They admit that "rationality means different things in different contexts (18) and focus on a specific form of rationality that equates to "reason in accordance with the principles of reasoning that are based on rules of logic, probability theory and so forth." (19). After discussing the most prominent research (The Wason selection task, The conjunction fallacy and others) which shows how people systemically fail to reason according to the rules of logic and probability theory, they arrive at the pessimistic conclusion that humans may very well be less than rational.

In the second chapter *Self-Knowledge*, authors are asking; how much can we know about ourselves? Miyazono and Bortolotti define self-knowledge as "the sophisticated capacity to consider one's own beliefs and reflect on one's own choices (in deliberation and self-examination or in interpersonal settings such as discussion with others)" (46). The authors address array of concepts related to self-knowledge (most notably introspection and privilege access) and analyse various empirical studies to see how reliable self-knowledge is. They advocate for the moderate view in which self-knowledge is reliable in some cases, depending on a part of the mind the self-knowledge is associated with.

Building on the previous chapter, in chapter *Duality*, the authors argue that the mind is fragmented and the way to explain how is The Dual process theory. The Dual process theory states that the mind is fragmented in two distinct processing modes of reasoning – fast, automatic and non-conscious (type 1) and slow, controlled and conscious (type 2). At the end of the chapter, they conclude that while having certain problems, The Dual process theory is plausible.

The next topic Miyazono and Bortolotti tackle is moral judgment. The main aim of this chapter is to make sense of people's moral judgment by using the framework of The Dual process theory. They argue that "moral judgments are determined by the interaction between two kinds of processes: an emotion based process and reasoning based process" (105, 106), which corresponded to type 1 and type 2 processes. Although, there are several issues that arise from this (the interaction question and the processing question), authors maintain that this is a correct and fruitful way to discuss the nature of moral judgments.

In their next chapter, *Moral Motivation and Behaviour*, Miyazono and Bortolotti discuss the role of moral emotions and affective empathy as well as their relation to moral motivation and behaviour. The authors' approach is to examine two hypotheses: the empathy-altruism hypothesis and the empathy benefit hypothesis. The first states that our empathy induced behaviour is genuinely motivated by altruism and the second states that it is motivated by morally beneficial consequences. Relying on studies by Batson (2001, 2011, 2018), Miyazono and Bortolotti argue that the former hypothesis has more merits than the latter.

In the chapter *Free Will and Responsibility* Miyazono and Bortolotti are trying to give answers to two questions; are humans free agents?, and how much freedom do they have? Authors start by examining the best scientific arguments against free will (various arguments from epiphenomenalism) and argue that the best available psychological and neuroscientific evidence does not support epiphenomenalism. They, again, take the middle road by claiming that humans do not completely lack the free will but "are not as free as we might think" (165).

The last two chapters in the book—*Delusions and Confabulation* and *Autism and Psychopathy* are thematically unique. The former focuses on unusual behaviours that can be a product of a pathological or neurodivergent condition and the latter focuses on two neurodivergent conditions—autism and psychopathy. Miyazono and Bortolotti argue that these topics are important because they "(…) provide us with useful data for thinking about mind and cognition in general" (192).

In *Delusions and Confabulation*, the authors state that these are psychological and behavioural phenomena which are characterized by people misrepresenting reality. They admit that both concepts are hard to define. Delusions are defined by formation process; reasoning abnormally and perceiving abnormally. Confabulations are defined as symptoms "(…) of disorders involving severe memory impairments, such as dementia and amnesia" (214). Miyazono and Bortolotti investigate whether these sort of distortions of reality are distinct from other forms of irrationality. After considering various empirical research and arguments they conclude that they do not differ from other forms of irrationality and that their pathology cannot be accounted for by relying on rationality alone.

In the chapter *Autism and Psychopathy*, Miyazono and Bortolotti explore the nature of these neurodivergent conditions. Autism spectrum disorder refers to a spectrum of conditions that are characterised by "persistent deficits in social communication and social interaction across multiple contexts, as manifested by deficits in social-emotional reciprocity, non-

verbal communication and developing, maintaining and understanding social relationships" (229). Psychopathy refers to "(…) a condition in which a person demonstrates remarkable tendency to engage in anti-social behaviour, which is often diagnosed according to the Psychopathy Checklist-Revised (PCL-R) by Robert Hare" (241). The authors claim that autism is usually defined by mindreading deficits (mindreading is an ability to attribute mental states to other people) and psychopathy by empathy deficits. They argue that autism and psychopathy are complex phenomena and that while mindreading and empathy are important for understanding both phenomena, they are not essential.

In the conclusion, I would like to say that this is precisely written and well-structured book. The book covers various topics, but each topic serves a unique and relevant purpose for understanding the philosophy of psychology. The structure of the book offers a natural progression from highly abstract topics, such as, rationality, to very specific topics, such as autism and psychopathy. I would like to recommend this book to anyone who is interested in implications of psychological/empirical studies for traditional philosophical topics regarding what makes us human.

BOOK REVIEW

José Luis Bermúdez
FRAME IT AGAIN: NEW TOOLS FOR RATIONAL DECISION-MAKING,

Cambridge University Press, 2020, pp. x + 330
ISBN-13: 978-1107192935
ISBN-10: 1107192935
Hardcover, $ 18.22
e-book, $15.49

Viktor Ivanković[1]

[1] Institute of Philosophy, Zagreb

A susceptibility to framing effects, i.e., arriving at different decisions pertaining to how a choice of the same content is presented, has standardly been rendered a mark of the irrational. In "Frame It Again: New Tools for Rational Decision-Making", José Luis Bermúdez offers convincing arguments that being sensitive to frames in this way is often not only compatible with rationality, but may be a requirement for it.

Too much attention, Bermúdez believes, has been paid to the "easy cases" (2). For instance, basketball players are rated more highly in the positive frame ('shots *made*') than in the negative frame ('shots *missed*'), despite this arising from the same datum (23). In the same way, individuals are more likely to buy beef presented as '25 percent *lean*' than '75 percent *fat*', despite leanness being the flipside of fatness (24). Or, in the famous Asian disease experiment, subjects are shown to respond differently to the same dilemma about survival estimates when it is presented in terms of 'lives *lost*' as opposed to 'lives *saved*' (20-21). The increased focus on these cases and many like them in psychological studies, described by Bermúdez in great detail and within various contexts in the first three chapters of the book, have painted the picture of defective human reasoning. It has amounted to "a litany of human irrationality", a narrative within the various social sciences "that human reasoning is fundamentally flawed" (10).

But the litanist, at least substantively, is not Bermúdez's opponent, as she may appear to be at first. Rather, Bermúdez seems to ground his case for the compatibility between the susceptibility to framing effects and rationality on an extended conception of framing effects. His most reoccurring example in the book, breaking away from the seemingly

Correspondence: viktor.ivankovic@gmail.com

narrower conception of relevant framing effects, is that of the Greek king Agamemnon, who frames the possible outcome of bringing about the death of his daughter Iphigenia in two different ways—*Murdering his Daughter* and *Following Artemis's Will*, whereas refusing to bring about her death is presented within a single frame—*Failing his Ships and People*. Bermúdez presents Agamemnon as simultaneously preferring to *Follow Artemis's Will* over *Failing his Ships and People* and preferring to *Fail his Ships and People* over *Murdering his Daughter*, while knowing full well that *Following Artemis's Will* and *Murdering his Daughter* are "the same outcome, differently framed" (7-8).

Bermúdez later seems to agree with the litanists that as far as their observed cases are concerned (those within an 'intensional context', as Bermúdez calls it (99)), individuals often demonstrate irrationality. If it is pointed out to individuals that '25 percent lean' is the same as '75 percent fat', or that survival estimates are in fact equal in the 'lives lost'/'lives saved' frames in the Asian disease case, and they fail to adjust their preferences, then they exhibit irrationality (98). Conversely, in the Agamemnon case and those similar to it (within an 'ultra-intensional context' (99)), Bermúdez argues that it is not irrational to stick to the same valuations even after finding out that the different frames are representative of the same outcome. It might be difficult to shake off the sense that cases within an intensional context, as opposed to those within an ultra-intensional one, are simply of a different kind. After all, Bermúdez gives up rather easily on even attempting to offer 'a single way of thinking what a frame is' (11), given the different ways frames are 'discussed and deployed […] in psychology, economics, linguistics, sociology, political science, and philosophy' (12). What we are left with is the broadest and most inclusive possible understanding of frames as unavoidable aspects of communication, manifesting themselves as 'descriptions' and 'narratives' (12). Yet, despite this conceptual indeterminacy in his account, Bermúdez is still persuasive in that it makes sense to accommodate the seemingly different kinds within the same account. Perhaps this is because Bermúdez's point that frames highlight some reasons while downplaying others, as in the Agamemnon case, is a very commonsense way of thinking about frames, and one which is often not captured by the narrower conceptions of framing effects.

In Chapters 4 to 6, Bermúdez puts forward the gist of his argument. First, in Chapter 4, he explains that litanists are in the business of undermining, through descriptive accounts such as prospect theory (83-89), the predominant normative theory of rationality often known as 'rational choice theory' or 'Bayesian decision theory'. Litanists point out myriad ways in which actual humans fail to live up to the requirements of such a normative theory (67), particularly its requirement of internal consistency

(74). To litanists, susceptibility to framing effects seems incompatible with standard versions of rational choice theory. Having transitive preferences requires that when we prefer A to B and B to C, we also prefer A to C, lest our preferences become *cyclical* (i.e., placing us in an endless loop that seems irreconcilable with rationality) (79). The famous money-pump argument shows that acting on cyclical preferences opens us up to a series of transactions that may well leave us bankrupt; strictly preferring A to B, B to C, and C to A, and agreeing to swap them in a loop can slowly drain us dry (e.g., if A is 1 $, and C is later sold for 0.99 $) (see, e.g. Schick 1986). Breaking with transitivity in such a case seems hardly consistent with rationality.

What about Agamemnon's preferences to *Follow Artemis' Will* over *Failing his Ships and People*, and simultaneously to *Fail his Ships and People* over *Murdering his Daughter*? Bermúdez says there is nothing erroneous about framing the death of Iphigenia in the two aforementioned ways, and that these frames merely reflect "different ways of thinking about the same basic outcome" (81). This is why Agamemnon only has *quasi-cyclical* preferences, which are not an affront to rationality (82). However, another concern about conceptual indeterminacy emerges here. For the most part, Bermúdez's examples of quasi-cyclical preferences are those of agents in the grip of indecision, as in the case of Agamemnon. In Bermúdez's descriptions, these are not preferences of the kind that simply move agents to act. If they were, and were reversible in the way money-pump scenarios work, then perhaps we would be able to conceive cases where agents with quasi-cyclical preferences end up in similar loops.[1] This raises the question whether quasi-cyclical preferences in Bermúdez's examples are in fact quasi-preferences, that is, mere considerations or mere aspects of a complex decision-making situation.

Still, it is on quasi-cyclical preferences that Bermúdez builds his normative account of rationality in Chapter 5.[2] Susceptibility to framing can be perfectly rational, he argues, when frames bring about reasons not previously considered; in such cases, "frame-sensitivity can be both rational and in fact very useful" (94). In the ultra-intensional context, Bermúdez says, "*it can be* rational to have values and preferences that shift according to how things are framed" (100; emphasis in original). He shows

[1] In fact, Bermúdez uses the notion of 'strictly preferring' only once in the book, to explain the original money-pump argument. Quasi-cyclical preferences are never once presented as 'strict preferences'.

[2] Although, as Niker (2021) writes in an earlier review of this book, it is not exactly clear whether Bermúdez aims to extend rational choice theory or replace it with an altogether new theoretical paradigm.

this by way of examples, most notably the case of George Orwell in the Spanish Civil War, who first views enemy soldiers as 'fascists', but then also comes to frame them as 'fellow human beings' when he sees one fleeing half-naked with his trousers hanging (101-105). On Bermúdez's description, Orwell can hold onto both frames even after reaching the decision to shoot (or not to shoot). In such cases, the other frame manifests *ex post* as 'mixed feelings' (105) for not having done otherwise. Yet, this is perfectly compatible with rationality on Bermúdez's account. In fact, he later argues that this kind of quasi-cyclicality might not just be compatible with rationality, but an important requirement for it. Here, some concerns could be raised. It could have been the case that after very short consideration, one frame clearly wins out for Orwell and no residual emotions are produced after the fact; this hardly seems to offend rationality.[3] It is also uncertain how erratically switching back and forth between decisions in response to lingering frames bears on the rationality of an agent.

But Chapter 6 sheds more light on these matters by explaining how we may transition from the claim that framing effects are compatible with rationality to the claim that they may be a requirement for it. Two important factors about rationality are thereby introduced. First, Bermúdez argues that holding onto multiple frames is often the consequence of considering the outcomes of different courses of actions—of displaying *due diligence* before arriving at decisions (121). For agents to think about complex decisions is likely to produce different frames from which to view options and develop a sensitivity to them. If Agamemnon failed to frame the death of Iphigenia in one of the two deeply relevant ways earlier described, but instead operated from a single frame, he would be "missing something very important" about the "complexity of the decision situation" (117). Despite being consistent in a way that the original Agamemnon is not, his failure to assume more than a single perspective amounts to "a failure of rationality" (117).[4] Second, he shows that emotional engagement with a given decision can depend crucially on how that decision is framed, often bringing about rationality-compatible quasi-cyclical preferences (128). Additionally, Bermúdez also mentions that it is a failure of rationality to

[3] Niker (2021) makes a related point to this one.

[4] As in previous cases, this part of Bermúdez's argument might also be somewhat lacking in conceptual depth. Because he does not explain exactly the kind of property rationality is, it is not clear whether Agamemnon's is a failure of rationality, and not some other failure of reasoning, such as the failure of thoughtfulness. Bermúdez only states that "since this is a failure of thought in the service of practical reasoning, it qualifies […] as a failure of rationality" (122).

ground a frame on false beliefs or fail to reject it in the face of countervailing evidence (133).

Chapter 7 explains how the exercise of self-control depends crucially on how rewards are differently framed for agents. In fact, Bermúdez argues that the pull of temptation is difficult to account for on the standard framework of rational choice theory. Instead, we need to appeal to multiple frames and quasi-cyclical preferences to express how agents can prefer larger rewards later over smaller rewards sooner, while simultaneously being tempted by the latter at the expense of the former.

In Chapter 8, Bermúdez turns to rationality in strategic decision-making, namely to how our "actions are interdependent with respect to rationality" (165). In particular, he tests Michael Bacharach's idea that there is a different perspective to rationality when we adopt the "we"-frame, as opposed to the "I"-frame, the latter of which predominates game-theoretical thought experiments. Switching to the "we"-frame does not change any of the values in standard pay-off tables, but "different aspects and properties […] become salient" (176). If agents approach strategic decision-making as *team reasoners*, then they are more likely to arrive at Pareto-optimal outcomes. Agents who are *committed team reasoners* (and are confident that the other person is as well) "would each play their part in the Pareto-optimal strategy profile" (181). However, according to Bermúdez, Bacharach does not manage to prove the rationality (in a normative sense) of adopting the "we"-frame, but merely offers descriptive points for why adopting it is a common occurrence (such as best common interests and strong interdependence) (184-185). Nor is the adoption of the "we"-frame always desirable, as Chapter 9 shows, given all the dark campaigns in human history launched from the "we"-frame (194-195). But more importantly, the Chapter tackles the problem that the perspectives of the two frames cannot be rationally compared—that they are incommensurable, making it impossible to explain why it would be rational to abandon one frame for the other. The mistake of assuming incommensurability, according to Bermúdez, is caused by the loaded terminology of standard reasoning within the "I"-frame ('defection', 'cooperation', 'free-riding'), leaving the impression that "the "I"-frame/"we"-frame distinction [maps] onto the selfish/altruistic distinction" (204). Bermúdez argues that reasoning into the "we"-frame can be grounded in the prior valuing of fairness, although he does not explain why 'valuing fairness' is part of a normative theory of rationality, and not a descriptive claim like those offered by Bacharach, namely that it is a fact that people often value fairness.

In the final two chapters, Bermúdez sketches the profile of a model, frame-sensitive reasoner able to tackle discursive deadlocks. For instance, such a reasoner acknowledges the difference between framing effects within intensional contexts, which require an adjustment of preferences for rationality to be retained, and ultra-dimensional contexts, where it is rational to hold onto quasi-cyclical preferences (217-218). Such a reasoner is able to: 1.) detach from her own perspective and reflect on her mental states "as separate from oneself" (246-247), thereby reducing her affectivity and letting go of emotional baggage (248); 2.) simulate the entire experience of assuming alternative perspectives (254-255); 3.) operate flexibly within multiple frames at once (261); and 4.) understand how frames make competing reasons salient and which values underpin them (264-267). Bermúdez acknowledges that there will be limits to what a frame-sensitive reasoner can and should be expected to achieve, e.g., in the case of holding onto clearly repugnant frames, such as those of serial killers or child molesters (271). Nevertheless, it is not always clear on the picture of Bermúdez's model reasoner how close we must approximate the model to be considered rational thinkers.

Bermúdez's book is very important, not only for remedying some shortcomings of rational choice theory, but also for acting as the bridge between the various disciplines in the social sciences and humanities delving into the matter of rationality. In philosophy, it is a necessary read for rationality theorists as well as moral and political philosophers exploring the ethics of nudging. However, due to its ambition, it adopts a very wide conceptual framework, leaving several loose ends. I have already mentioned some ambiguity about the very concept of 'frames', as well as that of 'preferences'. Other ambiguities, some of which could easily be cleared up in future work, relate to how the crucial concepts in the book come together—'frames', 'framing effects', 'preferences', 'rationality', 'reasons', 'perspectives', 'narratives'. But these conceptual drawbacks do not eclipse the very convincing case for rationality-compatible framing effects on offer.

References

Niker, Fay. 2021. "*Frame It Again: New Tools for Rational Decision-Making*, José Luis Bermúdez. Cambridge University Press, 2020, x + 330 pages". *Economics and Philosophy*. https://doi.org/10.1017/S0266267121000316

Schick, Frederic. 1986. "Dutch bookies and money pumps". *The Journal of Philosophy* 83(2): 112–119.

ABSTRACTS (SAŽECI)

EXTREME PERMISSIVISM REVISITED

Tamaz Tokhadze
University of Sussex

ABSTRACT

Extreme Permissivism is the view that a body of evidence could rationally permit both the attitude of belief and disbelief towards a proposition. This paper puts forward a new argument against Extreme Permissivism, which improves on a similar style of argument due to Roger White (2005, 2014). White's argument is built around the principle that the support relation between evidence and a hypothesis is objective: so that if evidence E makes it rational for an agent to believe a hypothesis H, then E makes it rational to believe H, for all agents. In this paper, I construct a new argument against Extreme Permissivism that appeals to a logically weaker, less demanding view about evidential support, Relational Objectivity: whether a body of evidence E is more likely if H is true than if H is false is an objective matter and does not depend on how any agent interprets the relationship between E and H. Relational Objectivity is solely concerned with the conditional probabilities called likelihoods and does not put substantive constraints on an agent's prior and posterior credences. For this reason, the presented argument avoids the standard permissivist criticism levelled against White's argument.

Keywords: permissivism; uniqueness; Roger White; objectivity of evidential support; relational objectivity; epistemic standards; likelihoods

PONOVNO RAZMATRANJE EKSTREMNOG PERMISIVIZMA

Tamaz Tokhadze
University of Sussex

SAŽETAK

Ekstremni permisivizam je gledište prema kojemu skup dokaza racionalno dopušta stav vjerovanja i nevjerovanja prema propoziciji. Ovaj rad iznosi novi argument protiv ekstremnog permisivizma, koji poboljšava sličan stil argumentacije Rogera Whitea (2005, 2014). Whiteov argument je izgrađen oko principa da je odnos podrške između dokaza i hipoteze objektivan: ako dokaz E čini racionalnim da djelatnik vjeruje u hipotezu H, onda E čini

racionalnim vjerovanje H, za sve djelatnike. U ovom radu izgrađujem novi argument protiv ekstremnog permisivizma koji se poziva na logički slabije, manje zahtjevno gledište o dokaznoj potpori, koje se može nazvati Relacijska objektivnost: je li skup dokaza E vjerojatniji ako je H istinit nego ako je H neistinit je objektivna stvar i ne ovisi o tome kako bilo koji djelatnik tumači odnos između E i H. Relacijska objektivnost bavi se isključivo uvjetnim vjerojatnostima koje se nazivaju likelihudovi i ne postavlja suštinska ograničenja na djelatnikova prethodna i posteriorna uvjerenja. Iz tog razloga, predstavljeni argument izbjegava standardnu permisivističku kritiku upućenu Whiteovom argumentu.

Ključne riječi: permisivizam; jedinstvenost; Roger White; objektivnost dokazne potpore; relacijska objektivnost; epistemički standardi; vjerojatnosti

GNOSTIC DISAGREEMENT NORMS

Domingos Faria
University of Lisbon

ABSTRACT

Our main question in this paper is as follow: (Q) What are the epistemic norms governing our responses in the face of disagreement? In order to answer it, we begin with some clarification. First, following McHugh (2012), if we employ a useful distinction in normativity theory between evaluative and prescriptive norms, there are two readings of (Q)—we explore such distinction in section 2. And secondly, we accept gnosticism, that is, the account that the fundamental epistemic good is knowledge. It is with this assumption that we want to answer (Q). So, if gnosticism is true, what is the plausible answer to (Q)? In section 3 we argue for gnostic disagreement norms as response to (Q) and in section 4 we apply such norms to particular cases of disagreement.

Keywords: disagreement; evaluative norms; prescriptive norms; disagreement norms; gnostic disagreement norms; knowledge-first epistemology

GNOSTIČKE NORME NESLAGANJA

Domingos Faria
University of Lisbon

SAŽETAK

Glavno pitanje u ovom radu je sljedeće: (P) Koje epistemičke norme upravljaju našim odgovorima u slučaju neslaganja? Kako bismo odgovorili na njega, započinjemo s pojašnjenjem. Prvo, slijedeći McHugha (2012), pod pretpostavkom da koristimo razliku iz teorije normativnosti između evaluativnih i preskriptivnih normi, postoje dva čitanja (P)-a — takvu razliku istražujemo u odjeljku 2. I drugo, prihvaćamo gnosticizam, tj. gledište da je temeljno epistemičko dobro znanje. Ovom pretpostavkom želimo odgovoriti na (P). Dakle, ako je gnosticizam istinit, postavlja se pitanje koji je prihvatljiv odgovor na (P)? U odjeljku 3 zagovaramo norme gnostičkog neslaganja kao odgovor na (P), a u odjeljku 4 takve norme primjenjujemo na posebne slučajeve neslaganja.

Ključne riječi: neslaganje; evaluacijske norme; preskriptivne norme; norme neslaganja; norme gnostičkog neslaganja; epistemologija prvenstva znanja

DEBUNKING DOXASTIC TRANSPARENCY

Ema Sullivan-Bissett
University of Birmingham

ABSTRACT

In this paper I consider the project of offering an evolutionary debunking explanation for transparency in doxastic deliberation. I examine Nicole Dular and Nikki Fortier's (2021) attempt at such a project. I suggest that their account faces a dilemma. On the one horn, their explanation of transparency involves casting our mechanisms for belief formation as solely concerned with truth. I argue that this is explanatorily inadequate when we take a wider view of our belief formation practices. I show that Dular and Fortier overstate the extent to which adaptive non-evidentially supported beliefs are rare, and the implausibility of disjunctive evolutionary systems. They should allow a role for the non-truth directed behaviour of our mechanisms of belief formation. On the other hand, we might restrict the explanation offered by Dular and Fortier to the deliberative context, that is, we might understand them as allowing for non-evidential belief formation outside of the deliberative context, but as

identifying the key to explaining transparency in the truth-directed evolutionary mechanisms as they operate in the deliberative context. However, this would land them on the second horn of the dilemma: we would then have no different an explanation to one I have offered elsewhere (2018), an explanation which Dular and Fortier explicitly put aside as engaged in a project different from their own. I finish by briefly considering some broader implications relating to explaining transparency, the nature of belief, and the prospects for pragmatism. I conclude that Dular and Fortier's debunking explanation of transparency bestows an implausible role for truth in fixing our beliefs, or, if it doesn't, then we simply have the restatement of a view explicitly disavowed by the authors. We are left, then, with an explanation we ought not want, or an explanation we already had.

Keywords: belief; transparency; doxastic deliberation; evolutionary debunking; reasons for belief

OSPORAVANJE DOKSASTIČKE TRANSPARENTNOSTI

Ema Sullivan-Bissett
University of Birmingham

SAŽETAK

Ovaj rad se bavi osporavanjem objašnjenja za transparentnost u doksastičkom promišljanju na temelju evolucijskih razmatranja koje nude Nicole Dular i Nikki Fortier (2021). Tvrdim da se njihovo objašnjenje suočava s dilemom. S jedne strane, njihovo objašnjenje transparentnosti pretpostavlja da su mehanizmi za formiranje vjerovanja isključivo orijentirani prema istini. Argumentiram da je to gledište neuvjerljivo kada uzmemo u obzir šire prakse formiranja vjerovanja. Pokazujem da Dular i Fortier ne uzimaju u obzir koliko su zapravo česta adaptivna vjerovanja koja nisu evidencijski potkrijepljena te da evolucija disjunktivnih kognitivnih sustava nije toliko nevjerojatna. Oni bi trebali dopustiti da ponašanja koja nisu usmjerena na otkrivanje istine igraju ulogu u našim mehanizmima za formiranje vjerovanja. S druge strane, mogli bismo ograničiti objašnjenje koje nude Dular i Fortier na deliberativni kontekst, odnosno mogli bismo ih shvatiti kao da dopuštaju mogućnost ne-evidencijalnog formiranja vjerovanja izvan deliberativnog konteksta, ali kao da identificiraju ključno svojstvo za objašnjenje transparentnosti u evolucijskim mehanizmima koji su usmjereni prema istini kada funkcioniraju unutar deliberativnog konteksta. Međutim, to dovodi do drugog roga dileme: tada ne bismo imali drugačije objašnjenje od onoga

koje sam ponudila na drugom mjestu (2018), objašnjenje koje Dular i Fortier izričito ne uzimaju u obzir jer smatraju da se odnosi na neki drugi kontekst. Rad zaključujem kratkim razmatranjem širih implikacija koje se odnose na objašnjenje transparentnosti, prirode vjerovanja i perspektive za razvoj pragmatizma. Zaključujem da Dularovo i Fortierovo osporavajuće objašnjenje transparentnosti daje preznačajnu ulogu istini kada govorimo o utemeljenju naših vjerovanja, ili, ako ne daje, onda zapravo samo ponovno izlažu gledište koje sami na drugom mjestu eksplicitno odbacuju. Ostali smo, dakle, s objašnjenjem koje ne bismo smjeli željeti, ili s objašnjenjem koje smo već imali.

Ključne riječi: vjerovanje, transparentnost, doksastičko promišljanje/deliberacija, evolucijsko osporavanje, razlozi za vjerovanje

A REPLY TO S SIDDHARTH'S 'AGAINST PHENOMENAL BONDING'

Gregory Miller
Independent scholar, Atlanta, Georgia, USA

ABSTRACT

In this journal S Siddharth has recently argued that the phenomenal bonding response to the subject summing argument for panpsychism is question begging, therefore we should reject constitutive forms of panpsychism. The argument specifically focuses on the proposals of Goff and Miller. In this reply, I show that the argument is unsound.

Keywords: panpsychism; combination problem; phenomenal bonding; subject summing; combinationism; S Siddharth

Odgovor na S Siddharthov 'Against Phenomenal Bonding'

Gregory Miller
Nezavisni istraživač, Atlanta, Georgia, USA

SAŽETAK

U ovom časopisu S Siddharth je nedavno argumentirao da je cirkularan panpsihistički odgovor fenomenalnog povezivanja na argument zbrajanja subjekata, te da bismo stoga trebali odbaciti konstitutivne oblike panpsihizma. Argument se posebno usredotočuje na prijedloge Goffa i

Millera. U ovom odgovoru pokazujem da je Siddharthov argument nepouzdan.

Ključne riječi: panpsihizam, problem kombinacije, fenomenalno spajanje, zbrajanje subjekata, kombinacionizam, S Siddharth

Translation / prijevod:
Iva Martinić
Marko Jurjako

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.uniri.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.

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 € 50

institutions € 100

Croatia:

individuals € 30

institutions € 60

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, Portal of Scientific Journals of Croatia (HRČAK), ANVUR (Italy), Sherpa Romeo, and SCImago

(AG)10