SELF-FAVORING THEORIES AND THE BIAS ARGUMENT

Bálint BÉKEFI

ABSTRACT In a recent article, Bernáth and Tőzsér (2021) defend what they call the Bias Argument, a new skeptical argument from expert peer disagreement. They argue that the best contrastive causal explanation for disagreement among leading experts in philosophy is that they adopt their positions in a biased way. But if the leading experts are biased, non-experts either are also biased or only avoid bias through epistemic inferiority. Recognizing this is expected to prompt one to decrease one's confidence in one's philosophical beliefs. This paper argues that some beliefs are immune to a key premise of the Bias Argument. To show this, the paper develops the concepts of self-favoring theories, decisive support, and standing-incommensurable disagreements. A plausible example of a self-favoring theory, dubbed Mere Reformed Protestantism, is sketched. Many disagreements over self-favoring theories and over beliefs decisively supported by self-favoring theories are shown to be standing-incommensurable. It is then argued that when non-experts are in standing-incommensurable disagreements with experts, the standards of assessing expertise are themselves controverted. This result undercuts the move in the Bias Argument from expert bias to non-expert bias. Finally, a couple reservations about the role of self-favoring theories in philosophy are addressed.

KEYWORDS: the Bias Argument, expert peer disagreement, philosophy, self-favoring theories, Bernáth, Tőzsér

Introduction

The Bias Argument proposed by Bernáth and Tőzsér (2021) seeks to show that philosophers ought significantly to decrease their confidence in their "substantive, positive, and factual philosophical beliefs" (372)[1] upon realizing that these are subject to disagreement among experts, and that the best explanation of this fact is that philosophers are motivated by arational biases in choosing their positions. The authors shirk away from accepting this argument, realizing that it is plausibly self-defeating, but they state that they "know no plausible way to refute its premises" (363). The argument runs as follows (364–365):

(1) In debates about the truth of a given philosophical proposition *p*, those participants who are recognized as the leading experts in the field are *epistemic peers to each other* (with respect to *p*), even though they have conflicting beliefs

---

[1] Page number-only references refer to Bernáth and Tőzsér (2021).

about the truth of p.

(2) If the conflicting philosophical beliefs of epistemic peers have contrastive causal explanations, then these beliefs are biased—their formation and persistence are decisively influenced by factors that do not indicate the truth of these beliefs.

(3) All beliefs (including philosophical ones) have contrastive causal explanations…

(C1) The conflicting philosophical beliefs of the leading experts in philosophy are biased.

(4) If the conflicting philosophical beliefs of the leading experts are biased, then my philosophical beliefs that are heavily debated among experts are biased, too…

(C2) My philosophical beliefs that are heavily debated among experts are biased…

(5) If my philosophical beliefs that are heavily debated among experts are biased, then I have a strong reason to significantly reduce my confidence in them…

(C3) I have a strong reason to significantly reduce my confidence in my philosophical beliefs that are heavily debated among experts.

In response, I argue that one need not accept (4) if one's view under dispute implies that one is epistemically superior to the so-called "leading experts." I call disagreements with this feature standing-incommensurable. In the first section, I show how standing-incommensurable disagreements (SIDs) arise from beliefs decisively supported by self-favoring theories (SFTs). While doing so, I attempt to give a plausible characterization of SFTs and develop an example. In the second section, I consider whether people in SIDs should doubt (1)—and conclude that shouldn't—and elaborate on why they have no reason to accept (4). Finally, I address two reservations one might have about attributing such philosophical relevance to SFTs.

1. How Self-favoring Theories Give Rise to Standing Incommensurability

This section proceeds in four parts. First, Bernáth and Tőzsér's notion of epistemic peerhood is extended to develop an analogous notion of epistemic superiority. Second, this notion is used to characterize SFTs, and an example is introduced. Third, the ways a belief might be "decisively supported" by an SFT are explored. Fourth, it is argued that disagreements over SFTs and many disagreements over beliefs decisively supported by SFTs are standing-incommensurable.

1.1. Epistemic Superiority

Bernáth and Tőzsér define epistemic position or standing in terms of *explanation*: their skeptical conclusion follows when differences in epistemic position cannot

explain a disagreement. (The authors later clarify that by explanation they mean *contrastive causal explanation* (368).) This approach gives rise to the following, original, definition of epistemic peerhood:

> Two epistemic agents are epistemic peers to each other in relation to the truth of *p* if and only if… the difference between their evidence bases, their competences, and their resistance to biases fails to explain their disagreement about the truth of *p*. (366)

This allows us to propose a parallel definition of epistemic superiority:

> Given epistemic agents *S* and *V*, *S* is an epistemic superior of *V* in relation to the truth of *p* if and only if (1) *S* has, on balance, better evidence base, competences, and strength of resistance to biases in relation to the truth of *p* than *V*, and (2) their disagreement about the truth of *p* is explained by (1).

1.2. Self-favoring Theories

This section develops the point that some theories claim or imply that some who accept them are epistemic superiors to all who reject them. Consider *Mere Reformed Protestantism* (MRP), a religious theory. It includes theism, the death and resurrection of Christ, life after death, biblical inspiration, and perhaps some other commitments. It also renders very plausible the following two theses:

> *The noetic effects of sin*. Everyone is sinful; moreover, sin has affected the human epistemic faculties such that all non-regenerate people have a diminished capacity for acquiring and correctly assessing the evidence for MRP.

> *Regeneration*. God has supernaturally restored, at least partially, those epistemic faculties of some people which are relevant for acquiring and correctly assessing the evidence for MRP; moreover, this set of regenerate people is coextensive with those who affirm MRP in some independently identifiable way (the "right way").[2]

(The "right way" is left unspecified; it might rule out obviously unreliable methods or postulate a special requirement, such as having undergone a conversion experience.)

It seems plausible that if *S* accepts MRP, recognizes the conditional plausibility of these two theses, and takes themselves to accept MRP in the right way, then *S* will rationally come to believe that *S* is in a better epistemic position concerning the truth of MRP to all people who reject it. *S* might also rightly think that this difference in epistemic positions contrastively explains the disagreement over MRP. This leads us to the definition of a SFT:

> A theory is self-favoring if and only if it implies that an independently

---

[2] For a contemporary development of these theses see Plantinga (2000, Ch. 7–8).

> identifiable subclass of those who affirm it ("proper believers") are epistemic superiors in relation to the truth of the theory to all who reject the theory.

One might suspect that SFTs are inherently implausible: how could an uninformed and unintelligent proper believer in a SFT be an epistemic superior to a well-informed and intelligent critic? If SFTs explained disbelief in terms of a systemic lack of communicable information or a systemic prevalence of identifiable errors in reasoning, SFTs would either have to be universally persuasive or become quickly falsified. Instead, there seem to be two viable possibilities. First, a SFT may involve a commitment to a non-standard epistemic principle—a claim that some doxastic method or practice is epistemically reliable (Lynch 2010, 264)—which in turn confirms the SFT. Second, a SFT might be such that disagreements over it must be reducible to disagreements over *basic* doxastic attitudes: basic beliefs, intuitive judgments of plausibility that cannot be further analyzed, or "ur-priors" (Pittard 2019).

Note that this contention is incompatible with some non-explanatory notions of epistemic superiority. Frances, for example, lists a set of individually necessary and jointly sufficient conditions for epistemic superiority, which includes being more informed, intelligent, and intellectually careful, less biased, having thought about the relevant issues longer and more deeply, and having considered the interlocutor's evidence and reasons (Frances 2010, 420–421). It seems that a proper believer in an SFT might be epistemically superior to someone who rejects the SFT in the explanatory sense—his evidence base and competences are better on balance, and this explain the disagreement—while failing to meet one of Frances' necessary conditions, such as having thought longer and more deeply about relevant issues than his interlocutor. This is not a problem, however, because the Bias Argument is intimately wedded to explanatory conceptions of epistemic standing.

Notions similar to that of a SFT have appeared in recent literature on the epistemology of religious disagreement. Pittard (2014, 90) talks about *self-favoring theories of epistemic credentials*, which are distinctively *partisan*; Choo (2021, 1143) talks about disagreement over "the relevant credibility-conferring features" which may yield *unconfirmed superiority disagreements*; and Moon (2021) develops the notions of *epistemically self-promoting* and *others-demoting propositions* that allow for implications of unequal epistemic reliability among those who accept and those who reject a given proposition. Compared to these accounts, the two options sketched above shed further light on the possible sources of such epistemic asymmetry.

1.3. Decisive Support

S's belief that p is *decisively supported* by q if S believes that p in view of his total evidence *including* q but would *not* believe that p if his total evidence *lacked* q but were otherwise identical.[3] This notion can be applied to any pair of propositions, but in order to keep its relevance in sight for our purposes, I shall stick to examples involving beliefs decisively supported by MRP. Most trivially, decisively supported beliefs might include propositions that are entailed by q—such as theism by MRP. But there may be cases where the addition of q tips the scales in favor of p even if it does not entail p, and even if plenty of other evidence goes into assessing p.

Consider substance dualism as an example. There are formidable philosophical and scientific arguments both for and against substance dualism; yet it is probably safe to say that philosophers who lack any religious background are very unlikely to be substance dualists today. Suppose that Doug, an irreligious philosopher, assigns substance dualism an epistemic probability of 0.2 given his knowledge of the relevant evidence. Suppose further that Doug later undergoes a religious conversion and comes to affirm MRP with considerable confidence.[4] It is perfectly possible that Doug will eventually conclude that the epistemic probability of substance dualism conditional on his original background knowledge in conjunction with MRP is much higher—somewhere around 0.8, say. In such a case Doug would presumably be rational in accepting substance dualism, and this belief of his would be decisively supported by his belief in MRP.[5]

There is at least one further way a belief can be decisively supported by a SFT. Consider Patricia, a moral philosopher who is agnostic about moral realism. Now, Patricia has a specific reason for being agnostic: she has strong moral intuitions which she takes to be *prima facie* evidence for moral realism, but she is also aware of evolutionary debunking arguments against moral intuitions which she takes to undermine any inference from moral intuitions to moral realism. Suppose that upon coming to accept MRP, Patricia realizes that evolutionary debunking arguments now have much less weight, given that she now takes the evolutionary process to have been superintended by a God invested in our moral intuitions. The original defeater for her inference from her moral intuitions to

---

[3] There are a few complications with subtracting a proposition r from one's evidence base—issues like conjunctive propositions including r, or multiple propositions jointly entailing r. Plantinga addresses these complications and eventually suggests: "let's say that EB$_{me}$-(B) is any subset of EB$_{me}$ [my background evidence] that doesn't entail (B) [the belief to be deleted] and is otherwise maximally similar to EB$_{me}$" (Plantinga 2011, 187). This is what is meant here.

[4] The diachronic nature of these examples only serves illustrative purposes.

[5] Points along similar lines have been made by Plantinga (1984; 2011, Ch. 6).

moral realism may thus be defeated, and she may rationally come to affirm moral realism—decisively supported by her belief in MRP.

We have thus seen three ways beliefs can be decisively supported by SFTs: entailment, increased conditional probability, and defeater-defeat. Note that these ways can combine and iterate to generate further decisively supported beliefs; however, if a chain of support extends too far, with independent evidence always feeding in, it is likely that the support will eventually cease to be decisive.

1.4. Standing Incommensurability

A disagreement might be said to be a standing-incommensurable disagreement (SID) if the relative epistemic standing of the agents cannot be assessed independently of the truth or falsity of the disputed proposition. I shall argue that disagreements over SFTs and over propositions which for one party are decisively supported by a SFT will often be standing-incommensurable.

Notice that these definitions of epistemic peerhood and superiority in section 1.1 don't say anything about how one might go about assessing epistemic peerhood and superiority. However, the most neutral approach that is also most susceptible to challenges and is hinted at by Bernáth and Tőzsér later (374) is to assess epistemic standing "independently of... views that are closely related to the philosophical issue at hand" (Ibid., emphasis removed).[6] That is, one ought not reason in the following way: I have very good reason to believe that p; my interlocutor does not believe that p; therefore, he just cannot be as bright or well-informed as I am concerning p. Though it is difficult to specify the nuances of this notion of independence (cf. Christensen 2019), the gist of the idea is clear.

A problem emerges, however, when one tries to assess epistemic standing in a disagreement over a SFT. Suppose that "Believer" is a proper believer in such a theory, while "Unbeliever" rejects it. Then, it can be independently recognized that if the theory is true, then Believer is an epistemic superior in relation to its truth to Unbeliever. It should be emphasized that *one need not accept the theory or think that the theory is plausible to accept this conditional*. Suppose further that as far as it can be assessed by ordinary means, Unbeliever is at least as well-informed, intelligent, and (un)biased as Believer. Thus, if the theory is false, then Unbeliever is either Believer's epistemic peer or Believer's epistemic superior.

In such a situation, the relative epistemic standing of "Believer" and "Unbeliever" cannot be decided without taking a position on the truth of the

---

[6] In context, the authors are expressing what they take to be the only legitimate way of demoting one's interlocutor.

contested SFT. An independent assessment *yields no result*. It is such cases that are here called SIDs.

Does this just "[push] the problem further down" (373)? Can the standing incommensurability involved in disagreements over SFTs be resolved on the next level? Consider Doug again. Doug and Dan disagree over MRP. Doug believes that this disagreement is explained by Dan's suffering from the noetic effects of sin, while Dan offers some other explanation. Why do they disagree about the explanation? Both parties may agree that *this* disagreement is explained by Doug's belief in MRP and Dan lack of such a belief—they have different evidence. Why do they have different evidence? Well, Doug thinks it's because Dan's suffering from the noetic effects of sin… and on and on it goes.

More abstractly, the dialectic can be depicted as a circle having two components (Figure 1): (1) disagreement over *the SFT*; and (2) differing explanations of the disagreement. Since Believer's explanation of their disagreement in terms of Believer's epistemic superiority is decisively supported by the theory (and Unbeliever presumably rejects this explanation), Believer and Unbeliever's disagreement over the explanation is explained by their disagreement over the theory (1). No agreement seems to be forthcoming.
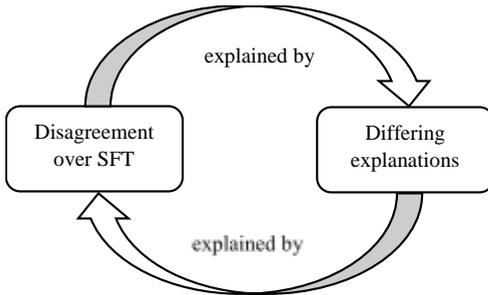

Figure 1: Circularity of explanation in SIDs

In addition to disagreements over SFTs, some disagreements over beliefs decisively supported by SFTs also turn out to be SIDs.[7] Recall Doug, convert to MRP and substance dualism. Recall further that Doug's belief in substance dualism is decisively supported by his belief in MRP. Suppose that Dan, an equally sharp and well-informed philosopher, rejects both substance dualism and MRP. Since—by our earlier definition of decisive support—Doug would reject substance dualism if

---

[7] That is, if the person appealing to a SFT has reason to consider themselves a proper believer. This qualification will be omitted in subsequent discussions of beliefs decisively supported by SFTs.

he did not accept MRP, their disagreement about substance dualism turns on a disagreement about MRP. Now, it seems that if a disagreement over p turns on a disagreement over q, then the disagreeing parties have the same relative epistemic standing in relation to p as they have in relation to q. But, since MRP is a SFT and Doug is a proper believer in MRP, there is no dispute-independent way to assess their relative epistemic standing—their disagreement is a SID.

Not all disagreements over beliefs decisively supported by SFTs turn out to be SIDs. Consider a modified version of Doug, Dougie: an intelligent layperson who accepts substance dualism and thinks its conditional probability on MRP alone is 0.8. He has a rudimentary understanding of the other evidence, which he judges to support substance dualism to a 0.5 degree. When Dougie learns that Dan, a professional philosopher, rejects substance dualism (and so do most of his colleagues!), he should suspect that there probably is relevant negative evidence that he has not considered and therefore should not take himself to have a better evidence base than the leading experts. To put it another way: because he realizes that he is unfamiliar with Dan's evidence, he is not justified in thinking that their disagreement over MRP contrastively explains their disagreement over substance dualism. Such a situation seems to require two conditions: first, the proper believer needs to think it likely that their interlocutor has weighty arguments and evidence that they are unaware of; and second, the support that the SFT lends to the given philosophical position needs to be moderate, such that other evidence has a realistic chance of offsetting it. Such cases do not generate SIDs.

2. Challenging the Bias Argument

Having introduced SFTs, decisive support, and standing incommensurability, we shall move on to consider the impact of SIDs on the Bias Argument. We will first consider its impact on (1), and conclude that SIDs could in principle, but do not in fact, call it into question. Then we will consider (4) and argue that someone whose belief is decisively supported by a SFT has no reason to accept it, and is thereby immune to the argument.

2.1. Considering (1)

Recall premise (1) of the Bias Argument (365):

> (1) In debates about the truth of a given philosophical proposition p, those participants who are recognized as the leading experts in the field are *epistemic peers to each other* (with respect to p), even though they have conflicting beliefs about the truth of p.

Bernáth and Tőzsér defend this premise indirectly (373–374). They argue that the only alternatives to considering the leading experts epistemic peers is to assume one side's superiority either arbitrarily or at best based on one's first-order evidence. The authors object that such a move can be mirrored by one's opponents, and this fact exposes it as unreliable reasoning. The only tenable attitude is to regard experts on both sides of a debate as epistemic peers.

This line of argument seems inapplicable to SIDs, for in such cases, the epistemic superiority of one party is implied by their position. This has two ramifications: first, it is not arbitrary in the sense that Bernáth and Tőzsér decry; and second, to regard both parties in SIDs as epistemic peers is not to be neutral or independent, but in fact to claim that the SFT involved is false (for its implication is denied). The only independent assessment would be to suspend judgment about the relative epistemic standing of the disagreeing parties—and then (1) would be unsupported.

But the fact is that most disagreements among leading experts are not SIDs. Most advocates of most or all positions don't have a SFT decisively supporting their view. Even if 5 out of 10 leading substance dualist experts support their belief decisively by a SFT, that theory is unlikely to explain the disagreement between the *rest* of the leading substance dualists experts and their interlocutors in epistemically relevant terms. Thus, the proper believer may consider their side of the leading experts unbiased only if the disagreement concerns a position which (almost) only proper believers in that theory adopt. Puzzlingly, this becomes more likely the more esoteric a philosophical position is, and the smaller the group of experts defending it is. Even so, it is hard to think of plausible candidates. It seems, then, that since most disagreements among leading experts in philosophy are not SIDs, not even the believer in a SFT has reason to question (1).

2.2. Against (4)

Recall premise (4) of the Bias Argument (365):

> (4) If the conflicting philosophical beliefs of the leading experts are biased, then my philosophical beliefs that are heavily debated among experts are biased, too.

Bernáth and Tőzsér motivate this premise by arguing that, assuming that the leading experts' philosophical beliefs are biased as stated in premise (1), "[t]here are only two things that can save philosophical beliefs from being biased. One is being *uninformed*. … Another relevant case is *incompetence*" (369, emphasis original). That is to say, if the leading experts are biased but the non-expert's belief is to be explained by epistemically relevant factors, those factors will involve the non-expert's epistemic inferiority.

This argument transparently assumes that the leading experts are the epistemic superiors of non-experts. The authors define "the leading experts" as "those philosophers who are more or less consensually regarded by the community of philosophers as the ablest participants in the debate concerning *p*" (365). Thus, the standard of expertise is the consensus of the philosophical community.

At least one party in a SID will deny that this standard of expertise is accurate with respect to the disputed view.[8] For example, recall Patricia, moral realist and proper believer in MRP. She believes that most philosophers miss a crucial bit of evidence—namely, MRP—and that their missing this bit of evidence contrastively explains their failure to accept moral realism. She also believes that she is in a better position to judge the truth of MRP than those who reject it because of *The noetic effects of sin* and *Regeneration*. Naturally, she also realizes that the consensus of the philosophical community does not take belief in MRP into account when judging who the leading experts are—in fact, she considers it very likely that most philosophers considered to be leading experts do *not* accept MRP. In such a case, she will have no reason to consider the leading experts her epistemic superiors, and thus will not sympathize with Bernáth and Tőzsér's arguments for (4).

To generalize, someone who accepts a "philosophical [belief]… heavily debated among experts" in a way that puts them in a SID with said experts will be unmotivated to accept (4). The Bias Argument does not, therefore, bear on beliefs held in such a way.

3. Two Reservations About Self-favoring Theories

The most likely reaction to the argument up to this point is neither enthusiastic agreement nor outright disagreement. Rather, I expect many readers to be *vaguely unsettled* by this focus on SFTs in philosophy. This section is devoted to addressing a couple possible reservations about belief in SFTs: first, that it is incorrigible; and second, that philosophers should "bracket" it when doing philosophy.

3.1. Is Belief in Self-favoring Theories Incorrigible?

One might worry that SFTs insulate their proper believers into a sort of epistemic circularity by allowing them to reject any objections raised against the theories as resulting from poor evidence, poor competences, or bias. Consequently, proper believers will never have to consider counterevidence to their theories—in a sense,

---

[8] For any standard of expertise, at least one party in a SID will deny that it is accurate— otherwise, the standing incommensurability could be resolved.

proper belief in SFTs turns out to be incorrigible (and resulting SIDs irresolvable). If this were so, that would certainly be concerning. As we shall see, however, the objection is hasty. I will show this through considering two distinctions in recent epistemology relevant to theory assessment.

When considering arguments against Christian belief, Alvin Plantinga distinguished *de facto* and *de jure* objections. *De facto* objections, such as many versions of the problem of evil, offer reasons to think that Christian belief is *false*. *De jure* objections, on the other hand, like the Freudian projection theory of theistic belief, argue that Christian belief is *irrational, unjustified*, or *unwarranted*—no matter whether true or false (Plantinga 2000, viii–x). Clearly, the objection that one is biased in holding a belief falls in the latter category: the belief's truth value plays no role in the argument. The thesis of this paper, then, is that SFTs are immune to one kind of *de jure* objection—a bias objection from peer disagreement—but they are not, or at least we have seen no reason to think they are, immune to *de facto* objections one might bring against them.

This point is further borne out by considering a similar distinction in recent literature on the epistemology of disagreement: the distinction between first-order evidence and higher-order evidence. First-order evidence E_1 for p is evidence relevant to assessing the truth of p. Higher-order evidence E_H about E_1 for p is evidence relevant to judging how well E_1's relation to the truth of p has been assessed (cf. Kelly 2016, Section 2). The concept of higher-order evidence is useful for connecting distinct kinds of epistemologically relevant information that seem to threaten with defeat, such as peer disagreement, poor track record, unreliable cognitive faculties, and having taken a rationality-distorting drug (cf. Christensen 2010). On such a schema, this paper argues that SFTs are immune to a certain kind of higher-order evidence (namely, peer disagreement)—but debate concerning the first-order evidence is left completely open to discussion.

We have thus seen that SFTs in principle allow for considering objections and challenging the evidence that purportedly supports them. Two points should nevertheless be acknowledged in favor of the objection. First, it may be the case that one has less rational obligation to consider the objections of one's epistemic inferiors than one's epistemic peers—though, perhaps, the obligation is less diminished in SIDs than in ordinary cases. This is far from incorrigibility, but it does mean that SFTs are more resistant to revision than more trivial beliefs. Second, *some* disagreements over SFTs may turn out to be rationally irresolvable. But this will not be merely in virtue of the theories being self-favoring, but rather some more specific feature, such as their involving contested fundamental epistemic principles (Lynch 2010).

3.2. Should Self-favoring Theories Be Bracketed?

Decisively supporting one's philosophical beliefs by a SFT might seem *unphilosophical*, perhaps illegitimately mixing religion with philosophy (cf. Schellenberg 2018). Maybe we should ask philosophers whose beliefs are decisively supported by SFTs to join those whose are not in a conditional project: they should strive to assess the probability of a given philosophical thesis *conditional on their shared evidence*, ignoring the support some of them might derive from theories not all of them share.[9] In a sense, they are asked to "bracket" their theories while doing philosophy. Then, if they fail to come to a consensus on this conditional probability, then they should significantly decrease their confidence that their shared evidence supports the given philosophical thesis to that certain degree.

Should the proper believer agree to this? Not always, for her theory might claim that she is in a better epistemic position to assess not just the theory itself, but also other issues close to its core claims. Thus, even if she were to bracket her theory *as evidence*, her theory might still provide a contrastive explanation of the disagreement over the assessment of the agreed-upon evidence in terms of her epistemic superiority. For example, a believer in MRP who accepts *The noetic effects of sin* may well be justified in thinking that those who reject MRP tend to be more biased in their assessment of the evidence for theism (entailed by MRP) than she is. So, if a SFT implies that its proper believers are epistemic superiors to those who reject the theory with respect to assessing the support their shared evidence lends to some philosophical view in domain *D,* then the Bias Argument does not demand that proper believers decrease their confidence in their initial assessment.

But suppose the given SFT does *not* imply its believer's epistemic superiority with respect to some *p* (other than the epistemic superiority gained by having the theory in one's evidence base), belief in which may nevertheless be decisively supported by the theory. Then, the philosopher accepting it may have to accept the Bias Argument and decrease her confidence in the probability of *p* conditional on the evidence shared by the philosophical community. This *would* impact her confidence and, in some cases, even her belief in *p*.

It is worth laying out three further considerations that limit the SFT believer's steadfastness. First, SFTs tend to restrict the claim to epistemic superiority to some limited domain. Second, epistemically relevant differences may explain some amount of disagreement while not explaining all amount. For example, MRP might give an epistemically relevant explanation for why Doug

---

[9] Here we diverge from Bernáth and Tőzsér's binary (belief–disbelief) model of disagreement.

would assign a 0.6 probability to the thesis that an infinite past is metaphysically impossible, while Dan would judge it to be 0.4 probable.[10] But MRP might *not* have an explanation in terms of epistemically relevant differences for a disagreement between probability judgments of 0.9 and 0.1. Third, the argument of this paper does not exempt believers in the *same* SFT of the skeptical consequences of disagreement.

4. Conclusion

This paper has argued that there is a special category of beliefs that is immune to Bernáth and Tőzsér's Bias Argument. To describe this category of beliefs, new concepts had to be proposed: those of a SFT, decisive support, and SIDs. It was argued that all disagreements over SFTs between proper believers and disbelievers and many but not all disagreements over beliefs decisively supported by SFTs are SIDs: there is no dispute-independent way to assess the relative epistemic standing of the two parties.

This conclusion was then applied to the Bias Argument. It yielded the conclusion that if the disagreement over some philosophical view between a person and "the leading experts" in philosophy is a SID, then that person has no reason to think that "the leading experts" are their epistemic superiors, and thus no reason to accept premise (4). Moreover, it was shown both through the concept of SFTs and through illustrations that such SIDs are perfectly possible.

The concept of SFTs may provoke objections; therefore, the final section of the paper was devoted to addressing two of these. It was argued that SFTs need to be incorrigible, as standing incommensurability does not prevent the believer in a SFT from examining *de facto* objections (objections concerning first-order evidence) to their view. The suggestion that belief in SFTs should be bracketed while doing philosophical work was shown to be ill-conceived, as it would not secure the success of the Bias Argument. This is because SFTs may claim that their proper believers are epistemic superiors to others not just with respect to the theory itself, but also with respect to other closely related domains of inquiry.

The paper raises questions for future inquiry about how the concepts proposed here relate to other discussions. What is the relationship between SIDs and deep disagreements? What views qualify as SFTs in philosophical, religious,

---

[10] How might MRP explain this? Most philosophers agree that the finitude of the past provides at least *some* evidence for theism; but, as we have said, *The noetic effects of sin* suggests that those who deny MRP are more biased in their assessment of the evidence for theism than those who accept it in the right way.

perhaps even psychological or economic thought?[11] These questions lie beyond the scope of the present work.

References

Bernáth, László and János Tőzsér. 2021. "The Biased Nature of Philosophical Beliefs in the Light of Peer Disagreement." *Metaphilosophy* 52:363–378. https://doi.org/10.1111/meta.12501.

Choo, Frederick. 2021. "The Epistemic Significance of Religious Disagreements: Cases of Unconfirmed Superiority Disagreements." *Topoi* 40:1139–1147. https://doi.org/10.1007/s11245-018-9599-4.

Christensen, David. 2010. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81 (1):185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x.

———. 2019. "Formulating Independence." In *Higher-Order Evidence: New Essays*, edited by Mattias Skipper and Asbjørn Steglich-Petersen, 13–34. Oxford: Oxford University Press.

Frances, Bryan. 2010. "The Reflective Epistemic Renegade." *Philosophy and Phenomenological Research* 81 (2):419–463. https://doi.org/10.1111/j.1933-1592.2010.00372.x.

Kelly, Thomas. 2016. "Evidence." In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2016/entries/evidence/.

Lynch, Michael P. 2010. "Epistemic Circularity and Epistemic Incommensurability." In *Social Epistemology*, edited by Adrian Haddock, Alan Millar, and Duncan Pritchard, 262–277. Oxford: Oxford University Press.

Moon, Andrew. 2021. "Circular and Question-Begging Responses to Religious Disagreement and Debunking Arguments." *Philosophical Studies* 178:785–809. https://doi.org/10.1007/s11098-020-01458-3.

Pittard, John. 2014. "Conciliationism and Religious Disagreement." In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain, 80–97. Oxford: Oxford University Press.

———. 2019. "Fundamental Disagreements and the Limits of Instrumentalism." *Synthese* 196:5009–5038. https://doi.org/10.1007/s11229-018-1691-1.

---

[11] For a few imaginative non-religious suggestions see Turnbull (2021).

Plantinga, Alvin. 1984. "Advice to Christian Philosophers." *Faith and Philosophy* 1 (3):253–271. https://doi.org/10.5840/faithphil19841317.

———. 2000. *Warranted Christian Belief*. Oxford: Oxford University Press.

———. 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. Oxford: Oxford University Press.

Schellenberg, J. L. 2018. "Is Plantinga-Style Christian Philosophy Really Philosophy?" In *Christian Philosophy: Conceptions, Continuations, and Challenges*, edited by J. Aaron Simmons, 229–243. Oxford: Oxford University Press.

Turnbull, M. G. 2021. "Religious Disagreement Is Not Unique." In *Religious Disagreement and Pluralism*, edited by Matthew A. Benton and Jonathan L. Kvanvig, 91–107. Oxford: Oxford University Press.