Logos & Episteme

an international journal
of epistemology

TABLE OF CONTENTS

RESEARCH ARTICLES

STRANDED RUNNERS: ON TRYING TO BRING JUSTIFICATION HOME

Christopher T. BUFORD

ABSTRACT: Those who endorse a knowledge-first program in epistemology claim that rather than attempting to understand *knowledge* in terms of more fundamental notions or relations such as *belief* and *justification*, we should instead understand knowledge as being in some sense prior to such concepts and/or relations. If we suppose that this is the correct approach to theorizing about knowledge, we are left with a residual question about the nature of those concepts or relations, such as justification, that were thought to be first but are now second. Jonathan Jenkins Ichikawa has recently proposed that we understand justification in terms of *potential* knowledge. Ichikawa combines his view of knowledge and justification with what initially seems to be a natural complement, *epistemological disjunctivism*. While Ichikawa focuses on hallucination, I shift the focus to illusion. I argue that the combination of justification as potential knowledge and epistemological disjunctivism entails that perceptual beliefs that arise from illusions are not justified.

KEYWORDS: illusion, disjunctivism, Jonathan Jenkins Ichikawa

Those who endorse a knowledge-first program in epistemology claim that rather than attempting to understand *knowledge* in terms of more fundamental notions or relations such as *belief* and *justification*, we should instead understand knowledge as being in some sense prior to such concepts and/or relations.[1] If we suppose that this is the correct approach to theorizing about knowledge, we are left with a residual question about the nature of those concepts or relations, such as justification, that were thought to be first but are now second. While it is open to the knowledge-first theorist to become an eliminativist about justification, many appear to want to still make room for justification. Jonathan Jenkins Ichikawa has recently proposed that we understand justification in terms of *potential* knowledge.[2]

---

[1] Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000), esp. Ch 3.

[2] I do not discuss Ichikawa's endorsement of epistemic contextualism. Ichikawa himself notes that it doesn't "play a central role in the discussion or defense of JPK." See Jonathan Jenkins Ichikawa, "Justification is potential knowledge," *Canadian Journal of Philosophy* 44, 2 (2014): 184-206 and Jonathan Jenkins Ichikawa, *Contextualising Knowledge: Epistemology and Semantics* (Oxford: Oxford University Press, 2017), 120.

Christopher T. Buford

Ichikawa combines his view of knowledge and justification with what initially seems to be a natural complement, *epistemological disjunctivism*.[3]

According to the epistemological disjunctivist, there is an important and fundamental epistemological difference between certain cases of veridical perception and corresponding cases of hallucination.[4] In standard cases of veridical perception, say looking at a maple tree in autumn, a subject sees the tree and comes to know that there is a tree on the basis of the factive mental state – *seeing that*. In contrast, in a case of hallucination, where there is no maple tree, there is no *seeing that* state on which to base a perceptual belief. The epistemological disjunctivist claims that the support provided by the veridical perceptual experience of seeing the tree is both different in kind and epistemically superior to the mere appearance of a tree. That there is a significant epistemological difference between the two cases would seem to fit well with the knowledge-first program and has not surprisingly been endorsed by some knowledge-firsters.[5] If both subjects are equally well positioned with respect to justification or evidence, then we may be tempted to understand knowledge as true belief + the relevant shared epistemic properties. Such a maneuver is in tension with the claim that knowledge cannot be understood in terms of more fundamental notions such as justification. As we will see, the combination of justification as potential knowledge and epistemological disjunctivism has an apparent cost; perceptual beliefs that arise from illusions are not justified.

I.

Here is Ichikawa's statement of the view that justification is potential knowledge.

> (JPK) S's belief is justified iff there is a possible individual, alike with respect to all relevant basic evidence and cognitive processing, whose corresponding belief is knowledge.[6]

---

[3] The criticism leveled here is admittedly narrow in that it applies to a fairly specific combination of positions, but as I note, the combination should be for some an attractive one.

[4] A difference over and above the fact that in standard cases of hallucination, knowledge is ruled out by factivity. See Alex Byrne and Heather Logue, "Either/Or," in *Disjunctivism: Perception, Action, Knowledge*, eds. Adrian Haddock and Fiona Macpherson (Oxford: Oxford University Press. 2008), 314-319, John McDowell, *Perception as a Capacity for Knowledge* (Milwaukee, WI: Marquette University Press, 2011), John McDowell, "Criteria, Defeasibility, and Knowledge," *Proceedings of the British Academy* 68 (1982): 455-479, and Duncan Pritchard, *Epistemological disjunctivism* (Oxford: Oxford University Press, 2012).

[5] Williamson, *Knowledge and its Limits*.

[6] Ichikawa, *Contextualizing Knowledge*, 119.

Of course, to fully understand JPK and its implications, we need to ask about basic evidence and cognitive processing. Ichikawa provides further explanation by revising an earlier view[7] of which facts are relevant

> [i]t [JPK] has it that there can be no difference in justification facts without a corresponding difference in the subject's relevant situation. Although I do not (any more) assume that all such relevant facts are geographically internal- I allow that factive perceptual states are relevant features- JPK is still a *mentalist internalist* approach to justification; differences in the external world that do *not* make a difference with respect to a subject's mental states do not affect whether a belief is justified.[8]

JPK so understood is able to capture two aspects of justification with which it has often been associated. JPK allows that there can be subjects with false, but justified beliefs. To see this suppose that my extremely trustworthy friend Aman tells me that she went to see a movie last Thursday. Suppose further that Aman is right that she went to a movie last week but it wasn't on Thursday but Wednesday. It seems easy to imagine an alternative scenario in which all of my relevant evidence and cognitive processing remains the same, but Aman goes to the movie on Thursday. In such a case, I would presumably know that she did. And if this is in fact possible, JPK delivers the result that my false belief is in fact justified. JPK is also consistent with Gettier's[9] true, justified beliefs that do not amount to knowledge. My knowledge that Trip is in a European capital, on the basis of his postcard from Madrid, might be destroyed by his last minute trip to Berlin, but not necessarily my justification. Had Trip remained in Madrid, I might have known that he was in a European capital. Thus, my current belief is justified according to JPK.

Ichikawa is aware that on our reading of JPK, it conflicts with what he labels the New Evil Demon Intuition,[10] the intuition that me and my subjectively indiscernible BIV-twin are equals with respect to the justificatory status of our beliefs about the external world. My BIV twin, who lacks certain factive mental states, fails to be justified in his beliefs about the external world. If JPK is correct, and my twin *is justified*, then there must be a possible individual sharing all the same basic evidence and cognitive processing who also possesses knowledge about the world. But this possible individual will not share all the same basic evidence due to the fact that the knowledge will be based on factive mental states; states that my BIV

---

[7] Ichikawa, "Justification is potential."

[8] Ichikawa, *Contextualizing knowledge*, 116.

[9] Edmund L. Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23, 6 (1963): 121-123.

[10] Stewart Cohen, "Justification and Truth," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 46, 3 (1984): 279-295.

twin does not have as part of his basic evidential set. Ichikawa does not see respecting the intuition as a requirement of a satisfactory account of justification.

> So I no longer consider respecting the New Evil Demon intuition in generality a central desideratum for a theory of justification. I continue to recognize it as somewhat intuitive, but I now categorize it with other internalist intuitions that I am comfortable rejecting if necessary.[11]

Given that there are extant views of justification, certain forms of reliabilism for example, that also reject the New Evil Demon intuition, the combination of JPK and epistemological disjunctivism should not be rejected solely for this reason.[12] We have also seen that JPK, even when combined with epistemological disjunctivism, allows for justified beliefs to possess features thought to be required of the correct theory of justification.

II.

Consider now the case of illusion. Suppose that I am looking at a table in what I take to be normal circumstances. I do not know it but the table is white and lit with a red light. Since I am unaware of the non-standard conditions, I believe that the table is in fact red. Further, since the table is not red, I cannot know that it is. I am though justified in believing the table to be red. What is the status of my belief according to JPK? Here is the relevant bi-conditional.

> (JPK-Table) My belief that the table is red is justified iff there is a possible individual, alike with respect to all relevant basic evidence and cognitive processing, whose corresponding belief is knowledge.

The problem is that once we have endorsed epistemological disjunctivism, we have ruled out such a possible individual. Notice that my evidence cannot include the factive mental state seeing that the table is red given that the table is not in fact red. I, looking at the red-lit table, and a possible individual who knows that the table is red according to the epistemological disjunctivist will thus not be alike with respect to all relevant basic evidence. We will differ in at least one important respect; he will *see that* the table is red.[13] Further, introducing the possibility that my epistemic

---

[11] Ichikawa, *Contextualizing Knowledge*, 110

[12] Not all forms of reliabilism must reject the intuition. Leplin (Jarrett Leplin, "In Defense of Reliabilism," *Philosophical Studies* 134, 1 (2007): 31–42) and Graham (Peter J. Graham, "Epistemic Entitlement," *Noûs* 46, 3 (2012): 449–483) propose ways of understanding the relevant notion of reliability that allow my BIV twin to have justified perceptual beliefs.

[13] There of course might be some overlap in evidence between myself and this possible individual. For example, I see that the surface of the table appears red and he does as well. However, the epistemological disjunctivist is committed to there being a significant difference in the evidence

twin, with respect to basic evidence and cognitive processing, might know that the table is red without seeing that the table is red will not help. The reason is that my twin will have to have some additional (non-basic) evidence, perhaps testimonial, to know that the table is red.[14] Counting this individual as relevant to determining whether my belief is justified makes it much too easy to arrive at a justified belief. Imagine a subject who makes a perceptual judgment about an object in the distance under far from ideal circumstances. That we can also imagine a similar subject with knowledge due to being told that the judgment is correct should not lead us to think the original judgment is justified.

It should be stressed that denying justified belief in the case of the table appears worse than denying the New Evil Demon intuition. That this is so allows us to respond to a worry about the argument offered above.[15] The concern that might be voiced is that all that we have accomplished is to point out a consequence of Ichikawa's view, a consequence that Ichikawa might well be aware of and happy to accept. I don't however believe that being aware of a commitment and being content to accept it are necessarily sufficient to defuse the strength of the objection to Ichikawa on offer. In fact, Ichikawa's own discussion of a version of JPK (applied to perception) is telling.

> But given the disjunctivist neo-Moorean position gestured at in the previous chapter, we may also wish to be open to factive perceptual states as among the relevant respects. If so, then JPK will not entail JPK*i*. It is possible, for instance, for two intrinsically identical subjects to differ with respect to their factive perceptual states; JPK allows that this may suffice for a difference in justification. If Despard sees that a villager is sick, and Margaret has an intrinsically identical hallucination, Despard's belief may be justified (indeed, it is knowledge), even though Margaret's is not (since her basic-evidence counterparts are hallucinating).So I am now more attracted to this *perception* emphasizing version of JPK:
>
> **JPK***p* S's belief is justified iff there is a possible individual, alike to S with respect to all relevant factive perceptual states and cognitive processing, whose corresponding belief is knowledge.[16]

There are features present in the above case that may help mitigate any theoretical costs to rejecting what amounts to a variant of the New Evil Demon intuition (i.e. that Margaret possesses a justified belief that the villager is sick). First, Margaret (like

---

or facts available.

[14] And if the table seems red to my twin, and he knows that the table is red without further evidence, then presumably my twin sees that the table is red. Thus, he is not alike with respect to all basic evidence.

[15] Thanks to an anonymous referee for pressing this objection.

[16] Ichikawa, *Contextualizing Knowledge*, 115-6

my BIV twin) is cognitively detached from her environment; she is not perceiving anything since she is hallucinating. Since Margaret is not in perceptual contact with her environment, many may find the denial that she has a justified perceptual belief in this case acceptable. However, notice that in the case of the table, one is in fact in perceptual contact with the table; and as others have noted, one sees the table in part by being seeing the (apparent) color of the table.[17] This difference makes JPK's commitment in the case of the table less appealing. Second, many of us have never experienced hallucinations that are qualitatively identical to everyday perceptual experiences. The fanciful nature of the thought experiment, as with the BIV case, can be used to sow doubt as to the accuracy of the intuitions it generates. This feature is also not present in the table case. Most of us have been the victim of some type of perceptual illusion.

If you haven't, please compare the circles below.[18]
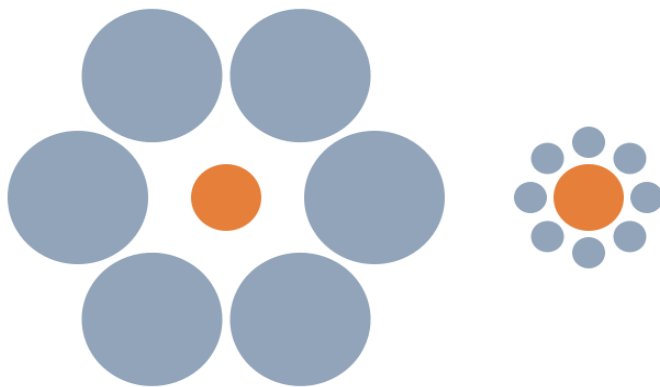
Given the context, it may not be too surprising to learn that the two orange circles are the same size even though the circle on the right appears larger. Imagine confronting the images without knowing about the illusion. Suppose also that forming a justified perceptual belief that the circle on the right is larger is possible.[19]

---

[17] Arthur David Smith, "Disjunctivism and Illusion," *Philosophy and Phenomenological Research* 80, 2 (2010): 384-410.

[18] This is the Ebbinghaus Illusion, named for its discoverer Hermann Ebbinghaus (https://en.wikipedia.org/wiki/Ebbinghaus_illusion).

[19] As an anonymous referee rightly notes, the table case and the circle case are not exactly similar. In the latter, the explanation for the illusion involves reference to biases built into our perceptual systems. I agree but would add that such biases are the product of the history and development of the human perceptual system. Thus, the illusions are in part explained by the fact that our systems, given facts about history and development, are functioning properly in such cases. Also, denial of justified belief in illusion due to bias cases brings along a commitment to unjustified belief not just

Is there a possible individual, like you with respect to all basic evidence and cognitive processing, who knows that the circle on the right is larger? If basic evidence includes factive states such as seeing that, the answer is no. For once again, that individual will see that the circle on the right is larger. And this you cannot do.

III.

The preceding considerations highlight an arguably inescapable feature of knowledge-first approaches to justification that are coupled with epistemological disjunctivism.[20] Combining the two positions yields the result that knowledge is both prior to justification conceptually and of a significantly different epistemic kind than a merely justified belief. It is perhaps then natural to move towards a view of justification as a type of broken or failed knowledge, a state one gets oneself into when one does everything right but the world happens not to cooperate. However, there are numerous ways for the world to fail to cooperate. Sometimes such lack of compliance is consistent with my being in the same position with respect to my basic evidence. The case involving Aman above is of this type. We can alter the truth-value of what is believed without changing the basic evidence. This will not always be the case. In cases of perceptual illusion, the evidence possessed by the possible knower will necessarily be different once we require factive states to be part of the evidential set. It then becomes difficult to see how we can derive a proper theory of justification by focusing *solely* on cases of knowledge, be they actual or potential. To help us see this, let us consider another proposal[21] to understand justification in terms of potential knowledge.

> (JuJu) If in a world w_1 S has mental states M and then forms a judgment, that judgment is justified if and only if there is some world w_2 where, with the same

---

in unlikely but possible lighting conditions, but also in more mundane cases (e.g. illusions of angle of presentation caused by textural features). For more on both of these points see Tyler Burge, "Perceptual Entitlement," *Philosophy and Phenomenological Research* 67, 3 (2003): 503-548 and Tyler Burge, "Disjunctivism and Perceptual Psychology," *Philosophical Topics* 33 (2005): 1-78.

[20] Whether *every* view that combines knowledge-first epistemology with epistemological disjunctivism will possess this feature of course depends in part on which proposals are to count as knowledge-first versions epistemological disjunctivism. For example, an anonymous referee suggests the following and wonders whether it can avoid the commitments of Ichikawa's approach: my total justificatory support for my belief that P in the good case is better than my total justificatory support for my belief that P in the bad case. Given the characterization of epistemological disjunctivism offered here, it isn't obvious that this view is a version of epistemological disjunctivism.

[21] Alexander Bird, "Justified Judging," *Philosophy and Phenomenological Research* 74, 1 (2007): 81-110.

> mental states M, S forms a corresponding judgment and that judgment yields knowledge.[22]

If JuJu is combined with the requirement that basic evidence include factive mental states, my belief that the table is red, when it is white but lit to look red, will not be justified. In order to know that the table is red, S must see that the table is red. Yet possession of this mental state will necessitate that S and myself do not share all the same mental states; thus ensuring that my current belief is not justified.

As Ichikawa does with respect to the New Evil Demon intuition, a proponent of knowledge-first epistemological disjunctivism might respond by rejecting the intuition that the false perceptual beliefs in question (i.e. that is a red table; that circle is larger) are justified. Such a move though is in danger of moving the position from one that is inconsistent with some of our intuitions concerning justification to one that is unacceptably counterintuitive. I agree with Ichikawa that an acceptable theory of justification should not be too "stingy;" theories of justification that rule out many of our ordinary beliefs as justified are *prima facie* troubling.[23] Since at least some of our ordinary beliefs involve illusions of the types discussed above, the combination of JPK and epistemological disjunctivism risks being cheap to a fault.

Finally, while a knowledge-first epistemologist need not endorse epistemological disjunctivism, the argument offered here helps stoke a general worry about any attempt to reconcile knowledge-first views with a plausible account of justification. One way to effect such a rapprochement is to understand justification in terms of knowledge. This strategy risks, as we saw with Ichikawa's proposal, placing unacceptable demands on justified belief. Another route is to attempt to give an account of justification that does not rely on an account of knowledge. However, since knowledge has been assumed to be prior to justification, it is hard to see what would guide such an analysis.[24]

---

[22] Bird, "Justified Judging," 84.
[23] Ichikawa, *Contextualizing Knowledge*, 111.
[24] Thanks to Jonathan Jenkins Ichikawa and numerous anonymous referees for helpful comments.

WHY RIP MATTERS?
REEXAMINING THE PROBLEM OF COGNITIVE DYNAMICS

Filip ČUKLJEVIĆ

ABSTRACT: The aim of this paper is to reexamine the importance of Rip van Winkle's case for the problem of cognitive dynamics. First I shall present the main problem of cognitive dynamics. Then I shall explain the relevance of Rip's case to this problem. After that I shall provide a short presentation of the main solutions to this problem. I shall explicate the problem concerning the manner in which philosophers who propose those solutions defend their response to the question of Rip's case. My argument shall be that they defend their response either in overly dogmatic or in circular way. Finally, I shall suggest a way out of that problem.

KEYWORDS: cognitive dynamics, belief retention, indexicals, propositions

1. The Problem of Cognitive Dynamics

In short, *cognitive dynamics* is an investigation of conditions required for either the retention of or the change in propositional attitudes. More extensively, the subject of cognitive dynamics is an investigation of conditions needed for the persistence of a propositional attitude – such as tokens of belief, hope, fear etc. – through time, as well as of those needed for such propositional attitudes to cease to exist at a given time.[1]

Following Joao Branquinho, I shall note some basic assumptions about propositional attitudes that shall be assumed throughout this paper. I shall suppose that a propositional attitude is a relational mental state. As its relata we have subjects on one side and a certain type of abstract objects, called *propositions*, on the other.[2]

---

[1] Cf. Joao Branquinho, "The Problem of Cognitive Dynamics," accessed January 12, 2021, http://www.joaomiguelbranquinho.com/uploads/9/5/3/8/9538249/cog.pdf, 1, David Kaplan "Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals," in *Themes From Kaplan*, eds. Joseph Almog, John Perry and Howard Wettstein (New York, Oxford: Oxford University Press, 1989), 537-8.

[2] Branquinho also refers to them as *thoughts*. Branquinho, "The Problem of Cognitive Dynamics," 1. Here I refer to them as propositions, not because I propose Russellian way of explaining the structure of propositions, as opposed to Fregean (Frege uses the term "thought"), but because I

Filip Čukljević

A proposition is the content of a propositional attitude. The most important type of propositions for this paper shall be the *singular* propositions. These are the propositions that are about a particular object.[3] A propositional attitude that has this type of proposition as its content – such proposition that can be expressed by a sentence which contains at least one indexical referring term (these propositions shall be called *indexical* propositions) – shall be the most interesting case of a propositional attitude for those who dwell on cognitive dynamics.[4] The focus of this paper shall be on indexical propositions that are about particular times. However, it should be noted that there are indexical propositions that are about particular places, as well as those that are about objects that can be identified by means of several sensory modalities.[5] Also, a belief shall be considered as a paradigmatic case of a propositional attitude.

Philosophers who deal with cognitive dynamics notice the following problem regarding beliefs that have indexical propositions as their content. There are cases in which a subject has to readjust the verbal term they use in order to express their belief at a specific time – such beliefs have indexical propositions as their content – in order to retain that belief at some later time. We shall see that situation becomes even more problematic when we find out that the verbal terms which are *prima facie* considered as appropriate means for a subject to use in order to retain their initial belief, are actually inappropriate.[6]

Before I proceed to that, due to the existence of the aforementioned problematic cases, a thesis that Branquinho refers to as *the central problem of cognitive dynamics* can be formulated. Branquinho formulates the problem by asking the following question: which circumstances must obtain in order for us to say that a subject has retained their belief from a time t; that is which sentence (with an appropriate indexical expression) should a subject be inclined to accept at time t' – which is after time t – in order to retain the belief from a time t?[7]

Let me clarify the situation with a help of the famous example. Suppose that a person on a particular day – that shall be called d – says "Today is a beautiful day" while truly believing in that. Which conditions must be satisfied, that is, which

---

consider the term "proposition" more neutral than the term "thought", at least given this type of problematics.

[3] Branquinho, "The Problem," 1.

[4] Cf. Branquinho, "The Problem," 1-2.

[5] Branquinho, "The Problem," 2.

[6] Branquinho, "The Problem," 2.

[7] Branquinho, as well as Kaplan, notices an analogue problem that arises when a propositional attitude is changed. However, I shall focus on the problem that arises due to its retention. Cf. Branquinho, "The Problem," 2-3, Kaplan, "Demonstratives," 537-8.

sentence should that person be inclined to accept on the following day – that shall be called *d+1* – in order to retain their previous belief? The answer that probably first comes to mind, to people like Gottlob Frege, David Kaplan and others, is the following – that person must be inclined to accept the sentence "Yesterday was a beautiful day" on *d+1*.[8] This claim, which hints at the possible solution to the problem of belief retention, Branquinho names the *natural realignment claim*.[9]

He offers two different readings of this claim. The first one is called the *necessity claim*, and the second one the *unqualified sufficiency claim*.[10] According to the necessity claim, in order for a person from the previous example to be able to retain their belief it is *necessary* for them to be inclined to accept the sentence "Yesterday was a beautiful day" on *d+1*. On the other hand, according to the unqualified sufficiency claim, in order for this person to retain their belief it is *sufficient* that they are inclined to accept the given sentence on *d+1*.[11] Branquinho uses this difference between two readings of the natural realignment claim in order to define a more precise role for Rip's case in the problem of cognitive dynamics. In the following chapter, I shall analyze to what extent is this Branquinho's attempt successful.

2. The Case of Rip van Winkle

The natural realignment claim can be criticized either by criticizing the necessity claim or by criticizing the unqualified sufficiency claim. According to Branquinho's interpretation, the first type of critique is important in order to demonstrate the relevance of Rip's case to the problem of cognitive dynamics. In order to become clear what exactly is meant by the necessity claim and what exactly is reconsidered when we criticize it, it is useful to expose the argument that Branquinho offers as a possible justification for this claim. Branquinho thinks that this argument can be used by representatives of different solutions to the problem of cognitive dynamics – obviously, as long as they accept the necessity claim itself. This is the argument:

---

[8] Branquinho, "The Problem," 3. Frege did not consider the problem of cognitive dynamics specifically, although many commentators attribute such intuition to him and use it either to defend their own or to dispute others' solutions to this problem. Cf. Gareth Evans, *The Varieties of Reference* (New York, Oxford: Oxford University Press, 1982), 192, Gareth Evans, "Understanding Demonstratives," in *Collected Papers*, ed. Gareth Evans, 291-321 (Oxford: Clarendon Press, 1985), 291-2, Kaplan, "Demonstratives," 501. On the other hand, Kaplan was the first to formulate the problem of cognitive dynamics and to consider the aforementioned solution as *prima facie* solution, although a problematic one. Cf. Kaplan, "Demonstratives," 537-8.

[9] Branquinho, "The Problem," 3.

[10] Branquinho, "The Problem," 3-4.

[11] Branquinho, "The Problem," 3-4.

1. Tracking an object over time and/or space is necessary in order to be able, at a later time, to retain the singular indexical belief that was expressed at some earlier time by a certain type of sentences (such as "Today is…" type of sentence from the previous example).
2. Having a disposition for accepting sentences of a certain kind (such as "Yesterday was…" type of sentence from the previous example) at some later times is necessarily involved in tracking.
3. Hence, having such disposition is necessary for belief retention.[12]

Here we should notice something that Branquinho does not mention explicitly enough, but is nevertheless important for the correct understanding of the necessity claim. Namely, the solution to the problem of cognitive dynamics should be able to show us what is necessary for a subject to keep a belief they had on *d* not just on *d+1*, but on some later day as well. For a subject to retain belief on some day after *d+1* their inclination to accept the sentence "Yesterday was…" on *d+1* is not enough; a subject should also be inclined to accept some other appropriate sentences on the days following *d+1*. It seems that Branquinho understands the necessity claim in the following manner: in order for a subject to retain their belief about *d* on some random later day – that shall be called *d'* – it is necessary for a subject to be disposed to accept an appropriate sentences on each day between *d* and *d'*, including on *d'*.[13] Keeping in mind previously formulated argument, having such disposition is necessary for belief retention because this disposition is necessary for tracking days over time.

Hence the understanding regarding the property that should be shared by sentences which a subject has to be inclined to accept at later times. All of those sentences should share some temporal indexical term that refers to a day the original belief is about.[14] By *temporal indexical term* I mean each indexical term that includes some temporal determinant, that is each indexical term whose appropriate use presupposes the possession of knowledge regarding how much time has passed – if not exactly, then at least approximately relative to a particular term – since the day we wish to refer to by that specific expression to the day that this expression is used. Those expressions include phrases such as 'yesterday,' 'the day before yesterday,' 'exactly 17 days ago,' 'last Thursday,' 'on the first Saturday of the last month' *etc*. The point is that a belief can be retained not just because we are inclined to accept sentences like "Yesterday was…" on *d+1*, "Two days ago was…" on *d+2*, "Three days ago was…" on *d+3*, *etc*. – which would be too tedious task as time goes on – but also

---

[12] Branquinho, "The Problem," 9.

[13] Cf. Branquinho, "The Problem," 7-9.

[14] In a couple of places Branquinho seems to be very close to explicitly accepting this claim. Cf. Branquinho, "The Problem," 7-9.

by using our knowledge regarding certain properties of those days, as well as our knowledge regarding the positions those days occupy in a general timeline. I believe that this interpretation makes the necessity claim more plausible.

However, Branquinho thinks that not all philosophers accept the necessity claim. According to him, the strongest critique of the necessity claim is provided by Kaplan. Branquinho believes that this critique, which relies on Rip's case, reveals the fact that someone who accepts the necessity claim must give a negative answer to the question regarding Rip's belief retention. If, however, someone were to give a positive answer to this question, they would oblige themselves to reject the necessity claim.[15]

I shall not get into the entirety of Rip's story. For my purposes, it shall be enough to say that our unfortunate protagonist – through an unusual series of events – fell asleep one evening and woke up in a morning, not the next one, but twenty years later. Let's suppose that on the day he fell asleep – which I shall call D – Rip had come to a belief that he had been inclined to express with the sentence "Today is a beautiful day." The question that arises is the following: has he retained such belief after waking up from his twenty-year-long sleep, on the day that I shall call D'?[16]

According to Branquinho's interpretation of Kaplan's critique, if we accept the necessity claim we have to give a negative answer to this question. The reason for that lies in Rip's inability to sincerely and reflectively accept, or be inclined to accept, sentences such as "Yesterday was a beautiful day" on $D+1$ – the day following D – which he, according to the necessity claim, should be able to do in order to retain his belief in the following days, including D'. It seems that Branquinho claims that Rip should not be able to do such thing since he had *systematically and massively* lost track of the time.[17] According to this author, the impossibility of Rip's belief retention is an unacceptable consequence for Kaplan, so he rejects the necessity claim. More precisely, Branquinho claims that Kaplan rejects the 1st assumption of the previous argument, according to which tracking objects over time is necessary for belief retention.[18] The relevance of Rip's case to the problem of cognitive dynamics would thus consist in the revelation of inconsistency between accepting the necessity claim and giving a positive answer to the question regarding Rip's belief retention.

---

[15] Branquinho, "The Problem," 9.

[16] Cf. Branquinho, "The Problem," 9, John Perry, "Rip van Winkle and Other Characters," in *Cognitive Dynamics*, ed. Jerome Dokic (Stanford: *European Review of Philosophy* 2, 1997), 35-6.

[17] Branquinho, "The Problem," 9.

[18] Branquinho, "The Problem," 9.

Filip Čukljević

Does Branquinho's interpretation of Kaplan's critique provide a proper view of the relevance of Rip's case to the given problem? According to the necessity claim, in order for Rip to be able to retain his initial belief on *D'* it is necessary – among other things – for him to be inclined to accept the sentence "Yesterday was a beautiful day" on *D+1*, and to do it sincerely and reflectively. Branquinho thinks that the positive answer to the question of Rip's belief retention brings on suspicion regarding the necessity claim, since in that case Rip retains his initial belief on *D'* without being inclined to accept the sentence "Yesterday was a beautiful day" on *D+1*, at least not sincerely and reflectively.

But, is Rip truly not inclined to accept the given sentence on *D+1* in a sincere and reflective way? In order for one to be inclined to do something, it is enough that they would do it given the appropriate circumstances, but it is not necessary to have such circumstances obtained. In order for Rip to be inclined to accept the sentence "Yesterday was a beautiful day" on *D+1* he does not have to find himself in circumstances in which he could actually do it. It is supposed that Rip has not found himself in such circumstances, since he had slept through *D+1* and one cannot accept the sentence while asleep, at least not sincerely and reflectively. However, it is unclear why he would not have accepted the given sentence in a sincere and reflective manner had he been awake on *D+1*. This is the reason why the positive answer to the question regarding Rip's case does not shed the doubt on the necessity claim – at least not by shedding doubt on his inclination to accept the sentence "Yesterday was a beautiful day" – since it seems that Rip does indeed have the appropriate inclination.

Does then Rip's case shed doubt on the (unqualified) sufficiency claim, since it does not shed the doubt on the necessity claim? In order for that to be the case, Rip must satisfy the sufficient conditions for belief retention, without having such belief. This is problematic for two reasons. First of all, Rip does not fulfill such conditions, at least not if the (unqualified) sufficiency claim is understood as analogous to my more precise formulation of the necessity claim; and there is no reason why it shouldn't be understood as such. In that case, in order for Rip to retain his belief on *D'*, it is sufficient for him to have a disposition for the acceptance of appropriate sentences on those days between *D* and *D'*, including *D'*. We do not have to get into details regarding which sentence would be appropriate on *D'*, although it would probably go something like this: "Exactly twenty years and one day ago, it was a beautiful day" or "On a particular date twenty years ago, it was a beautiful day." What would, however, be a necessary condition that this type of sentences ought to satisfy is that such sentences should be about *D*. Nevertheless – as it is assumed – on *D'* Rip is inclined to accept the sentence "Yesterday was a beautiful

day" that is not about *D*. Since he is inclined to accept this sentence it appears that, on such particular day, it seems to be impossible for him to be inclined to accept any other sentence that includes a temporal indexical expression that refers to *D*. Thus Rip's case could not qualify as a counterexample to the (unqualified) sufficiency claim.[19]

On the other hand, it is problematic that Kaplan himself did not consider such a case as a counterexample to the (unqualified) sufficiency claim and instead considered it as a counterexample to something like the necessity claim. The reason for this is that he considers a positive answer to the question of Rip's belief retention as a difficulty for the solution to the problem of cognitive dynamics, instead of a negative one. According to him, a negative answer is in accordance with the solution.[20] That is why it seems that Rip's case, *i.e.* a positive answer to the question regarding the case, must somehow put in question the necessity claim.

At first, a possibility arises from the considerations of the previous paragraph. According to the necessity claim, in order for Rip to retain his belief, he must be inclined to accept a sentence that includes a temporal indexical expression that refers to *D* on *D'*. However, as it was exposed in the previous paragraph, on *D'* Rip is inclined to accept a sentence that is about *D'-1*, so it seems impossible for him to satisfy this necessary condition. That seems to be why Kaplan claims that Rip has lost track of the time, so it seems as if we have to deny his belief retention – if we accept the necessity claim, that is. Kaplan is not bothered by Rip's lack of inclination to accept the sentence "Yesterday was a beautiful day" – as Branquinho claims to be the case – but is instead bothered by Rip's inclination to accept such sentence on *D'*, which leads to his lack of inclination to accept an appropriate sentence on that day.

I think that it is useful to present the results of the considerations so far, as Branquinho does, in the form of an argument in favor of the (more precisely formulated) necessity claim. The assumptions of such argument shall be specified by referring to tracking the specific and relevant-for-our-case type of objects – days – instead of referring to the objects in general. It is important to note that tracking days is not exactly analogous to tracking three-dimensional objects. The latter requires more or less constant perceptive contact with the objects.[21] However, when it comes to days, it is impossible to keep track of them in such a manner. When a certain day passes, there is no way for us to run into it again in the same way, at some time in the future, as it is the case with three-dimensional objects. This is why

---

[19] This is the reason why it cannot serve as a counterexample to the qualified sufficiency claim, that is created by including some additional conditions. Branquinho, "The Problem," 7.
[20] Cf. Kaplan, "Demonstratives," 537-8.
[21] Cf. Evans, "Understanding," 310-1.

it is impossible to retain the constant direct contact with a certain day.[22] How is it, then, possible to keep track of day over time? Possible ways to do this will be discussed later in this paper. For now, there is one more thing left to note – that "keeping track of day over time" means to keep a day in our cognitive view and in that way be able to have a belief regarding it at some time in the future.

Also, instead of the two standard assumptions, my argument will include three of them. The advantage of such argument will become clear in the following chapter, in which I shall consider different solutions to the problem of cognitive dynamics. Then the importance of Branquinho's hidden assumption for the problem of cognitive dynamics shall become obvious. Until then, the new argument in favor of the necessity claim is as follows:

1. Tracking a day *d* over time is necessary in order for one to be able to, on some day in the future, retain a singular indexical belief that was, on *d*, expressed by a sentence that included a specific temporal indexical expression that referred to *d*.
2. Keeping track of the time regarding *d* is necessary for tracking *d* over time.
3. The possession of disposition to accept sentences that include specific temporal indexical expression that refer to *d* on the future days is necessary for keeping track of the time regarding *d*.
4. Hence, the possession of such disposition is necessary for belief retention.

Philosophers concerned with the problem of cognitive dynamics as a rule do not make a difference – or, at least, nothing in their writings suggests such difference – between keeping track of the time regarding *d* and tracking *d* over time. Moreover, it seems that the general view is that tracking *d* presupposes keeping track of the time regarding *d*. I do not think that this is a trivial assumption. Keeping track of the time regarding *d* is one way to track *d*. However, it is not clear why it should be the only one. We shall see that some philosophers claim that a belief regarding a certain day can be retained by memory. I believe that in such case we would not have to keep track of the time regarding *d* in order to be able to track *d* over time. Keeping track of the time regarding *d* presupposes the possession of certain knowledge regarding the place occupied by *d* relative to other (particular) days. If we accept the assumption 3 of the argument in favor of the necessity claim, that knowledge is manifested in the possession of disposition to accept appropriate sentences on certain days. Such knowledge, however, is not necessary for belief retention through memory.

After this consideration of the correctness of Branquinho's interpretation of Kaplan's view on the relevance of Rip's case for the given problem, we are finally

---

[22] Perry, "Rip van Winkle," 35.

able to ask the following question – is Rip's inability to retain his belief an acceptable consequence? The views of the philosophers concerned with the problem of cognitive dynamics differ regarding this issue. Kaplan, who was the first to react to this consequence, considers it problematic. As he states, Rip's inability to retain his belief – given his loss of track of the time – *seems strange*.[23] Kaplan does not give a particular explanation for this impression; thus it seems to be an intuitive one. John Perry agrees with Kaplan regarding the unacceptability of this consequence. However, unlike Kaplan, he bases his belief on a certain argument that will be examined more closely in the following section.[24]

On the other hand, Gareth Evans considers this consequence as unproblematic. As the reason for this acceptability, he refers to the obvious fact that someone who lost track of the time cannot retain their temporal beliefs.[25] Similarly to Evans, Branquinho claims to be inclined to think that Rip most likely is not able to retain his belief. He claims that someone who systematically and massively losses track of time can hardly retain their temporal beliefs. Moreover, he thinks that maybe it is best to say that someone in the given situation cannot form temporal beliefs at all.[26]

We have seen why the answer to the question of Rip's belief retention is relevant to the problem of cognitive dynamics. If someone who offers a solution to this problem accepts the necessity claim they commit themselves to the negative answer to the previous question. If, however, they claim the positive answer they would have to reject the necessity claim. Respectively, they would have to reject at least one of the three assumptions of the argument in favor of this claim. Now I shall briefly consider some of the main solutions to the problem of cognitive dynamics. We shall see that those solutions are, at least in part, results of the stances taken towards the Rip's case and the necessity claim, *i.e.* towards the assumptions of the argument in favor of it.

3. Solutions to the Problem of Cognitive Dynamics

Although it cannot be said that Kaplan had provided a specific solution to the problem of cognitive dynamics, his view of this problem should be mentioned because of its influence on the latter solutions. Namely, as we have noticed, Kaplan considers the possibility of Rip's belief retention as an intuitive one, so the necessity claim comes off as problematic to him. Moreover, Branquinho claims that Kaplan

---

[23] Kaplan, "Demonstratives," 538.

[24] Cf. Perry, "Rip van Winkle," 35-6.

[25] Evans, "Understanding," 311.

[26] Cf. Branquinho, "The Problem," 9-10.

rejects it; more precisely, Kaplan rejects the 1st assumption of the given argument in favor of this claim, since he does not consider keeping track of days over time as necessary for belief retention.[27] However, I do not think that we are given enough textual evidence in order to be able to ascribe the rejection of any specific assumption to Kaplan. All of the assumptions of the argument in favor of the necessity claim seem to be plausible for Kaplan – hence the problem – since he finds the possibility of Rip's belief retention as equally plausible.[28] The idea of a possible solution to this problem can hardly be found in Kaplan; there is only a strong conviction that a problem regarding the original assumptions exists.

Unlike Kaplan, Evans provides a solution to the given problem by accepting the necessity claim and denying the possibility of Rip's belief retention. On the one hand, there is strong textual evidence in favor of Evans' acceptance of the 2nd assumption of the necessity claim, since he equates tracking day over time with keeping track of the time.[29] Given that, Evans explicitly accepts the other two assumptions – he considers keeping track of the time as necessary for belief retention and the disposition to accept certain sentences as necessary for keeping track of the time.[30] As a consequence of accepting the necessity claim Evans accepts Rip's inability to retain his belief, without noticing any problems. According to him, Rip has lost track of the time and it is not a bit unusual for someone who has lost track of the time to be unable to retain their temporal beliefs.[31]

It can be noticed that Evans defends his answer to the question of Rip by referring to the assumption that is a part of his solution to the problem of cognitive dynamics. Here I shall not get into the more detailed examination of Evans' solution.[32] Just as it is the case with other solutions, my main goal here is to examine a solution given its context, that is based on the authors' views regarding Rip's case and the assumptions of the argument in favor of the necessity claim.

Perry's solution to the problem of cognitive dynamics follows, in general, Kaplan's view on this problem. Similarly to Kaplan, Perry considers Rip as able to retain his belief and, thus, is inclined to reject the necessity claim.[33] At first, it seems unclear which of the assumptions of the argument in favor of the necessity claim Perry would reject. The answer to this question can be examined more properly after

---

[27] Branquinho, "The Problem," 9.
[28] Cf. Kaplan, "Demonstratives," 537-8.
[29] Cf. Evans, *Varieties*, 194-6, Evans, "Understanding," 309-11.
[30] Cf. Evans, *Varieties*, 194-6, Evans, "Understanding," 309-11.
[31] Evans, "Understanding Demonstratives," 311.
[32] More on that topic can be found in Evans, *Varieties*, 192-6, Evans, "Understanding," 306-11.
[33] Perry, "Rip van Winkle," 14.

examining how, according to Perry, belief retention is possible. Here Perry follows one of Kaplan's ideas as well. When we believe in a certain proposition we always do so under a certain – in Perry's case doxastic – *character*. Character is, simply told, a way in which we believe in a certain proposition. In order to be able to keep our belief in a proposition even after a few changes in the context, we must have an appropriate doxastic character under which we believe in it.[34] Besides the notion of doxastic character, another relevant part of Perry's solution to the problem of cognitive dynamics is the notion of *information games*. Information games involve the acquisition and later application of a belief about an object.[35] Among other things, they show us ways in which we can retain our beliefs about objects. Here I shall not go into a detailed account of the idea of information games, and shall limit myself to Perry's use of it and the notion of doxastic character in order to describe Rip's case.

Perry claims the following: Rip had, on the day he had fallen asleep, formed a belief that it had been a beautiful day. He believed in it under the character "Today (the day of this thought) is a beautiful day." After he had woken up twenty years later, Rip, thinking that he had woken up after just one night, tried to update his belief in accordance with his thought of the context change by using the character "Yesterday (the day before the day of this thought) was a beautiful day."[36] Since his opinion on the context change is wrong, he cannot retain his belief by using this character. However, this does not mean that he cannot do so by using another character. If he has some memories of that day he can retain his belief by using the character "That day (the one that I remember) was a beautiful day." If he, however, does not have any memories of that day, he can still retain his belief by using the character "That day (the day this belief had been formed) was a beautiful day."[37]

We can notice that the key difference between Perry's take on necessary conditions for belief retention and that of those who accept the necessity claim lies in the different understanding of what makes a sentence an appropriate candidate to accept at later times in order to be able to retain the original belief. According to the necessity claim – as I had previously noticed – that sentence must include a temporal

---

[34] The notions of proposition (or content), context and character Perry inherits from Kaplan, but while Kaplan applies them to sentences Perry applies them to beliefs (thus the notion of *doxastic character*). Cf. Perry, "Rip van Winkle," 19-24.

[35] Cf. Perry, "Rip van Winkle," 24-31.

[36] Updating is, according to Perry, an information game in which we infer something about an object about which we have a belief; however, this inference is not based on observed or inferred movements or changes in object our belief is about, but is instead based on changes in our situation or some general change (such as the passage of time). Cf. Perry, "Rip van Winkle," 29-30.

[37] Perry, "Rip van Winkle," 35-6.

indexical term that, in Rip's case, refers to D. Perry accepts the condition that such sentence has to be about D. However, he does not think that an indexical expression that refers to that day has to be a *temporal* indexical expression. Characters such as those that Perry suggests in order for Rip to be able to retain his belief include indexical expressions ("that day," "the day I remember," "the day that this belief was formed"), but these are not temporal indexical expressions. The temporal indexical expressions refer to a particular day relative to the day of their utterance. By using a particular temporal indexical expression we cannot refer to the same day independently of the day of its utterance.

This is not the case with Perry's expressions. By using Perry's proposed characters Rip could – given that Perry is indeed right – retain his belief not just on D', but on any other day as well. Moreover, in order to be able to refer to any day we wish to refer to by the use of some temporal indexical expression, we ought to have at least some vague knowledge of the positions that the day of our uttering the temporal indexical expression and the day we wish to refer to hold in the entire timeline. In other words, we need to have at least vague knowledge of the time that had passed between those days. That is, we need to keep track of the time. This is not the case with Perry's expressions. In order to be able to use them to refer to the day we wish to refer to we do not need to have any idea of how much time has passed since that day.

That is why I think that Perry's solution to the problem of cognitive dynamics rejects the 2nd assumption of the argument in favor of the necessity claim. Perry claims that belief retention is possible even once track of the time is lost by some other way of tracking a day (by memory, for example).[38] Although Perry does not state this explicitly, given our previous considerations – as well as the fact that Perry does not say anything that would support the claim that he rejects the 1st or the 3rd assumption, nor that he accepts the 2nd one – I think it would be for the best to ascribe to him such view of the argument in favor of the necessity claim.

Finally, there is Branquinho's solution, unique among the solutions to the problem of cognitive dynamics insofar as he rejects the necessity claim yet claims that Rip cannot retain his belief. Branquinho, like Perry, rejects the necessity claim. He does so since he believes that someone, Jones for example, who accepts the sentence "Today is a beautiful day" on d at 23:58 p.m., could retain belief expressed by such sentence without being inclined to accept the sentence "Yesterday was a beautiful day" on $d+1$ at 00:01 a.m., since he did not know if midnight had passed.[39]

---

[38] This is supported by the textual evidence according to which it is certain that Perry thinks that Rip had lost track of the time. Cf. Perry, "Rip van Winkle," 35-6.
[39] Branquinho, "The Problem," 10.

Branquinho thinks that Jones tracks *d* in a certain manner, without having the appropriate disposition, since he believes that tracking the relevant day in some way is necessary for belief retention. Moreover, Branquinho explicitly accepts the 1st assumption of the argument in favor of the necessity claim.[40] Which assumption he rejects then? Just like in Perry's case, I shall not immediately provide the answer to this question, but shall instead focus on the positive things Branquinho has to say about belief retention. Like Perry, Branquinho thinks that memory is important for belief retention. Namely, he thinks that in order for a subject to be able to retain their belief that *p*, it is necessary that they remember that *p* (this is an explanation for Jones' case as well).[41] If someone remembers that *d* was a beautiful day, it is one way in which they can track *d* and thus retain their belief.

Which assumption of the argument in favor of the necessity claim does Branquinho reject then? Keeping the original argument (the one with two assumptions) in mind, I would claim that Branquino would reject the 2nd assumption. However, as was previously mentioned, I believe that such argument is a consequence of the failure to make a difference between notions of tracking day over time and keeping track of the time. Thus, keeping in mind the improved version of the argument, it is best to understand Branquinho as claiming that keeping track of the time is not necessary for belief retention, thus rejecting the 2nd assumption. The rejection of the 3rd assumption would be problematic since it is unclear how we could keep track of the time without a disposition to accept appropriate sentences. On the other hand, Branquinho, similarly to Perry, proposes a way for tracking a day without keeping track of the time – through memory.

I think that Branquinho's views of the problem of cognitive dynamics face certain difficulties. First of all, it is questionable whether the case of Jones represents a better argument against the necessity claim than the Rip's case. Namely, Branquinho seems to understand the necessity claim as an assertion that someone, like Jones, could retain their belief about *d* at some time *t* on *d+1*, only if they were inclined to accept the sentence "Yesterday was…" at the same time *t*. However, I fail to see why the necessity claim could not be understood in a less rigid manner. This claim could be understood as an assertion that in order to be able to retain their belief about *d* at *t* on *d+1* a person has to be inclined to accept the sentence "Yesterday was…" on *d+1*, but not necessarily at *t*, but somewhat later.

Our inclination to claim that Jones had retained his belief at 00:01 a.m. on *d+1* could be explained in this manner. It is supposed that he is not inclined to accept the sentence "Yesterday was a beautiful day" at a given moment, since, knowing that it

---

[40] Branquinho, "The Problem," 10.
[41] Branquinho, "The Problem," 11.

is somewhere around the midnight, he is unsure whether midnight had already passed. However, Jones most likely would be inclined to accept such sentence a bit later, as soon as he thinks that enough time has passed. Memory can play a certain role in Jones' having this inclination; however this explanation of Jones' belief retention seems, unlike Branquinho's, to be relying primarily on Jones having a certain disposition. Since Jones is inclined to accept the appropriate sentence at a later time, we shall also say that Jones had that very same belief even at the previous time *t*. Otherwise, it would remain unclear in which way the belief's vanishing at first and later reappearance could be explained. I think that this, somewhat less strict, understanding of the necessity claim, seems appropriate, at least at first glance. In that case, we could wonder whether Rip's case presents a bigger problem for the necessity claim than Jones'.

This leads us to the second problem. It concerns Branquinho's view of Rip's case. Is Branquinho's denial of the possibility of Rip's belief retention in accordance with his solution to the problem of cognitive dynamics? Nothing in Rip's narrative seems to suggest that he could not remember that *D* was a beautiful day. Moreover, given Rip's inclination to express the belief he had on *D* with the sentence "Yesterday was a beautiful day" on *D'*, and if we assume – just as Perry did – that such belief is the only explicit belief that Rip had formed on *D*,[42] it seems as though he does indeed remember that *D* was a beautiful day. Couldn't we, then, say that Rip had been tracking the day he fell asleep, and, thus, had retained his belief about that day?

Here Branquinho becomes somewhat unclear, by claiming that we could not say that Rip retained his belief, since he had lost track of the time systematically and massively.[43] Branquinho seems to claim that, although the loss of track of the time is not *necessarily* incompatible with the possibility of belief retention, it is so in cases of systematic and massive loss of track of the time. I do not think that this claim is of any help to Branquinho's argument. His reasons against Rip's ability to retain his belief remain unclear. What does it mean to say that someone had lost track of the time *systematically*? Does that mean that it happens to them often, *i.e.* that they more than once had overslept more than one night at the time? If that were the case, then it would be untrue that Rip had systematically lost track of the time. It is presupposed that the only way that his track of the time differs from other people's is that he had once, without realizing and under the influence of some mysterious forces, slept for twenty years. Other than this there were never any similar incidents

[42] Perry, "Rip van Winkle," 36.
[43] Cf. Branquinho, "The Problem," 9-10.

involving him. Is one such incident enough to claim that he has systematically lost track of the time?

Perhaps the importance of this incident lies in its duration, as Branquinho seems to suggest.[44] In that case, this objection would be more suited for the case of Rip's *massive* loss of track of a time. However, why would the duration of his sleep turn out to be of such great importance? Would Branquinho claim that someone who sleeps over the period of two continuous nights, due to being extremely tired, could not retain their belief after waking up as well? It seems not, since he claims that the main reason for Rip's inability to retain his belief is his *massive* loss of track of the time, and not just any loss of track of the time. I am not sure whether sleep that lasted for two nights could qualify as a *massive* loss of track of the time. In any case, couldn't Rip, at least in principle, be able to update his beliefs after realizing that he had slept for twenty years, in the same way as someone who had slept for two nights? It seems as though the difficulties that Rip would face would be more of a psychological type, rather than a conceptual one. It would be way more difficult for someone to update their beliefs had they slept for twenty years than it would have been had they slept for two days. However, it is unclear why it shouldn't be *possible at least in principle* for someone to do so. Thus, it remains unclear why does it really matter whether Rip had slept for twenty years or for two days.

This is why Branquinho's argument against Rip's ability to retain his belief does not seem convincing. Moreover, Rip's ability to retain his belief – in case it is possible – looks like a much bigger issue for the necessity claim than the case of Jones. I do not see a convincing manner in which we could understand the necessity claim as if Rip could have the disposition to accept the appropriate sentences, *i.e.* those including temporal indexical expression referring to *D*. This is another confirmation of Rip's relevance to the consideration of the given problem.

4. Rip's Deeper Relevance to the Problem of Cognitive Dynamics

We have seen why Rip's case is relevant to the problem of cognitive dynamics. Now I shall provide a deeper understanding of this relevance. Kaplan and Perry think that Rip can retain his belief. Evans and Branquinho disagree. Although different, I believe that both of those types of answers to the question of Rip's belief retention share *two important characteristics*.

The first one is that all the mentioned philosophers think that *the adequate solution to the problem of cognitive dynamics should have the correct answer to this question as a consequence*. If a solution to this problem as its consequence has an

---

[44] Cf. Branquinho, "The Problem," 9-10.

answer to this question that is perceived as wrong that solution would be deemed as inadequate. This could be formulated in a more direct manner, without referring to someone's opinion – *the adequate solution to the problem of cognitive dynamics must have the correct answer to the question regarding Rip as a consequence*. I believe that all of the previously mentioned philosophers would agree with this statement. Due to this, giving the correct answer to the given question is of great importance to the problem of cognitive dynamics. Since the question, as I have previously noticed, cannot be answered in a manner everyone would immediately agree with, it is of great importance to look closer at the manner in which the answers are defended.

This leads us to another important characteristic that is shared by the different philosophers' answers. This similarity lies *in the manners their answers are defended*. Namely, they do so in one of the two following manners: they either rely on their own *intuitions* regarding the case in question; or they rely on the fact that their – allegedly adequate – solution to the problem of cognitive dynamics *has their own answer as a consequence*.

We have seen that Kaplan sees the problem in denying the possibility of Rip's belief retention. Although he does not explicitly rely on intuitions, this seems like the most appropriate understanding of his justification for his answer to the question of Rip. Evans, however, bases his answer on the obviousness of the assumption that someone who loses track of the time cannot retain their temporal beliefs. Evans accepts this assumption as a part of his solution to the problem of cognitive dynamics. Perry provides a different answer, but defends it in the same manner. In defending his answer, Perry relies on the notions of doxastic character and information games, which he uses to formulate his own solution to the problem of cognitive dynamics.

Finally, the only answer to the question of Rip that perhaps might be characterized as not sharing this other characteristic is Branquinho's. He does not claim that he simply feels that Rip cannot retain his belief – thus, he does not rely on his intuitions – nor his answer seems to be justified by referring to the ideas he bases his solution to the problem of cognitive dynamics on. However, his justification of his (negative) answer to the given question is problematic, as we have already seen. The positive answer to this question seems more in accordance with his solution to the problem of cognitive dynamics. However, if that were the case, Branquinho would have to defend his (now positive) answer by referring to his own solution to the given problem. But, even if it were that way, what exactly is the problem with the way Branquinho and others defend their answers?

A justification of a certain claim that relies on intuitions becomes especially problematic in cases of a serious conflict over the given claim. We have seen that philosophers disagree on the matter of whether Rip could retain his belief. Referring to intuitions in order to solve this problem does not seem like a fruitful project. Evans, like Kaplan, could probably refer to his intuitions, although their answers differ.

If relying on intuitions turns out to be a problem in this case, what is problematic with relying on one of the solutions to the problem of cognitive dynamics? A justification of a certain claim that is based on the fact that such claim is a consequence of a solution to a certain problem or a theory is not problematic *per se*. However, I do think that it is so in this case. If Rip's case should serve as an adequacy test for the solution to the problem of cognitive dynamics – and we have seen that it should according to philosophers engaged in this problem – we shall encounter a difficulty in justifying the answer to the question regarding Rip by referring to the solution to the given problem. Namely, since Rip's case represents an adequacy test for the solutions to the problem of cognitive dynamics, the justification of a certain solution will, in part, consist of it having the correct answer to the question regarding Rip as a consequence. But what *is* the correct answer to this question? If the answer to the question is defended by referring to it being the consequence of a given solution, it seems as if we are trapped in a vicious circle of reasoning. A certain solution is defended by referring to it having a certain answer as a consequence, while that same answer is defended by referring to it being the consequence of a given solution. This does not seem right. That is why I believe that this type of justification of an answer to the given question is not satisfying.

But is there a better way available? My answer to this question is positive. However – someone could perhaps claim – would not any argument that has a certain answer to the question regarding Rip as its conclusion also serve as a solution to the problem of cognitive dynamics? I do not think that this is the case for two reasons. First, the problem of cognitive dynamics does not refer to the Rip's case exclusively, but to many other cases as well. It seems unlikely that the argument – or some version of it – that has certain answer to the question regarding Rip as its conclusion, would also provide an answer to the questions regarding other cases. The other reason is that if the Rip's case should serve as an adequacy test for the solutions to the problem of cognitive dynamics then from that solution should follow not only *that* Rip can(not) retain his belief, but *how* is that (im)possible as well. On the other hand, the type of argument that I find desirable would show that Rip can(not) retain his belief, but not necessarily how it is (im)possible as well. This is a task for the solution to the problem of cognitive dynamics.

Filip Čukljević

It should be noted here that Kaplan, Evans and Perry have a special motivation to take a specific stance regarding the question of the possibility of Rip's belief retention and the necessity claim. Namely, Kaplan and Evans, just like Frege, think that certain types of expressions – that include the indexical expressions – include a specific semantic component that has a double role. Firstly, it should determine the reference of a given expression and – secondly – it should do so by the use of cognitively significant mental content. Kaplan, as well as Perry, has *character* as such a semantic component, while Evans uses the Fregean notion of *sense*.[45]

In spite of the differences between those notions, the possibility of Rip to retain his belief presents the threat for both of the solutions in the same manner. More precisely, if he could retain his original belief (the one that was about D) even on D' by the use of the expression "yesterday," than a given semantic component could not satisfy both of the previous conditions. If an object that Rip's belief is about is determined by the expression "yesterday," *i.e.* if he thinks of that object as of the day that preceded D', then his belief would not be about D, but about the day right before D', that I shall name D'-1. However, in that case he would have not retained his original belief, for the simple reason that the belief that D had been a beautiful day could be true while the belief that D'-1 had been beautiful could be false at the same time. If, however, Rip's belief is about D due to the use of an expression "yesterday," then it cannot follow from his thinking of an object that his belief is about as of a previous day. Since he fails to see which other expression could Rip use in order to retain his belief, while believing that he can do so, Kaplan encounters a problem to which he does not provide a solution. Evans sees a simple solution to this problem – Rip cannot retain his belief. On the other hand, Perry finds an adequate (at least according to him) alternative to the expression "yesterday" in the expression "that day."

Now we see that more is at stake when considering Rip's case and the necessity claim than it perhaps seemed to be the case at first. However, this is not the case against what was previously claimed in this chapter. Kaplan, Evans and Perry do not defend – nor they could do so adequately – their answers to the question regarding Rip by referring to the considerations of this paragraph. That rather stands as their motivation behind the answers to the given question. We have seen, however, the type of reasons they actually use in order to defend their answers. Now that I had them presented, as well as what would be a more desirable type of an argument, I shall try to provide such an argument.

---

[45] Cf. Evans, *Varieties*, 14-7, Evans, "Understanding," 124, Kaplan, "Demonstratives," 505-6, 520-1, 523-4, 529-39, Perry, "Rip van Winkle," 14-24.

5. An Argument in Favor of Rip's Ability to Retain His Belief

The argument I shall provide here satisfies the conditions I proposed before for this type of argument. As we shall see, the conclusion of this argument shall be that Rip is able to retain his belief. In order to deliver the argument more easily, I shall somewhat change Rip's circumstances, although the argument itself – if one uses enough imagination – can be applied to the original version as well. First, I shall change our hero's original 18th century background into contemporary times. Then, instead of supposing that Rip formed the belief expressed by the sentence "Today is a beautiful day" on D, I shall propose that on D he had formed the belief expressed by the sentence "Today is the last day for paper submission." The belief was formed by his superiors at the university telling him so when he came to inquire on the paper submission on D (in this changed scenario Rip is a university student). As in the original version, Rip had, unknowingly, woken up not on $D+1$, but twenty years later, on D' (some heavy partying might have been involved). The attempt to express his original belief by accepting the sentence "Yesterday was the last day for paper submission" – similarly to the original version – is determined to fail. A question arises – can Rip retain his original belief?

Let me include some additional details into this story. Suppose that Rip, on the day he had woken up, knowing that there is still some chance for his (very tolerant) superiors to accept his paper just one day after the deadline, rushed to the university to deliver his quickly finished paper. How could we explain this act? The usual philosophical method would require referring to Rip's beliefs and desires. In this case, we would claim that he had desire to submit his paper and thus pass the exam and that he had – among other relevant beliefs, such as that there is still a chance for his paper to be accepted on the day after the deadline – the belief he is inclined to express by accepting the sentence "Yesterday was the last day for paper submission" on D'.

We must not yet suppose that he has knowledge that $D'-1$ is the last day for paper submission. It is yet to be determined what day his belief is about. Also, we must not think that we have committed ourselves to claim that the expression "yesterday" that Rip uses on D' does not satisfy the aforementioned two conditions requested by Kaplan and the others. If Rip turns out to be able to retain his original belief that does not mean that we have committed ourselves to the claim that he can do so by using the sentence "Yesterday was the last day for paper submission." His inclination to express the belief in such manner does not mean it is an adequate way to express this belief, nor that he cannot express his belief in some other, more appropriate, manner. What is, however, without a doubt is that if we accept this standard model of explaining human action then Rip has a belief that he is inclined

to express by accepting the sentence "Yesterday was the last day for paper submission" on D', that is that he has a belief that a certain day is the last day for paper submission. Without ascribing such belief to him we would be unable to explain his action. What is left to be determined is the exact identity of such belief, *i.e.* what day it actually is about.

I believe that the choice easily comes down to two candidates, days D and $D'-1$. If we suppose that Rip's belief is about any other day then it would not satisfy the condition of being that belief he would be inclined to express by the acceptance of the sentence "Yesterday was the last day for paper submission" on D'. If Rip would have believed that the last day for paper submission was $D+2$, for example, then he would not be inclined to express that belief in such manner – under, excluding his twenty-year-long sleep, usual circumstances that are supposed to be the case. That is why his belief is either about D or about $D'-1$.

Let us suppose that it is about $D'-1$. In that case the following question arises: in which way is it formed, that is, what is the cause of such belief? If that belief is about D, the answer to this question would be easy – the belief was formed as a consequence of Rip being told by his superiors that "Today is the last day for paper submission" on that day. But, if his belief is about $D'-1$, it seems as if there are no adequate candidates for an answer to such a question. What could cause Rip's belief regarding the day almost twenty years ago after D, the day he had fallen asleep? Moreover, Rip had been asleep throughout the entire day $D'-1$. There is, however, one answer that seems acceptable. Rip had, on D, formed the belief that it is the last day for paper submission. After he had woken up on D', believing that he had woken up on the day after D (the day $D+1$), he formed the belief that $D'-1$ was the last day for paper submission. This Rip's implicit reasoning on D' can be presented in the following manner:

1. Rip believes that D is the last day for paper submission.
2. Rip believes that $D = D'-1$
3. Rip believes that $D'-1$ is the last day for paper submission.[46]

However, according to the 1st assumption of this argument Rip has retained the belief he had formed on D. In order for this reasoning to have Rip's belief that $D'-1$ is the last day for paper submission as a consequence, it ought to have occurred on D'. Otherwise, it does not seem possible that he would mistake $D'+1$ for D. In that

---

[46] An example that is somewhat analogous to this one can be found in Perry, "Rip van Winkle," 31-2.

case, in order to arrive at the given conclusion, he would have had to believe that D was the last day for paper submission even on D'.[47]

Now we can see that if we suppose that belief necessary for explaining Rip's rushing to the university with paper on D' is about $D'-1$, then we also suppose that he believes that D is also the last day for paper submission. The other option is to claim that belief necessary for the explanation of his act is about D. No matter which option we choose, it will have as a consequence the possibility of Rip to retain his belief that D is the last day for paper submission. Since there are no other plausible options, given that we can hardly provide an acceptable explanation of his belief being about any day other than D and $D'-1$, it seems that Rip can retain his temporal beliefs even after his twenty-year long sleep.

6. Conclusion

What are the consequences of this argument to the problem of cognitive dynamics? If the argument is sound, then all the solutions that accept the necessity claim are inadequate. Out of the examined solutions, this argument would exclude the one provided by Evans. When it comes to Branquinho's and Perry' solutions, this argument does not give us the insight into which one of them is true. It merely shows that Rip can retain his belief, without telling us anything either about the way it is possible, or about the way in which belief retention, in general, is possible. The contribution of this type of argument to the problem of cognitive dynamics lies in its elimination of certain solutions to the given problem as inadequate. However, this type of argument cannot show us what solution is the adequate one. In order to fulfill this important goal of cognitive dynamics another type of argumentation is needed.

---

[47] It is interesting to notice that Evans thinks that this type of reasoning is impossible, since the subject is not able to think the 2nd assumption. Evans, "Understanding," 294. The question is whether this is really the case. Why a subject, for example, would not at the same time be able to think of D as of a particular day they remember and of $D'-1$ as of the previous day, while believing that the day they remember is actually the previous day? Even in the case in which such reasoning is indeed impossible, that would only support my argument, since the given belief could not be about $D'-1$ and instead has to be about D.

A MODIFIED SUPERVALUATIONIST FRAMEWORK FOR DECISION-MAKING

Jonas KARGE

ABSTRACT. How strongly an agent beliefs in a proposition can be represented by her degree of belief in that proposition. According to the orthodox Bayesian picture, an agent's degree of belief is best represented by a single probability function. On an alternative account, an agent's beliefs are modeled based on a set of probability functions, called imprecise probabilities. Recently, however, imprecise probabilities have come under attack. Adam Elga claims that there is no adequate account of the way they can be manifested in decision-making. In response to Elga, more elaborate accounts of the imprecise framework have been developed. One of them is based on supervaluationism, originally, a semantic approach to vague predicates. Still, Seamus Bradley shows that some of those accounts that solve Elga's problem, have a more severe defect: they undermine a central motivation for introducing imprecise probabilities in the first place. In this paper, I modify the supervaluationist approach in such a way that it accounts for both Elga's and Bradley's challenges to the imprecise framework.

KEYWORDS: formal epistemology, supervaluationism, imprecise probabilities, decision-theory

Introduction

How strongly an agent beliefs in a proposition can be represented by her degree of belief in that proposition. According to the orthodox Bayesian picture, an agent's degree of belief is best represented by a single probability function. In particular, the Bayesian claims that agents must assign numerically precise probabilities to every proposition that they can entertain. On an alternative account, an agent's beliefs are modeled based on imprecise probabilities. With that, imprecise degrees of belief can be represented by a set of probability functions. A central decision-theoretical motivation for introducing imprecise probabilities is their solution to the Ellsberg Problem. In this problem, the orthodox Bayesian framework fails to adequately model the aversion of seemingly rational agents towards ambiguous actions whereas decision rules based on imprecise probabilities can do so.

Jonas Karge

Recently, however, imprecise probabilities have come under attack. Adam Elga[1] claims that there is no adequate account of the way they can be manifested in decision-making. In response to Elga, more elaborate accounts of the imprecise framework have been developed. One of them is based on supervaluationism, originally, a semantic approach to vague predicates. Supervaluationism can very naturally be applied to imprecise probabilities. With that, it solves Elga's problem.

Still, Seamus Bradley[2] showed that some of those accounts that solve Elga's problem, including supervaluationism, have a more severe defect: they undermine a central motivation for introducing imprecise probabilities in the first place. That is, their solution to the Ellsberg Problem. In this paper, I modify the supervaluationist approach in such a way that it accounts for both Elga's and Bradley's challenges to the imprecise framework.

This paper is organized as follows: In section 1, I will lay out the basic terminology of orthodox Bayesianism and the imprecise probabilities framework. In section 2, I will introduce the Ellsberg problem as a decision-theoretical motivation for imprecise probabilities. In section 3, we will look at Elga's decision-theoretical counterargument to imprecise probabilities. In section 4, I will show how supervaluationism can be applied to imprecise probabilities as well as how it solves Elga's problem. Moreover, I will introduce Bradley's argument against supervaluationism. In section 5, I will present a modified version of supervaluationism which solves Bradley's as well as Elga's problem.

1. Basic Terminology

In this section, I will briefly outline the basic terminology of the two competing views: Namely, *orthodox Bayesianism* and the *imprecise probabilities framework*.

1.1 Orthodox Bayesianism

The starting point for both views is an agent with beliefs about the world and who is capable of decision-making. According to the orthodox Bayesian picture, a *belief* can be defined as follows:

> **Definition, Belief.** A belief is a ternary relation between an agent S, an object of belief and a real number between 0 and 1.[3]

---

[1] Adam Elga, "Subjective Probabilities should be Sharp," *Philosopher's Imprint* 10, 5 (2010): 1-11.
[2] Seamus Bradley, "A Counterexample to Three Imprecise Decision Theories," *Theoria* 85 (2019): 18-30.
[3] Franz Huber, "Belief and Degrees of Belief," in *Degrees of Belief*, eds. Franz Huber and Christoph Schmidt-Petri (Dordrecht: Springer, 2009), 1-33, here 2.

We assume the objects of belief to be propositions, i.e. sets of possible worlds.[4] Moreover, the real number assigned to a proposition by an agent is called her *degree of belief* in that proposition where 0 represents the lowest level of confidence and 1 the highest level of confidence in it. Consider, for instance, the proposition P that *it will rain tomorrow*. Assume, moreover, that agent S is 70% sure that it will, in fact, rain. We can then state that S's degree of belief in P is 0.7.

Bayesianism claims that degrees of belief ought to be represented by single probability functions that assign a precise number to propositions. In order to define probability functions, we have to be more precise about what we mean by *propositions*. For that purpose, we begin by defining the set of all possible worlds and calls this set *event space*:

> **Definition, Event Space.** $\Omega = \{w_1, w_2, \ldots, w_n\}$ is called an event space where each w_i in Ω is a state of affairs, or possible world.[5]

Since propositions are taken to be sets of possible words, we can define a proposition as follows:

> **Definition, Proposition.** A proposition (or event) A is a subset of set Ω.[6]

Taking a proposition as a subset of the set of possible worlds, we can, moreover, define the set of all the propositions the agent can possibly believe in:

> **Definition, The Set of Propositions.** The set of objects of beliefs (the propositions) is the power set of Ω: 2^{Ω}.[7]

Finally, based on our definition of the set of propositions, we can define a probability function as follows:

> **Definition, Probability Function.** A probability function pr is a function Pr: $2^{\Omega} \to \mathbb{R}$, satisfying the probability axioms.[8]

In a next step, we can apply this view to decision-making. When an agent has to choose between different actions, Bayesianism suggests as decision rule to choose the action that yields the highest expected utility. Thus, we not only need a precise

---

[4] *Ibid.*, 2.

[5] Anna Mahtani, "Imprecise Probabilities," in *The Open Handbook of Formal Epistemology*, ed. Richard Pettigrew and Jonathan Weisberg (PhilPapers Foundation, 2019), 107-130, here 108.

[6] *Ibid.*, 108.

[7] Seamus Bradley, "How to Choose Among Choice Functions," in *Proceedings of the Ninth Symposium on Imprecise Probability: Theories and Applications*, ed. Thomas Augustin, Serena Doria, Enrique Miranda, and Erik Quaeghebeur (2015), 57-66, here 57, http://www.sipta.org/isipta15/data/paper/9.pdf.

[8] *Ibid.*, 57.

probability function but also a precise utility function that assigns a number to each possible outcome of an action reflecting the agent's value of that outcome.[9] Given those functions, the expected utility for an action can be calculated as follows:

> Let Pr(S) be the degree of belief for an event to be the case, let u(O) be the utility value the agent assigns to the consequence of an action, given the event S. Let A_i be some action. Now, the expected utility of A_i can be calculated as follows:
>
> **Definition, Expected Utility.** $EU(A_i) = \sum_{j=1}^{m} Pr(S_j) \times u(O_{ij})$.

That is, we multiply the agent's degree of belief in an event by the utility value of the outcome of that action and, subsequently, sum those values for all possible outcomes.

To sum up, *orthodox Bayesianism* has two characteristics relevant to our discussion:

1) An agent's belief state is represented by a probability function. The probability function maps each relevant proposition to a real number between 0 and 1. This number is the agent's degree of belief in that proposition.

2) Rational agents must choose an action that has maximum expected utility based on the agent's degrees of belief in the relevant propositions.

1.2 Imprecise Probabilities

Assume, our agent S has to evaluate the proposition *the European Union will consist of exactly 27 member states in 20 years*. What precise probability should she assign to that proposition? Since orthodox Bayesianism represents an agent's belief with a single probability function, such a precise value has to be given.[10] Considering propositions of this type, it seems highly implausible to represent belief states with a single probability function.[11]

On an alternative account, degrees of belief can be defined based on imprecise probabilities. One way to construe imprecise probabilities is the following:

> **Definition, Imprecise Probabilities.** Imprecise probabilities are sets of probability functions.[12]

---

[9] Mahtani, "Imprecise," 119.

[10] Susanna Rinard, "A Decision Theory for Imprecise Probabilities," *Philosopher's Imprint* 15, 7 (2015): 1-16, here 1.

[11] Miriam Schoenfield, "Chilling out on epistemic rationality," *Philosophical Studies* 158 (2012): 197-219, here 199.

[12] Seamus Bradely and Katie Steele, "Should Subjective Probabilities be Sharp?," *Episteme* 11, 3 (2014): 277-289, here 277.

Moreover, we call each such set of probability functions the agent's *representor* $\mathcal{P}$.[13] Additionally, we assume that the set of values the distributions in the agent's representor assign to a proposition covers all of the interval [x, y] with x, y $\in \mathbb{R}$. With that, we can define an imprecise degree of belief:

> **Definition, Imprecise Degree of Belief.** An agent's imprecise degree of belief in a proposition H is represented by a representor, $\mathcal{P}$, with $\mathcal{P} = \{Pr(H) : Pr \in \mathcal{P}\}$.[14]

This can be illustrated as follows: Let A be the proposition that the European Union will consist of exactly 27 member states in 20 years. Assume, our agent S is 40-60% confident that this will be the case. With that, we can represent the agent's imprecise degree of belief in A with: $\mathcal{P}(A) = [0.4, 0.6]$.

Finally, an agent's representor can be understood as a credal committee where every probability function in that committee represents the opinion of one of its members.[15] Collectively, these opinions reflect the beliefs of an agent.[16] This idea will be significant for the last section of this paper.

2. The Ellsberg Problem

A central decision-theoretical motivation for introducing imprecise degrees of belief is the so-called Ellsberg Problem. For this problem, it is vital to distinguish between risky and ambiguous actions. We take an action to be risky in case the probabilities of the relevant outcomes are known. An action is ambiguous, in turn, if the probabilities are unknown, or, only partially known.[17] The Ellsberg Problem relies on the observation that, under specific circumstances, seemingly rational agents prefer taking risky decisions, instead of ambigious ones, even though it violates expected utility theory.[18]

---

[13] *Ibid.*, 227.

[14] Bradley, "How to Choose," 57.

[15] Seamus Bradley and Katie Steele, "Learning, and the 'Problem' of Dialation," *Erkenntnis* 79 (2014): 1287-1303, here 1291.

[16] Seamus Bradley, "Imprecise Probabilities," *The Stanford Encyclopdia of Philosophy* (Spring 2019 Edition), https:// plato.stanford.edu/archives/spr2019/entries/imprecise-probabilities.

[17] Bradley, "A Counterexample," 22.

[18] Katie Steele, "Distinguishing Indeterminate Belief from 'Risk-Averse' Preferences," *Synthese* 158 (2007): 189-205, here 190.

2.1 The Ellsberg Problem with Precise Probabilities

Now, to the problem itself: In the Ellsberg Problem, an agent is told that an urn contains 30 red balls and 60 balls that are either blue or yellow in some unspecified proportion.[19] The agent faces two decision problems: A and B.

In problem A, the agent can decide to bet on either (I) which yields $100 if the next ball drawn is red or (II) where she receives $100 if it is blue. Likewise, in problem B: The agent can bet on (III) where she gets $100 if the next ball drawn is not blue or (IV) she receives $100 if it is not red.[20] For the sake of simplicity, we can assume that receiving $100 yields a utility value of 1 and not receiving it yields 0 utility.[21] With that, we can summarize the payoffs as follows:[22]

	Red	Blue	Yellow
Problem A			
(I)	1	0	0
(II)	0	1	0
Problem B			
(III)	1	0	1
(IV)	0	1	1

Figure 1. Payoffs, Ellsberg.

A significant majority of apparently rational people chooses option (I) in problem A and option (IV) in problem B when the Ellsberg Problem is studied empirically.[23] In the following, we will call this combination of (I) and (IV) *Ellsberg preferences*. Since the probabilities for those actions are known to the agent, they classify as risky actions. By upholding to this pattern, the agent expresses an aversion towards the ambiguous options (II) in problem A and (III) in problem B.

The orthodox Bayesian, however, cannot rationalize the Ellsberg preferences since there is no precise probability an agent could possibly assign to drawing a blue ball such that EU(I) > EU(II) and, at the same time, EU(III) < EU(IV).[24] This can be

---

[19] *Ibid.*, 191.

[20] *Ibid.*, 191.

[21] Daniel Ellsberg, "Risk, Ambiguity, and the Savage Axioms," *The Quarterly Journal of Economics* 75, 4 (1961): 643-669, here 655.

[22] Figure based on: Bradley, "A Counterexample," 23.

[23] Steele, "Distinguishing," 191.

[24] *Ibid.*, 191.

shown as follows: Let EU(I) = Pr_1, EU(II) = Pr_2, EU(III) = $Pr_1 + Pr_2$, and EU(IV) = $Pr_2 + Pr_3$. However, there is no Pr_i such that $Pr_1 > Pr_2$ and $Pr_1 + Pr_3 < Pr_2 + Pr_3$.[25]

As ambiguity aversion seems to be a feature of rational decision-making, not being capable of adequately modeling it is a problem for the Bayesian account.

2.1 The Ellsberg Problem with Imprecise Probabilities

Now, let's analyze the Ellsberg Problem with imprecise probabilities. Since the proportion of blue and yellow balls is unknown, we can assume the proportion to lie somewhere between 0 and 2/3. It could be the case, for instance, that all non-red balls turn out to be blue or that there are as many blue balls as yellow balls. Thus, a natural distribution of probability functions is the following:

Imprecise Degrees of Belief, Ellsberg. $\mathcal{P}$(blue) = $\mathcal{P}$(yellow)= [0, 2/3] and Pr(red) = 1/3.[26]

Representing an agent's belief state with imprecise degrees of belief implies that the expected utility for a given action will be imprecise also. That is, it corresponds to a set of utility values given by the probability functions in the agent's representor. Since those can overlap, the possible actions under considerations can turn out to be incommensurable.[27] Given the imprecise probabilities for the Ellsberg problem, the expected utilities can be summarized as:

Imprecise Expected Utilities, Ellsberg. (I) = 1/3, (II) = [0, 2/3], (III) = 1/3 + [0, 2/3], (IV) = 2/3.[28]

Since expected utility theory cannot be applied to intervals, we have to look for alternative decision rules for imprecise probabilities.[29] One such possible rule is the *Maximin Rule*. It tells us the following: For each action, there is a lowest expected utility value. This lowest value is the minimum expected utility for an action. In a decision problem, Maximin recommends the agent to choose the action that has the maximum minimum expected utility.[30]

In order to apply Maximin to the Ellsberg Problem, we first have to state its minimum expected utility values:

Minimum expected utility values, Ellsberg. EU(I) = 1/3, (II) = 0, (III) = 1/3, and (IV)

---

[25] Ellsberg, "Risk, Ambiguity," 655.

[26] Steele, "Distinguishing," 195.

[27] Bradley and Steele, "Should Subjective," 278.

[28] Steele, "Distinguishing," 195.

[29] Nils-Eric Sahlin, "Unsharp," *Theoria* 80, 1 (2014): 100-103, here 100.

[30] Mahtani, "Imprecise," 121.

= 2/3.

With that, the Maximin Rule recommends choosing (I) in problem A and (IV) in problem B. This corresponds to the Ellsberg preferences. Hence, we have shown that there is at least one decision rule for imprecise probabilities that rationalizes the Ellsberg preferences, and, with that, it rationalizes a case of ambiguity aversion.

To sum up, empirical studies indicate that agents tend to have the Ellsberg preferences even though it violates expected utility theory. According to orthodox Bayesianism, those agents are irrational. However, it seems like they are, in fact, rational.[31] With that, we have a case where orthodox Bayesianism fails to adequately model an instance of rational decision-making: it does not succeed in modeling ambiguity aversion. Still, the Ellsberg preferences can be rationalized with the imprecise framework.[32] This is taken to be a decision-theoretical motivation for the imprecise probabilities framework.

3. Elga's Problem

Even though decision rules based on imprecise probabilities seem to perform well in cases of ambiguity aversion, they struggle in another type of decision problem. That is, when imprecise probabilities are applied to sequential decision problems.[33] When it comes to sequential decision problems, it can be the case that each decision taken individually is rationally admissible, the sequence of decision, however, can turn out to be rationally impermissible.[34]

The central sequential decision problem to this discussion has been presented by Adam Elga. Elga's argument is structured as follows: He considers three types of possible decision rules for imprecise probabilities and shows for each type that it either leads to absurd consequences in a specific decision problem, or, that it has some other severe defect. For this discussion, we will only consider the first type of decision rule.

The specific decision problem goes as follows: In a *great series of bets*, an agent is sequentially offered Bet A first, and, immediately after the agent decides whether to accept or reject Bet A, she is offered Bet B.[35] Now, let H be some proposition such as *it will rain tomorrow*. The agent is then offered the following series of bets:

Bet A. If H is true, S loses $10. Otherwise S wins $15.

---

[31] Steele, "Distinguishing," 190.

[32] Mahtani, "Imprecise," 125.

[33] *Ibid.*, 191.

[34] Bradley, "A Counterexample," 21.

[35] Elga, "Subjective," 4.

Bet B. If H is true, S wins $15. Otherwise S loses $10.

This series of bets is called *great* since S is guaranteed to win $5 in case she accepts both bets.[36] However, it is not rationally required for S to accept both bets. If she believes, for instance, that H is highly unlikely, she would be better off to accept Bet B only. Still, she is rationally required to accept at least one of the bets since rejecting both bets is strictly dominated by accepting them.[37] With that, any decision rule that permits to reject both bets rationalizes an obviously irrational action, and, thus, has to be rejected.

Let's apply Maximin to the great series of bets. The minimum expected utility for rejecting Bet A is 0. However, the one for accepting Bet A is -10. With that, Maximin suggests rejecting Bet A. Likewise, the minimum expected utility for rejecting B is 0 whereas the one for accepting it is -10.[38] Hence, our agent S should reject both bets according to Maximin. This, however, is irrational and Maximin does fail in this betting scenario.

Finally, this result can be extended to a number of decision rules for imprecise probabilities which all suggest rejecting both bets by Isaac Levi, Peter Walley, Teddy Seidenfeld, Gärdenfors and Sahlin as well as Gilboa.[39]

4. Supervaluationism

A defender of imprecise probabilities in decision-making can now choose one of two strategies:

> **Strategy 1.** It can be argued that the great series of bets does not show that decision rules for imprecise probabilities are irrational.

> **Strategy 2.** It can be accepted that it does, but that different decision rules for imprecise probabilities can be introduced that are not affected by Elga's argument.[40]

In this section, I will now introduce an approach following strategy 2: namely, supervaluationism. Supervaluationism is, originally, a semantic theory designed to handle vague predicates. The central idea is that vague predicates such as *tall* don't have a definite extension, but rather a variety of different extensions. Each possible extension of a vague predicate corresponds to a possible precisification of that

---

[36] *Ibid.*, 4.

[37] *Ibid.*, 4.

[38] Bradley, "A Counterexample," 20.

[39] Elga, "Subjective," 5.

[40] Richard Pettigrew, *Dutch Book Arguments (Draft)* (Cambridge University Press: 2019), 96, https://richardpettigrew.com/books/the-dutch-bookargument/.

predicate.[41] Since there is no definite precisification for the semantics of a vague language, the semantic value of a statement remains unclear unless there is complete agreement among the precisifications on that value.[42]

When it comes to the truth value of statements in a vague language, *complete agreement* is understood as a proposition being either *determinately true* or *determinately false* in supervaluationistic terms. This can be spelled out as follows:[43]

> **Definition, Determinately True.** If a proposition is true according to all admissible precisifications, then it is determinately true.

> **Definition, Determinately False.** If a proposition is false according to all admissible precisifications, then it is determinately false.

That said, if neither of those two is the case, it is possible for a statement to have no semantic value:

> **Definition, Indeterminately True.** If a proposition is true according to some, but not all, admissible precisifications, then it is indeterminate whether it's true.

Example. Consider the predicate *tall*. For this predicate, there are numerous possible precisifications. Each such precisification determines a threshold for what it means to be tall and not-tall. Assume, every threshold between 160cm and 200cm is an admissible precisification, but 220cm is not an admissible precisification. It that case it's determinately true that someone 220cm tall is tall whereas it's indeterminately true whether someone 170cm tall is tall.[44]

In a next step, we have to make more precise what we mean by an *admissible* precisification. In fact, supervaluationism can very naturally be applied to imprecise probabilities by giving such a precisification:

> **Definition, Admissible Precisifications.** The admissible precisifications are the functions in an agent's representor.[45]

Moreover, with that definition at hand, we can characterize a supervaluationist decision theory based on imprecise probabilities. Contrary to most decision theories, actions will now not only be permissible or impermissible, but also classifiable as indeterminately permissible. This can be seen from the following definitions:[46]

---

[41] Rosanna Keefe, "Vagueness: Supervaluationism," *Philosophy Compass* 3, 2 (2008): 315-324, here 315.

[42] Achille C. Varzi, "Supervaluationism and Its Logics," *Mind* 116 (2007): 633-676, here 634.

[43] Definitions according to: Rinard, "A Decision Theory," 2.

[44] *Ibid.*, 2.

[45] *Ibid.*, 2.

[46] Definitions according to: Rinard, "A Decision Theory," 3.

Definition, Determinately Permissible Action. If some action A has the highest expected value (or ties for highest) according to every function in the agent's representor, then it's determinately true that A is permissible.

Definition, Determinately Impermissible Action. If some action A has a higher expected value according to every function in the agent's representor than some alternative action B, action B is determinately impermissible.

Analogous to the evaluation of the semantic value of statements in the supervaluationist framework, we also have the case of indeterminate permissibility:

Definition, Indeterminately Permissible. If some action A has the highest expected value according to some, but not all, functions in the representor, it is indeterminate whether A is permissible.

4.1 Supervaluationism and Sequential Decision-Making

In a next step, we can apply these definitions to Elga's problem. Assume, again, the great series of bets:

Bet A. If H is true, S loses $10. Otherwise S wins $15.

Bet B. If H is true, S wins $15. Otherwise S loses $10.

Assume, moreover, the following representor for our agent S as suggested by Elga: $\mathcal{P}(H) = [0.1, 0.8]$.[47]

Now, according to our supervaluationist decision theory, it is indeterminate whether accepting Bet A is rationally permissible. This is the case because it is permissible according to some, but not all probability functions in the agent's representor. For instance, according to the function that represents the precise value of $\Pr(H) = 0.2$, accepting Bet A has the highest expected value. According to $\Pr(H) = 0.7$, though, the agent should reject Bet A. Likewise, it is indeterminate whether rejecting Bet B is permissible. However, it is determinately impermissible to reject both bets since there is an alternative action with a higher expected value according to every function in the agent's representor.[48] This alternative is to accept both bets. Since this analysis yields the desired result, the supervaluationist decision theory does succeed in Elga's problem.

4.2 The Sequential Ellsberg Problem

Even though supervaluationism succeeds in Elga's problem, Bradley recently showed that it fails in another: If we interpret the Ellsberg Problem sequentially,

---

[47] Elga, "Subjective," 4.
[48] Rinard, "A Decision Theory," 6.

supervaluationism cannot rationalize ambiguity aversion. If this is the case, supervaluationism undermines a central motivation for introducing imprecise probabilities in the first place.[49] For this reason, Bradley's argument is a major challenge for the supervaluationist account.

Let's see how supervaluationism fails in the sequential Ellsberg Problem. As in the original problem, we again have an urn that contains 90 balls. From those balls, 30 balls are red and the remaining ones are either blue or yellow in some unknown proportion. Now, according to the sequential interpretation, an agent is offered two decision problems in quick succession:

> **Problem A.** The agent faces two choices:
>
> (I), which wins the agent a utility value of 1 if the ball drawn in the first round is red and nothing otherwise.
>
> (II), which wins the agent a utility value of 1 if the ball drawn in the first round is blue and nothing otherwise.
>
> **Problem B.** The agent faces two choices:
>
> (III), which wins the agent a utility value of 1 if the ball drawn in the second round is not blue and nothing otherwise.
>
> (IV), which wins the agent a utility value of 1 if the ball drawn in the second round is not red and nothing otherwise.[50]

As degree of belief for our agent, we assume that $\mathcal{P}(\text{blue}_i) = \mathcal{P}(\text{yellow}_i) = [0, 2/3]$ and $\Pr(\text{red}_i) = 1/3$ with $i = 1, 2$ referring to the current round of the decision problem.[51] With that, we can begin by analyzing Problem A.

> **Analysis, Problem A.** According to supervaluationism, it is indeterminate whether it is permissible to choose (I) over (II) and (II) over (I). It is indeterminate to choose (I) over (II) because it is permissible according to those functions in the representor that assign a probability less than 1/3 to drawing a blue ball and impermissible to the other functions in the representor. Likewise, it is permissible according to those functions in the representor to choose (II) over (I) that assign a probability greater than 1/3 to drawing a blue ball, but impermissible according to the other functions. The same line of reasoning applies to problem B:
>
> **Analysis, Problem B.** It is indeterminate whether it is permissible to choose (III) over (IV) and (IV) over (III).

[49] Bradley, „A Counterexample," 18.

[50] *Ibid.*, 24.

[51] *Ibid.*, 24.

So far, every option is indeterminately permissible. But what about the Ellsberg preferences, i.e. the sequence of (I) in round 1 and (IV) in round 2?

> **Analysis, Ellsberg Preferences (I) + (IV).** No function in the representor is such that it yields a preference of (I) over (II) and (IV) over (III). With that, it's determinately impermissible to have the Ellsberg preferences.[52]

This is a major drawback for a supervaluationist decision theory for imprecise probabilities. Initially, the Ellsberg Problem was used in order to motivate imprecise probabilities since there are decision rules for them that can rationalize ambiguity aversion. Still, we have now shown that supervaluationism fails to do so.[53] With that, supervaluationism seems to undermine a central motivation for introducing imprecise probabilities in the first place.

Concluding this section, I want to briefly discuss the idea of *rationalizing* the Ellsberg preferences. To begin with, it is clear that any decision rule that classifies them as irrational, or determinately impermissible in this case, fails.

However, it remains unclear how much permissibility is necessary in order to rationalize the Ellsberg preferences. One option is to classify them as determinately permissible. However, this could be a too strong requirement. That is, we are looking for the right amount of permissibility that we should be assigning to them.[54] In fact, it can be argued that classifying them as indeterminately permissible is just the right amount. To support this idea, I want to give the following line of reasoning:

On the one hand, we do not want to rationalize the Ellsberg preferences by classifying them as determinately permissible since there is no precise probability that allows this pattern of preferences. Thus, in order to respect expected utility theory, this amount of permissibility is too much.

On the other hand, as we have seen, we want to take seriously the aversion towards ambiguous gambles among rational agents. Therefore, it would be too strong to classify the Ellsberg preferences as determinately impermissible.

Luckily, supervaluationism allows for a third class of actions. Thus, in order to solve this conflict, I introduce a modified version of supervaluationism that classifies the Ellsberg preferences as indeterminately permissible as what I take to be the right amount of permissibility.

---

[52] *Ibid.*, 25.

[53] *Ibid.*, 25.

[54] The idea of the right amount permissibility regarding the Ellsberg preferences goes back to personal correspondence with Dr. Seamus Bradley.

5. Modified Supervaluationism

In this section, I introduce a modified supervaluationist framework which has to meet two objectives: First, it has to rationalize the Ellsberg preferences in its diachronic version. Secondly, it must not rationalize the rejection of both bets in Elga's problem.

The starting point for the modified supervaluationist framework is to take literally the interpretation of the agent's representor as a credal committee. In this committee, I construe every member as a voter that votes for propositions or actions in decision problems. As voting method, I apply relative majority voting. That is, the alternative that accumulates the most votes wins.[55] Based on this idea, I will now modify the central concepts of supervaluationism as well as its decision rule.

According to the original supervaluationist account, a proposition is *determinately true* if it is true according to all admissible precisifications. In the following, I replace this notion by propositions being *predominantly true*. This can be defined as follows:

> **Definition, Predominantly True.** A proposition is predominantly true if it is true according to a relative majority of precisifications.

Consider the following example: Assume, that there are ten admissible precisifications for the predicate *tall*. According to one of those the threshold for being tall is 170cm, according to two precisifications it is 175cm, according to three it is 185, and according to four precisifications the threshold lies at 180cm. In this case, it is predominantly true that someone who is at least 180cm tall is tall.

In a next step, we apply this idea to imprecise probabilities by assuming that every probability function in the representor is represented by a member of a voting committee. With that, we can very naturally derive a novel decision rule for imprecise probabilities. We begin by defining *predominantly permissible* and *impermissible* actions:

> **Definition, Predominantly Permissible Action.** An action A is predominantly permissible if a relative majority of members in the representor vote for it.

> **Definition, Predominantly Impermissible Action.** An action A is predominantly impermissible iff there is a relative majority of members in the representor that vote for an alternative action B.

---

[55] Joachim Behnke, Florian Grotz, and Christof Hartmann, *Wahlen und Wahlsysteme* (Oldenbourg: De Gruyter, 2016), 8.

It is important to note that an available action does only count as impermissible in case there is an alternative that receives a majority of votes. If this is not the case, we have a case of indeterminacy:

Definition, Determinately Permissible Action. If there is no relative majority in the representor for any action, every action is indeterminately permissible.

Finally, it has to be defined how the members of the committees do, in fact, vote:

Definition, Vote for an Action. A member in the representor, representing a probability function, votes for the action with the highest expected utility. If all actions bear the same expected utility according to the probability function, this member refrains from voting.

Consider the following example:

Example, Vote for an Action. Let action U be: buy an umbrella; action (I): buy ice cream. Furthermore, let H be: It will rain in an hour with $\mathcal{P}(H) = [0.1, 0.6]$.

	H	¬H
U	1	-1
I	-1	1

Figure 2. Payoffs, Vote for an Action.

The members in the representor which represent $\mathcal{P}(H) = [0.1, 0.5)$ vote for (I) (80%). The members which represent $\mathcal{P}(H) = (0.5, 0.6]$ vote for U (20%). With that, (I) is predominantly permissible and U is predominantly impermissible.

5.1 Modified Supervaluationism and the Sequential Ellsberg Problem

In the final part of this text, I will first apply the modified supervaluationist framework to the sequential Ellsberg Problem, and, subsequently, to Elga's problem.

The agent is facing again problem A and B where she has to choose between (I) and (II) as well as (III) and (IV) respectively in two rounds. Moreover, we have as imprecise degrees of belief: $\mathcal{P}(\text{blue}) = \mathcal{P}(\text{yellow}) = [0, 2/3]$ and $Pr(\text{red}) = 1/3$. With that, we can analyze both problems as follows:

Analysis, Problem A. The members in the voting committee that represent $\mathcal{P}(\text{blue})$ = [0, 1/3) vote for (I) since it yields a higher expected utility than voting for (II).

> The ones representing $\mathcal{P}$(blue) = (1/3, 2/3] , in turn, vote for option (II). The member that represents the function with Pr(blue) = 1/3 refrains from voting. With that, there is no majority for either option, and, thus, both options are indeterminately permissible. The same holds for problem B:

> **Analysis, Problem B.** $\mathcal{P}$(blue) = [0, 1/3) vote for (III) and $\mathcal{P}$(blue) = (1/3, 2/3] vote for (IV). Both of these actions are indeterminately permissible.

Finally, since this is the sequential Ellsberg Problem, we have to consider the sequence of (I) in round 1 and (IV) in round 2. That is, the Ellsberg preferences. Similar to the original supervaluationist framework, no member in the representor votes for this sequence. However, according to modified supervaluationism, that does not imply its impermissibility.

> **Analysis, Ellsberg preferences.** The members in the representor that represent the probability functions $\mathcal{P}$(blue) = (1/3, 2/3] vote for (II) + (IV) and the ones that represent $\mathcal{P}$(blue) = [0, 1/3) vote for (I) + (III). With that, the Ellsberg preferences are indeterminately permissible.

This, I count as an advantage of modified supervaluationism because it confirms with the demanded *right amount of permissibility* that should be assigned to the Ellsberg preferences by neither classifying them as determinately permissible nor determinately impermissible.

5.2 Supervaluationism and Elga's Problem

In a final step, I will apply modified supervaluationism to Elga's problem. Our agent faces once more the great the series of bets:

> **Bet A.** If H is true, S loses $10. Otherwise S wins $15.

> **Bet B.** If H is true, S wins $15. Otherwise S loses $10.

We assume, moreover, the imprecise degree of belief in H given by Elga: $\mathcal{P}$(H) = [0.1, 0.8]. Now, we can analyze both bets as follows:

> **Analysis, Bet A and B.** The members in the representor representing [0.1, 0.6) vote to accept Bet A. The members representing (0.6, 0.8] vote to refuse Bet A. With that, 71% vote to accept Bet A.

Likewise, Bet B: The members representing [0.1, 0.4) vote to refuse Bet B and the ones representing (0.4, 0.8] to accept Bet B. With that, 57% vote to accept Bet B.

Thus, accepting Bet A and accepting Bet B are predominantly permissible. For the given representor this is the correct result. Moreover, this can be shown for any possible representor:

> **Assertion:** For no imprecise degree of belief, it is possible to refuse both bets.

Proof. Assume, that it is possible to refuse both bets. In particular, the agent has then to refuse Bet A. In order to refuse Bet A, the agent has to have more functions in his representor with $Pr(H) > 60\%$. This is the case for imprecise degrees of belief with $\mathcal{P} \subseteq (0.6, 1]$. Bet B, in turn, is voted to be accepted for any imprecise degree of belief with $Pr(H) > 40\%$. That is: $\mathcal{P} \subseteq (0.4, 1]$. With that, every member that represents a function with $Pr(H) > 60\%$ votes for A to be refused but for Bet B to be accepted. Thus, it is not possible to refuse both bets at the same time.

Summary

This paper's objective was to provide a decision-theoretical framework based on imprecise probabilities that solves Elga's and Bradley's challenge. By modifying the supervaluationist account such a framework could be found. Modified supervaluationism construes the agent's representor as a voting committee that applies relative majority voting to evaluate the truth of statements and permissibility of actions. Moreover, it relies on a weaker notion of truth and permissibility than standard supervaluationism. Instead of *determinate* truth and permissibility, modified supervaluationism only requires *predominant* truth and permissibility. With that, it succeeds in both cases: It rationalizes the Ellsberg preferences to a reasonable extend and it does not rationalize rejecting both bets in Elga's problem.

RELIABILISTS SHOULD STILL FEAR THE DEMON

B.J.C. MADISON

ABSTRACT: In its most basic form, Simple Reliabilism states that: a belief is justified *iff* it is formed as the result of a reliable belief-forming process. But so-called New Evil Demon (NED) cases have been given as counterexamples. A common response has been to complicate reliabilism from its simplest form to accommodate the basic reliabilist position, while at the same time granting the force of NED intuitions. But what if despite initial appearances, Simple Reliabilism, without qualification, is compatible with the NED intuition? What we can call the Dispositionalist Response to the New Evil Demon problem is fascinating because it contends just that: Simple Reliabilism *is* fully compatible with the NED intuition. It is claimed that all we need to do to recognize their compatibility is appreciate that reliability is a dispositional property. In this paper I shall critically evaluate the Dispositionalist proposal.

KEYWORDS: epistemic justification, New Evil Demon Problem, reliabilism

1. Introduction

Reliabilism is a family of views about the nature of epistemic justification. In its most basic form, Simple ("Crude") Reliabilism states that: a belief is justified *iff* it is formed as the result of a reliable belief-forming process. To be a little more precise, Simple Reliabilism is understood to be an instance of *Lone World Reliabilism*, the class of views which hold that: "S's belief that *p* is justified *iff* S's belief that *p* is the output of a belief-formation process type that is reliable in *w*."[1] Specifically, Simple Reliabilism is also an instance of *Same World Reliabilism* which "identifies *w* as the world in which S forms the belief. A process's performance in a world determines its reliability there and thus, on Same World Reliabilism, determines the justificatory status of its outputs *there*."[2]

  Reliabilism faces a number of objections, and several have been conflated under the label of the 'Generality Problem.' Perhaps the most discussed aspect of the Generality Problem concerns what we can call Reliability-of-What: which process

---

[1] Matthew Frise, "The Reliability Problem For Reliabilism," *Philosophical Studies* 175 (2018): 923-945.

[2] Ibid.

type is the relevant one to assess for reliability? The charge is that the reliabilist cannot provide an answer in a principled and non ad hoc way.[3]

But other aspects of the Generality Problem are no less pressing. For instance, reliabilists need an answer to the issue of Reliability-When: what is the relevant temporal interval to assess reliability?[4] In addition to Reliability-of-What and Reliability-When, reliabilists also need to address the issue of Reliability-Where: which worlds must the process be reliable in?[5] As noted above, Simple Reliabilism answers this question straightforwardly by endorsing Same World Reliabilism: what matters is that the process used is reliable in the very same world that it is actually used in. But many philosophers, including those sympathetic to some kind of reliabilism generally, think that this cannot be correct, due to what has become known as the New Evil Demon problem.

So-called New Evil Demon (hereafter 'NED') cases aim to show that Simple Reliabilism is false, since subjects with false unreliably produced beliefs can nonetheless still have justified beliefs.[6] From this it is concluded that reliability is not necessary for epistemic justification. For example, take a subject here in our actual world who has intuitively justified beliefs: she justifiably believes that there is a cat before her on the basis of seeming to see a cat before her; she justifiably believes what she had for breakfast on the basis of seeming to remember what she had for breakfast, and so on. Now compare this subject with her counterpart in a possible world inhabited by an evil demon of great power, so great that he could ensure that the subjects of that world have beliefs about the external world that are false, based on subjectively indistinguishable non-veridical perceptual experiences. The demon also ensures that the subject's memory beliefs are false, based on subjectively indistinguishable non-veridical memory experiences, and so on.

The subjects in the demon world, we can suppose, have all the same non-factive mental states as their non-deceived counterparts, but their beliefs are by and large false, and so presumably unreliably produced. But do the subjects of that demon world have *justified* beliefs? If so, are their beliefs justified to the exact same extent as their counterparts in a non-demon world? What I shall call the New Evil

---

[3] e.g. Richard Feldman and Earl Conee, "The Generality Problem for Reliabilism," *Philosophical Studies* 89 (1998): 1-29.

[4] Matthew Frise calls this the Temporality Problem for reliabilism. See Frise, "Generality,". See also Brian Weatherson, "The Temporal Generality Problem," *Logos & Episteme* 3 (2012): 117-122.

[5] Scott Sturgeon distinguishes between the reliability-of-what vs. reliability-where questions that arise for process reliabilism. See Scott Sturgeon, *Matters of Mind* (London: Routledge, 2000), 96.

[6] E.g. Keith Lehrer and Stewart Cohen, "Justification, Truth, and Coherence," *Synthese* 55 (1983): 191-207; Stewart Cohen, "Justification and Truth," *Philosophical Studies* 46 (1984): 279-95.

Demon intuition is an evaluative judgment about *sameness*: our counterparts have all the same justified beliefs as we do, despite their falsity.

Three main responses to NED cases are common. What we can call The Committed Reliabilist response maintains that we should keep Simple Reliabilism but reject NED intuitions.[7] At the other end of the spectrum, what we can call The Committed Internalist response suggests that we keep NED intuitions and reject all forms of reliabilism, thus maintaining that reliability is not necessary for justification.[8] Finally, one might endorse what we can call an Irenic Solution. An Irenic Solution is one which aims to preserve some form of (modified) reliabilism, while at the same time, granting the force of the NED intuition.

Most Irenic Solutions do this by departing from Same World Reliabilism and adopting some form of *Modal Reliabilism*: "S's belief that *p* is justified *iff* S's belief that *p* is the output of a belief-formation process that is reliable in *W*."[9] W is taken to be a special domain of worlds that need not include the same world the belief forming process is actually used in. Complicating reliabilism to accommodate the NED intuition has taken many forms. Notable examples include Normal Worlds Reliabilism;[10] Weak and Strong Justification;[11] Indexical Reliabilism;[12] Home World

---

[7] For example, Bach, Brewer, Engel, and Sutton have, for different reasons, denied that our demon world counterparts' beliefs are justified. See Kent Bach, "A Rationale for Reliabilism," *The Monist* 68 (1985): 246-63; Bill Brewer, "Foundations of Perceptual Knowledge," *American Philosophical Quarterly* 34 (1997): 41-55; Mylan Engel, "Personal and Doxastic Justification," *Philosophical Studies* 67 (1992): 133-51; Jonathan Sutton, "Stick to What You Know," *Nous* 39 (2005): 359-96; Jonathan Sutton, *Without Justification* (Cambridge, MA: MIT University Press, 2007).

[8] For example, on the basis of New Evil Demon cases, prominent internalists have denied that reliability is necessary for justification, such as Lehrer and Cohen, "Justification,"; Cohen, "Justification and Truth;" Earl Conee and Richard Feldman, *Evidentialism* (New York: Oxford University Press, 2004); Ralph Wedgwood, "Internalism Explained," *Philosophy and Phenomenological Research* 65 (2002): 349-369; Michael Huemer, "Phenomenal Conservatism and the Internalist Intuition," *American Philosophical Quarterly* 43 (2006): 147-158. Some epistemic *externalists* have even drawn this conclusion in response to the New Evil Demon Problem; see for example Michael Bergmann, *Justification without Awareness* (New York: Oxford University Press, 2006).

[9] Frise, "Generality," 940.

[10] Alvin Goldman, *Epistemology and Cognition* (Cambridge, MA: Harvard University Press, 1986).

[11] Alvin Goldman, "Strong and Weak Justification," *Philosophical Perspectives* 2 (1988): 51-69.

[12] Ernest Sosa, "Reliabilism and Intellectual Virtue," in *Knowledge in Perspective: Selected Essays in Epistemology* (New York: Cambridge University Press, 1991), 131-145; Ernest Sosa, "Goldman's Reliabilism and Virtue Epistemology," *Philosophical Topics* 29 (2001): 383-400.

Reliabilism;[13] distinguishing Personal v. Doxastic Justification;[14] Bergmann's version of Proper Functionalism,[15] among others.

But if one wants to endorse a form of reliabilism in the face of NED cases, is this research program into Modal Reliabilism needed? What if despite initial appearances, even the most basic and simple form of process reliabilism, without qualification, is compatible with the NED intuition? All these complications and refinements of reliabilism, insofar as they aim to reconcile reliabilism with the NED, would then be redundant and unmotivated.

What we can call the Dispositionalist Response to the New Evil Demon problem is fascinating because it contends just that: Simple Reliabilism *is* fully compatible with the NED intuition (and so versions of Modal Reliabilism, whatever their other virtues, are unmotivated by the NED problem). The Dispositionalist Response to the NED purports to be an Irenic Solution, but interestingly one that does not think that we need to modify Simple Reliabilism to reconcile it with New Evil Demon cases. It is claimed that all we need to do to recognize their compatibility is appreciate that reliability is a dispositional property. It is to the details of this proposal that I shall now turn.

2. The Dispositionalist Response to the New Evil Demon Problem

According to what I shall call the Dispositionalist Response, recently advanced by Umut Baysan, counterparts *can* have justified beliefs in a demon world because the beliefs *are* produced by a reliable belief forming process. It is just that the reliable disposition is blocked / masked in the demon world.[16] We are told that the key to the proposal is to recognize that 'reliable' is a dispositional concept, and *reliability* is a dispositional property.[17] In general, something can have a disposition, despite not manifesting it. Recognizing this general truth is meant to help us realize that victims in NED cases can have beliefs that are the product of reliable belief forming methods, but the reliability is simply not manifested, as the resulting beliefs are false.

To see that in general, one can be the bearer of a dispositional property, without ever manifesting the disposition in question, consider an analogy with the

---

[13] Brad Majors and Sarah Sawyer, "The Epistemological Argument for Content Externalism," *Philosophical Perspectives* 19 (2005): 257-280.

[14] Bach, "Rationale;" Engel, "Personal."

[15] Bergmann, *Justification without Awareness*.

[16] Umut Baysan, "A New Response to the New Evil Demon Problem," *Logos & Episteme* VIII (2017): 41-45.

[17] Baysan, "A New Response," 43.

property of fragility.[18] A vase can be fragile even if it never breaks. This could be simply because its owner is very careful and the fragile vase is never struck. But it is also true that something can be fragile even if it does not break when struck, even if struck repeatedly. A glass-faced smartphone is very fragile. It often breaks when dropped. But sometimes one is very lucky, and the face does not smash despite being dropped, even repeatedly. The phone need be no less fragile as a result.

Baysan suggests something similar with regard to reliable belief forming processes: a belief forming process may be disposed to produce true beliefs, but for whatever reason, at every attempt it may fail to do so. A belief forming process might instantiate the property of reliability without ever manifesting it. Baysan argues the following is a possible state of affairs: "(iv) *a* is a reliable belief-forming process; *a* is exercised; *a* doesn't produce true beliefs; this happens systematically."[19]

Baysan argues that if (iv) is possible, then it follows that one can be a Simple Reliabilist about epistemic justification and still hold that victims of the NED have beliefs that are justified: despite their systematic falsity, the beliefs are still produced by reliable belief forming processes – it is just that the reliability is not manifested in the demon world. If this is correct, we have an Irenic Solution that shows the compatibility of Simple Reliabilism and the NED intuition, without the need to resort to a form of Modal Reliabilism. I shall now critically evaluate the Dispositionalist proposal.

3. In What Sense Are Beliefs Produced by Reliable Faculties in Demon Worlds?

The key claim of the Dispositionalist Response is that NED victims can have beliefs that are produced by cognitive faculties that are reliable, but since reliability is a dispositional property, it can be instantiated without being manifested – and that is just what is going on in the NED case. But in what sense, if any, are beliefs produced by reliable faculties in demon worlds? Crucial here is what reliability amounts to. When applied to belief forming methods, "reliable" means, at the very least, that the method used is more likely to result in true beliefs than not. So the relevant sense of reliability here means reliably truth-conducive. But in what sense is a method more likely to produce true instead of false beliefs? At least two options initially present themselves, a frequency sense and a modal sense of reliability, both of which I shall argue are inadequate for the Dispositionalist Response to the NED.

First, reliability might be understood as the *frequency* of true beliefs generated by a process. Reliability in this sense is determined by the actual track

---

[18] Cf. Baysan, "A New Response," 44.
[19] Baysan, "A New Response," 45.

record of the belief forming process.[20] If the frequency results in a favorable ratio of true to false beliefs, then the process is a reliable one. New Evil Demon cases can be set up, however, so as to ensure a track record of falsity. So the belief forming processes employed in demon worlds are not reliable in this sense.

Furthermore, an actual high proportion of true beliefs to false beliefs is neither necessary nor sufficient for reliability. It is not necessary: suppose that a subject undergoes an operation that results in them having the perfect ability to add the sum of any numbers. Before they can exercise this incredible new ability, they die due to surgical complications. Nevertheless, for a time they possessed a reliable mathematical disposition. This is not because of anything to do with frequency, however, since the ability was never used. Rather, it seems that they possessed the reliable ability because if they *had* performed any possible addition, they *would* have always been correct.

Neither is an actual high proportion of true beliefs over false beliefs sufficient for reliability. This is because a process can be "luckily" truth conducive. For example, a student might be incredibly lucky in guessing correctly more often than not on a pop quiz. Even if such guessing resulted in a perfect score on the quiz, this would not make guessing a reliable process. A natural explanation of this is that while the student actually got all the correct answers, if they had used this same method more generally, it would presumably have resulted in many incorrect answers.

These considerations motivate a *modal* conception of reliability. According to a modal sense of reliability, a belief forming process is reliable just in case it *would* yield a favorable ratio of true to false beliefs (in some relevant range of circumstances; more on this later). But the NED case can be set up such that the beliefs of the victims of the New Evil Demon would not be reliably produced given a modal conception of reliability either, but according to the NED intuition, their beliefs would nonetheless still be justified.

To see this, suppose that, as a contingent matter of fact, all the worlds our counterparts occupy are also demon worlds. In that case, there would be no worlds where our counterparts exercise their cognitive faculties and the demon is not present. Now suppose that the demon is also not only extremely powerful, but it is also extremely evil, such that necessarily it will radically deceive its victims. It follows that the demon could ensure the falsity of these subjects' beliefs in all nearby worlds, if not all worlds, in which they exercise their cognitive faculties. Given all this, *if* the victims' belief forming processes were used, they would always result in

---

[20] And what temporal period is the relevant one to assess how successful the track record is? It is exceedingly difficult to say. See Frise, "Generality."

falsity (the demon sees to that). And yet, according to the NED intuition, our counterparts' same beliefs are just as justified as ours are.

So if not a frequency or modal conception of reliability, in what other sense might subjects in a NED case have beliefs produced by reliable faculties? Perhaps there is a different but related *dispositional* sense of reliability? Baysan writes, "A reliable belief-forming process is disposed to produce true beliefs."[21] And how should we understand what it is to be disposed to produce true beliefs? Baysan contends that "[…] the reliability of a belief-forming process is manifested in the fact that beliefs that are formed as a result of that process are mostly true, again, *in the right circumstances*."[22] (emphasis added)

To see how being disposed to result in true beliefs might work, consider again fragility, a paradigm case of a disposition. Take the case of the vase again: what makes it the case that the vase is fragile? Not that it broke, or will break, or that it would break under *any* possible conditions, but presumably that it would break, *in the right circumstances*, and if submitted to the right kind of stimulus conditions. It is true that a vase can be fragile even if it never breaks, and even if it does not break when struck. I grant that even if wrapped in packing material, and so in those circumstances it cannot break, that it might still nonetheless be fragile. But this, I submit, is because it is at least *possible* for the vase to break (including in whatever counts as the right circumstances).

If, on the other hand, it is now metaphysically impossible for the vase to break, that is, if there are now no worlds where it can break, I submit that the vase is no longer fragile.[23] As a vivid illustration: if God exists and promises to protect the vase, and never let it break, it would then be metaphysically impossible for it to break. Given God's omnipresence, omnipotence, and that he always keeps His promises, there are no worlds where the vase breaks. In such a case, I suggest that the vase has lost a disposition; the disposition is not just 'super-masked.' Rather, the vase is no longer fragile.

---

[21] Baysan, "A New Response," 44.

[22] Baysan, "A New Response," 43-44.

[23] While most will surely agree that something can have a disposition that is never actually manifested, is it really a necessary condition of having a disposition that it at least has metaphysically possible manifestation conditions? On the possibility of dispositions with impossible manifestation conditions generally, see C.S. Jenkins and Daniel Nolan, "Disposition Impossible," *Nous* 46 (2012): 732-753; see also Jack Spencer, "Able to Do the Impossible," *Mind* 126 (2017): 466-497. Note that several of the examples in Jenkins and Nolan turn on logical or nomic necessity, rather than metaphysical necessity, which is at issue here. For critical discussion of Jenkins and Nolan, see Barbara Vetter, *Potentiality: From Dispositions to Modality* (Oxford: Oxford University Press, 2015), chapter 7.

Similarly, a belief forming process may be disposed to produce true beliefs, but for whatever reason at any attempt, it may fail to do so. But it must at least be *possible* for the process to produce true beliefs if it is disposed to be reliable. Even if not physically possible, it must be at least metaphysically possible. But in the version of the NED case I am considering where the demon happens to exist in all the worlds our counterparts do, given the Demon's omnipresence, his vast power, and his unwavering evil intentions to deceive, there are no worlds where the belief-forming process yields true beliefs – the demon sees to that.

I agree with Baysan that the following is a possible state of affairs: "(iv) *a* is a reliable belief-forming process; *a* is exercised; *a* doesn't produce true beliefs; this happens systematically."[24] But it does not follow from the truth of (iv) alone that a belief forming process remains reliable even if it *never* produces true beliefs; and even stronger, even if it is *impossible* for it to produce true beliefs. Like the vase no longer being fragile if it is impossible for it to break, I submit that if it is impossible for a process to yield true beliefs, it is not reliable either.

One might object that one has the lingering intuition that a vase can be fragile, even if it is now impossible that it breaks. If so, might a process be reliable even if it is impossible that it produces true beliefs? Even if this were granted in this particular case of fragility, there is a key disanalogy between vases and belief forming processes: a vase's fragility is presumably determined in part by its *intrinsic* properties (e.g. its microstructure), and if this is not changed, then the vase is still fragile, even if it is now metaphysically impossible that it breaks. After all, it is a fairly widely held view that dispositions are fixed by a thing's intrinsic features.[25]

But a belief being produced by a token process, one that is of a reliable type, are wholly *extrinsic* features both of the belief and the process. If two subjects are exact intrinsic duplicates, and have the same belief forming processes, these processes need not be equally reliable – for instance, the subjects might be in radically different environments, as the NED cases make vivid. Whether a process produces true beliefs is partly determined relationally. This means that whether a process is reliable necessarily depends on the environment in which it is used. Baysan seems to implicitly recognize this, as reliable belief-forming processes are described as tending to produce true beliefs, *in the right circumstances*. What would the right circumstances be, in the case of reliable belief forming processes? Presumably, at the very least, ones where there is no deceiving demon.

But accommodating this point is inconsistent with Simple Reliabilism as this new proposal really amounts to a closet form of *Modal Reliabilism*, which demands

---

[24] Baysan, "A New Response," 45.
[25] Cf. David Lewis, "Finkish Dispositions," *The Philosophical Quarterly* 47 (1997): 143-158.

not the actual reliability of the process used, but only reliability in a special domain of worlds – namely ones that lack, amongst other things, deceiving demons. This is to relativize reliabilism in just the way that Baysan wants to avoid. The upshot is that reliabilism still needs an answer to not only questions of Reliability-of-*What*, and Reliability-*When*, but also Reliability-*Where*. That is, reliabilists still need to type-individuate environments; not only the type of physical environment, but also the relevant range of possible worlds the process needs to be truth-conducive in.

In short, reliability is determined not only by the relevant belief forming process, but also the relevant environment, and the Dispositionalist response to the NED problem overlooks this.

4. A Value Problem for the Dispositionalist Response

Besides under-appreciating the issue of reliability-*where*, another problem for the Dispositionalist Response to the NED problem is accounting for the value of epistemic justification. By maintaining that beliefs can be fully justified, despite being actually all false, we lose a main benefit of Simple Reliabilism. Namely, traditionally Same World Reliabilism can offer a straightforward account of the value of justification: the value of justification is instrumental as a means to truth.

As Laurence Bonjour asks, "Why should we, as cognitive beings, *care* whether our beliefs are epistemically justified? Why is such justification something to be sought and valued?"[26] Bonjour thinks that the answer to these questions is obvious. He writes,

> Once the question is posed this way, the following answer seems obviously correct, at least in first approximation. What makes us cognitive beings at all is our capacity for belief, and the goal of our distinctively cognitive endeavors is *truth*: We want our belief to correctly and accurately depict the world. […] The basic role of justification is that of a *means* to truth, a more directly attainable mediating link between our subjective starting point and our objective goal. […] If epistemic justification were not conducive to truth in this way, if finding epistemically justified belief did not substantially increase the likelihood of finding true ones, then epistemic justification would be irrelevant to our main cognitive goal and of dubious worth. It is only if we have some reason for thinking that epistemic justification constitutes a path to truth that we as cognitive beings have any motive for preferring epistemically justified beliefs to epistemically unjustified ones. Epistemic justification is therefore in the final analysis only an instrumental value, not an intrinsic one.[27]

---

[26] Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge, Mass.: Harvard University Press, 1985), 7.

[27] Bonjour, *The Structure of Empirical Knowledge*, 7-8; see also p. 157 for another clear expression

As is made clear above, Bonjour thinks that the value of justification is instrumental to the end of truth.[28] Simple Reliabilism can easily account for the value of justification in the following way: since truth is of value, and given Simple Reliabilism, beliefs formed by reliable processes are more likely to actually be true.

But according to the Dispositionalist Response, beliefs can be fully justified, even though all and always false. So one is left wondering: what is so great about epistemic justification? If it is also consistent with being the product of a reliable belief forming process that output belief tokens can all and always be false, then what is also so great about reliability? By having no answers to the questions of why justification and reliability are valuable, the Dispositionalist Response to the New Evil Demon loses one of the key advantages of Same World Reliabilism. This is of course not a decisive reason to reject the Dispositionalist Response, but it is a major strike against it. Having a clear and straightforward account of the value of justification would be a clear advantage over epistemically internalist views which reject reliability as necessary for justification.

In summary, I have argued that we have no adequate new solution to the problem of reconciling Simple Reliabilism with the NED intuition. The traditional options therefore remain. One can be a Committed Reliabilist and reject the NED intuition. One can be a Committed Internalist and reject reliability as a necessary condition on justification. Alternatively, one ought to seek an Irenic Solution by complicating the basic reliabilist proposal and developing some form of Modal Reliabilism consistent with NED cases.[29]

---

of this position.

[28] For a recent defense of the idea that the value of epistemic justification is not exhausted by its instrumental value, that justification is also valuable for its own sake, and that therefore truth value monism is false (i.e. that there is more of final epistemic value than mere true belief), see B.J.C. Madison, "Epistemic Value and the New Evil Demon," *Pacific Philosophical Quarterly* 98 (2017): 89-107.

[29] Thanks to an audience at New York University Abu Dhabi. Thanks also to Rhiannon James for helpful written comments on earlier drafts of this paper.

ALLEGED COUNTEREXAMPLES TO UNIQUENESS

Ryan ROSS

ABSTRACT: Kopec and Titelbaum collect five alleged counterexamples to Uniqueness, the thesis that it is impossible for agents who have the same total evidence to be ideally rational in having different doxastic attitudes toward the same proposition. I argue that four of the alleged counterexamples fail and that Uniqueness should be slightly modified to accommodate the fifth example.

KEYWORDS: uniqueness, permissivism, impermissivism, evidence, rationality

Introduction

There is now a standing debate about how many doxastic attitudes can be rational given a single body of evidence. This disagreement is about the thesis that

> **Uniqueness**: It is impossible for agents who have the same total evidence to be ideally rational in having different doxastic attitudes toward the same proposition.

I will say that a *permissivist* is someone who denies Uniqueness, whereas an *impermissivist* is someone who accepts Uniqueness.[1]

Uniqueness makes the claim that a certain state of affairs cannot obtain. A good way to object to such a claim is to present possible examples in which it appears plausible that the state of affairs does obtain. This is the strategy pursued by Kopec and Titelbaum.[2] They collect five alleged counterexamples to Uniqueness from the literature. I will argue that only one of these examples is problematic for Uniqueness and that even this example can be dealt with by making a slight modification to Uniqueness.

---

[1] The term *permissivist* was introduced by Roger White, "Epistemic Permissiveness," *Philosophical Perspectives* 19 (2005): 445– 459. The term *impermissivist* was introduced by Sophie Horowitz, "Immoderately Rational," *Philosophical Studies* 167 (2014): 41-56.

[2] Matthew Kopec and Michael Titelbaum, "The Uniqueness Thesis," *Philosophy Compass* 11 (2016): 189–200.

Ryan Ross

Clarifications about Uniqueness

Kopec and Titelbaum distinguish between several theses that might be labelled "Uniqueness."[3] The version of Uniqueness I defend (i) is interpersonal, (ii) applies to all doxastic attitudes, (iii) allows for rational dilemmas, and (iv) applies only to ideal rationality.

First, my version of Uniqueness makes an *inter*personal claim, not an *intra*personal claim. Some statements of Uniqueness are ambiguous between these two interpretations.[4] My version states that if agent A is ideally rational in having doxastic attitude D toward proposition P when A's total evidence is E, then no other doxastic attitude toward P is ideally rational for *anyone* to have when their total evidence is E. For example, my version of Uniqueness entails that, if believing that P is ideally rational given total evidence E, then *anyone* who has total evidence E but doesn't believe that P is not ideally rational.

Second, my version of Uniqueness makes a claim that applies to all doxastic attitudes; this includes believing, disbelieving, suspending judgment, and having credences. For example, according to Uniqueness, if two agents have the same total evidence, then it's impossible for one of them to be ideally rational in suspending judgment about whether P while the other is ideally rational in believing that P. Likewise, if two agents have the same total evidence, it's impossible for one of them to be ideally rational in having credence 0.6 in P while the other is ideally rational in having credence 0.7 in P.

Third, my version of Uniqueness allows for rational dilemmas (cases in which there is no rational response to one's evidence). In other words, my version does not presuppose that there is always at least one rational response to one's evidence. As Kopec and Titelbaum note, some versions of Uniqueness say that, for any body of evidence, there is *at least one* doxastic attitude that is rational; meanwhile, other versions say that, for any body of evidence, there is *at most one* doxastic attitude that is rational.[5] The latter, but not the former, allows for rational dilemmas. My version is in accord with the latter.

---

[3] Kopec and Titelbaum, "The Uniqueness Thesis," 190-2.

[4] E.g., White's statement of Uniqueness in "Epistemic Permissiveness" is ambiguous in this way. This point is made by Thomas Kelly, "Evidence Can Be Permissive," in *Contemporary Debates in Epistemology*, 2nd ed., eds. Matthias Steup, John Turri, and Ernest Sosa (Hoboken, NJ: John Wiley & Sons, 2013), 298-311; and Michael Titelbaum and Matthew Kopec, "Plausible Permissivism" (manuscript).

[5] Kopec and Titelbaum, "The Uniqueness Thesis," 190-1. An example of the former version is White, "Epistemic Permissiveness." An example of the latter version is Richard Feldman, "Reasonable Religious Disagreements," in *Philosophers without God: Meditations on Atheism and*

Fourth, my version of Uniqueness makes a claim about ideal rationality, not about subideal rationality.[6] My version of Uniqueness allows that two people who have the same total evidence may be rational *to some degree* in disagreeing; they might even be *equally* rational. However, they cannot both be *ideally* rational. Within the category of rationality simpliciter, there are two subcategories: ideal rationality and subideal rationality. Ideal rationality is rationality without epistemic mistakes; that is, without making any mistakes about what one's evidence supports. One is ideally rational in having a given doxastic attitude iff one's total evidence supports having that attitude.

Meanwhile, subideal rationality is a form of rationality that is consistent with making mistakes about what one's evidence supports. Let's look at a few examples in which it seems that there is an agent whose doxastic attitude is subideally rational. My first example is as follows:

> Jones and Smith have the same complex body of evidence. Jones concludes that P. Smith concludes that ~P. Jones concluded that P because he made the subtle mistake of putting too much trust in the testimony of Expert 1. Smith concluded that ~P because she made the subtle mistake of putting too much trust in the testimony of Expert 2.

Next, here is an example inspired by Cohen:

> Jones concludes that P on the basis of his total evidence E. Almost all intelligent people would agree that E supports believing that P. However, for subtle reasons that only a super genius could discern, E actually supports believing that ~P.[7]

The last example I will mention is based on a case that Podgorski discusses:

> Jones and Smith have the same evidence concerning whether P and are listening to the same radio program. The radio program mentions something that bears on whether P. Smith takes the new evidence from the radio program into account and increases her credence in P, which is what her evidence now supports. Meanwhile, Jones heard the same news from the radio; however, just afterward, his apartment caught on fire. Instead of increasing his credence in P (as his new evidence

---

the Secular Life, ed. Louise Antony (Oxford: Oxford University Press, 2007), 194-214.

[6] In this way, I follow Roger White's revised statement of Uniqueness, stated in terms of "fully rational" doxastic attitudes, from his "Evidence Cannot be Permissive," in *Contemporary Debates in Epistemology*, 2nd ed., eds. Matthias Steup, John Turri, and Ernest Sosa (Hoboken, NJ: John Wiley & Sons, 2013), 312-23. White's earlier formulation of Uniqueness from "Epistemic Permissiveness" is not explicitly restricted to full or ideal rationality.

[7] Stewart Cohen, "Defense of the (Almost) Equal Weight View," in *The Epistemology of Disagreement: New Essays*, eds. David Christensen and Jennifer Lackey (Oxford: Oxford University Press, 2013), 98-119.

requires), Jones ran for his life.[8]

What is the correct epistemic evaluation of the conclusions reached by Jones in these cases? One option is to say that, in each case, Jones made mistakes, so his doxastic attitudes are not rational. This is to assume a perfectionist view of epistemic rationality such that epistemic rationality is inconsistent with epistemic mistakes. The other option, which I adopt in this paper, is to take an imperfectionist view of epistemic rationality, according to which a doxastic attitude can be rational, despite being based on a mistake. This view allows that Jones' doxastic attitudes are *subideally* rational, but not *ideally* rational. For a doxastic attitude to be subideally rational is for it to fall short of ideal rationality but still not be so bad that it counts as irrational. According to this view, Jones can be subideally rational (because he approximates rational perfection closely enough), but not ideally rational (because he made mistakes).

Jury Example

Having made these clarifications, we can now consider the alleged counterexamples that Kopec and Titelbaum discuss. They draw their first example from something Rosen says:

> It should be obvious that reasonable people can disagree, even when confronted with the same body of evidence. When a jury or a court is divided in a difficult case, the mere fact of disagreement does not mean that someone is being unreasonable.[9]

The alleged counterexample in question goes like this:

> **Jury example**: The members of a jury hear the same evidence presented to the court. One juror concludes that the defendant is guilty. Another juror concludes that the defendant is not guilty. Both jurors are ideally rational in their respective beliefs.

The impermissivist should respond by noting that if the jury example is intended to be a realistic example of jury deliberation, then it won't be plausible that the jurors have the same evidence. Even after hearing closing arguments, the members of the jury may remember different things. Having different memories entails having different evidence. But let's assume that all the jurors remember every aspect of the case. Nevertheless, before the trial even begins, the jurors will come in with different memories based on having had different experiences. Because of such differences, the jurors may come in with different evidence of relevance to the case.

---

[8] Abelard Podgorski, "Dynamic Permissivism," *Philosophical Studies* 173 (2016): 1923-1939.

[9] Gideon Rosen, "Nominalism, Naturalism, Epistemic Relativism," *Philosophical Perspectives* 15 (2001): 69–91.

Therefore, in addition to the trial evidence, the jurors have the non-trial evidence that they came into the courtroom with. In other words, the *trial* evidence does not exhaust the juror's *total* evidence. So, it's implausible that the members of an actual jury have the same total evidence. For example, suppose that the case against the defendant has an epistemic property, EP. One juror might have non-trial evidence that EP is truth-conducive, whereas another juror lacks this evidence. More concretely, one juror might have non-trial evidence that people who act in a specific way are giving dishonest testimony, whereas another juror lacks this evidence. Thus, if this example is supposed to be realistic, it fails as a counterexample.

Can we idealize the example into a convincing counterexample? Let's stipulate that the jurors have the exact same evidence. But even the best version of this example doesn't give us a knockdown argument against Uniqueness. What the permissivist apparently wants to say is this:

1. Even if two agents have the same evidence and disagree, there can still be something epistemically good about each of their respective beliefs.

2. This "something epistemically good" is ideally rational belief.

3. Therefore, even if two agents have the same evidence and disagree, each of them can be ideally rational in their respective beliefs.

The impermissivist can grant the first claim, but deny that the "something epistemically good" amounts to ideally rational belief. The impermissivist can say, instead, that the "something epistemically good" is subideally rational belief. At this point, the permissivist will need to make the case that the type of rational belief is ideal, not just subideal. If nothing else, the example in question doesn't show this. Some other kind of argument is needed.

This response also shows that it's not the case that impermissivists lack the resources to give due credit to the jurors whose responses were not ideally rational. Suppose these jurors acted in good faith. Suppose they did their best. Suppose there are many intelligent people who agree with these jurors. The thought goes: "Shouldn't such responses count as rational? It's not like they concluded that aliens or witches committed the crime." The impermissivist can answer affirmatively: Such responses may well be rational (subideally rational), but not ideally rational.

In addition to the jury example, Rosen also uses the example of paleontologists who disagree about what killed the dinosaurs.[10] This example is different from the jury example at least inasmuch as it involves expertise rather than the sort of everyday epistemic abilities that we hope for jurors to have. And we don't need to limit ourselves to just paleontology. Scientific disagreement extends to all areas of

---

[10] Rosen, "Nominalism," 77.

science. Do we really want to say that so many scientists in so many fields have been failing to respond to their evidence in a rational way?

My response is basically the same. First, if we're being realistic, we can't rule out the possibility that the scientists in question had some different evidence despite having a great deal of overlap in their evidence. Next, even if we say they had the same evidence, there is little pressure to say that all the parties to these scientific debates were being ideally rational rather than subideally rational. Either way, we don't have a counterexample to Uniqueness.

Materialism Example

Next, let's consider an example that Kopec and Titelbaum borrow from Decker.[11] Here is their paraphrase:

> **Materialism example**: [S]uppose two initially identical agents spontaneously materialize, one on Earth and the other on Twin Earth. Both agents encounter perceptually identical worlds, and therefore are guaranteed to have all the same evidence. But further suppose that while the Earthling comes to form a strong conviction in a mind independent world composed of material objects, the Twin Earthling becomes convinced of a Berkelean world composed entirely of either minds or ideas in minds. So let P be the proposition that "The world is composed of physical objects." The Earthling [rationally] believes P, while the Twin Earthling [rationally] believes not P, and both have the very same evidence.[12]

We should begin by asking: What evidence do the agents have for and against materialism? Materialism and idealism (if justified at all) must be justified by philosophical arguments. If the agents are aware of different arguments, then they have different evidence. So, the example must stipulate that they are aware of the same arguments for materialism and idealism. The difference is that they disagree about which arguments are sound: One agent thinks that at least one argument for materialism is sound, while the other thinks that at least one argument for idealism is sound. In this way, the agents in question are similar to metaphysicians in the actual world who disagree about whether materialism or idealism is true. The main difference is that our actual metaphysicians are part of a wider philosophical community.

As a result, this example is more pregnant than it may initially appear. If these hypothetical metaphysicians are analogous to our actual metaphysicians, then this example requires saying that the materialism-idealism debate in the actual world is a permissive case. Is there anything special about the materialism-idealism debate

---

[11] Jason Decker, "Disagreement, Evidence, and Agnosticism," *Synthese* 187 (2012): 753–83.

[12] Kopec and Titelbaum, "The Uniqueness Thesis," 196.

that distinguishes it from most other long-standing metaphysical debates? It seems not. So, the implication is that many longstanding metaphysical debates are permissive. And are metaphysical debates so different from other philosophical debates? If not, we can generalize from metaphysics to the rest of philosophy and conclude that most long-standing philosophical debates are permissive.

Thus, if the materialism example is possible, then Uniqueness is not only false, but believing it might entail having a radically mistaken view about the rationality of philosophical disagreement. How should impermissivists respond? To see, let's note that the reasoning behind the materialism example seems to be similar to the reasoning behind the jury example. Do we really want to say that one of these agents is being *irrational*? After all, their reasoning may well be commendable in several ways. They may have tried their best. They may have come up with arguments that are by no means crazy or incoherent. They may even make use of arguments that actual metaphysicians find plausible and accept. So, why not say that the agents are both rational? My response to the materialism example is the same as my response to the jury example: Even if we make the (not-so-realistic) assumption that the agents have the same evidence, there is little pressure to say that they are both ideally rational rather than subideally rational.

Community Example

Let's consider a third example. This one is drawn from Schoenfield:

> **Community example**: You have grown up in a religious community and believe in the existence of God. You have been given all sorts of arguments and reasons for this belief which you have thought about at great length. You then learn that you only have the religious beliefs that you do, and only find the reasoning that you engaged in convincing, because of the influence of this community. If you had grown up elsewhere, you would have, on the basis of the same body of evidence, rejected those arguments and become an atheist.[13]

How should impermissivists respond? As with the jury example, to the extent that this case is realistic, it is not plausible that the agent would have the same evidence in both conditions. It's unlikely that a religious community would supply reasons in favor of atheism that are as good as one would hear in an atheistic community; and vice versa. But let's agree to work with an idealized case in which the religious community in question and the atheistic community in question do provide the same evidence for and against the existence of God. You, then, learn that, if you had grown up in a different community, then you would have formed

---

[13] Miriam Schoenfield, "Permission to Believe: Why Permissivism is True and What it Tells us about Irrelevant Influences on Belief," *Noûs* 48 (2014): 193–218.

different beliefs based on the same evidence. As a straight counterexample, this example seems to fail badly. The intuitive reaction is that there is something worrisome about the belief in question. One worries, "Should I continue to hold this belief while knowing that my having this belief is influenced by which community I grew up in, which is irrelevant to the truth of the matter?" This much shows that the community example fails as a straight counterexample, since counterexamples are supposed to be intuitively plausible.

Schoenfield acknowledges as much and makes it the burden of her paper to show that this intuitive reaction is not right; rather, she thinks one can rationally keep such beliefs.[14] Schoenfield's goal is to argue that, *contrary to appearances*, people in cases like the community example can be rational in sticking to their beliefs.[15] This is because epistemic rationality supervenes on (at least) two things: one's total evidence and one's epistemic standards (roughly, one's way of evaluating evidence). On her view, growing up in a religious community can imbue one with religious epistemic standards, and growing up in an atheistic community can imbue one with atheistic epistemic standards. The same evidence, filtered through these different standards, can lead to different rational responses to the evidence.[16] And we need not change our minds to accommodate others' epistemic standards; rather, we can stick to our beliefs as long as we live up to our own epistemic standards.[17] Therefore, people in cases like the community example can rationally stick to their beliefs.

Importantly, however, this line of reasoning wouldn't work as an objection to Uniqueness.[18] Such an objection would have to assert the premise that

> There are different epistemic standards that are all ideally rational to have

in support of the conclusion that

> Uniqueness is false.

However, such an argument would be question-begging without independent support of the premise, since clearly no impermissivist should grant that premise.[19]

---

[14] Schoenfield, "Permission to Believe," 193.

[15] *Ibid*.

[16] Schoenfield, "Permission to Believe," 199-200.

[17] Adam Elga, "Lucky to be Rational," (manuscript) talks in terms of living up to one's standards.

[18] Again, Schoenfield never says this line of reasoning should be used as an objection to Uniqueness.

[19] The point I make is this: Impermissivists already deny that two agents who have the same total evidence can be ideally rational in having different *doxastic attitudes*. So, do we really think they will grant that two agents who have the same total evidence can be ideally rational in having different *epistemic standards*? Another objection is to challenge the idea that all we need to do is

Reasoning Room Example

A fourth example that Kopec and Titelbaum discuss is taken from another one of their co-authored works.[20] Here is the case:

> **Reasoning room example**: You are standing in a room with nine other people. Over time the group will be given a sequence of hypotheses to evaluate. Each person in the room currently possesses the same total evidence relevant to those hypotheses. But each person has different ways of reasoning about that evidence (and therefore different evidential standards). When you are given a hypothesis, you will reason about it in light of your evidence, and your reasoning will suggest either that the evidence supports belief in the hypothesis, or that the evidence supports belief in its negation. Each other person in the room will also engage in reasoning that will yield exactly one of these two results. This group has a well-established track record, and its judgments always fall in a very particular pattern: For each hypothesis, 9 people reach the same conclusion about which belief the evidence supports, while the remaining person concludes the opposite. Moreover, the majority opinion is always accurate, in the sense that whatever belief the majority takes to be supported always turns out to be true. Despite this precise coordination, it's unpredictable who will be the odd person out for any given hypothesis. The identity of the outlier jumps around the room, so that in the long run each agent is odd-person-out exactly 10% of the time. This means that each person in the room takes the evidence to support a belief that turns out to be true 90% of the time.[21]

The problem with this example is that it starts off with the assumption that the agents in the reasoning room have different ways of reasoning. In order to be a counterexample to Uniqueness, we must strengthen this assumption to say that more than one of these different ways of reasoning is ideally rational. (We won't have a counterexample if only one way of reasoning is ideally rational.) Ways of reasoning are important to the example, because Kopec and Titelbaum think that epistemic rationality supervenes on (at least) two things: one's total evidence and one's way of reasoning. On their view, the same evidence can be filtered through two different ways of reasoning in order to reach two different but rational doxastic attitudes toward the same proposition.[22] Thus, Kopec and Titelbaum have to argue from the premise that

---

live up to our epistemic standards. For this objection, see White, "Epistemic Permissiveness," 451-2; and Feldman "Reasonable Religious Disagreements," 149.

[20] Titelbaum and Kopec, "Plausible Permissivism." The shorter, published version of which is Michael Titelbaum and Matthew Kopec "When Rational Reasoners Reason Differently," in *Reasoning: New Essays on Theoretical and Practical Thinking*, eds. Magdalena Balcerak-Jackson and Brendan Balcerak-Jackson (Oxford: Oxford University Press, 2019), 205-31.

[21] Kopec and Titelbaum, "The Uniqueness Thesis," 196.

[22] In this way, their view is like Schoenfield's, "Permission to Believe."

Ryan Ross

> There is more than one way of reasoning that is ideally rational

to the conclusion that

> Uniqueness is false.

But just as impermissivists should deny that there are multiple epistemic standards that are ideally rational, impermissivists should also deny that there are multiple ways of reasoning that are all ideally rational. To assume this premise without further argument begs the question.

Self-fulfilling Example

The last example that Kopec and Titelbaum discuss aims to make trouble for Uniqueness by applying it to self-fulfilling beliefs.[23] The example is as follows:

> **Self-fulfilling example**: God appears to you, and (rationally) convinces you of her omnipotence and omniscience. For example, she's able to read your mind perfectly, predict all of your actions, grant all your wishes, and change the weather at will. One day, God makes the following proposal: If you believe that it will rain in Canberra tomorrow, then she will make sure it rains in Canberra tomorrow. But if you believe it won't rain tomorrow, then she'll make sure it doesn't. She doesn't say what will happen if you suspend judgment on the matter (maybe she'll flip a coin?). Assume that before she made the proposal, you hadn't even considered whether it would rain. Supposing you rationally believe she'll deliver on the proposal, it seems like you're now in a permissive case. If you believe that it will rain in Canberra, then it certainly will rain, so that belief is surely justified. If you believe it won't, it certainly won't, so that belief is surely also justified. Uncertainty only creeps in if you suspend judgment. So if we let P be the proposition that "It will rain in Canberra tomorrow," then it's rationally permissible for you to form either a belief in P or, instead, a belief in not P.[24]

This example, if possible, would refute intrapersonal versions of Uniqueness and thereby refute interpersonal versions.

My response is to concede that the self-fulfilling example is a counterexample to Uniqueness, but to modify Uniqueness so that it only applies to act-state independent doxastic attitudes. These are doxastic attitudes whose accuracy does not

---

[23] For other arguments to this effect, see Morten Dahlback, "Infinitely Permissive," *Erkenntnis* (forthcoming); Jonathan Drake, "Doxastic Permissiveness and the Promise of Truth," *Synthese* 194 (2017): 4897-4912; Matthew Kopec, "A Counterexample to the Uniqueness Thesis," *Philosophia* 43 (2015): 403-409; Thomas Raleigh, "Another Argument Against Uniqueness," *Philosophical Quarterly* 67 (2017): 327-346.
[24] Kopec and Titelbaum, "The Uniqueness Thesis," 197.

depend on whether the agent comes to have that doxastic attitude.[25] The result is this:

> **Uniqueness***: It is impossible for agents who have the same total evidence to be ideally rational in having different act-state independent doxastic attitudes toward the same proposition.

This modification would still be in line with the considerations that motivate Uniqueness. By now, there are many different arguments for Uniqueness. But if there is one main motivation for Uniqueness, it is that epistemic rationality would allow for objectionably arbitrary beliefs if Uniqueness were false. However, the self-fulfilling example doesn't involve the kind of arbitrariness that impermissivists are concerned to avoid. So, this modification is relatively painless for the impermissivist.

Moreover, merely pushing impermissivists back from Uniqueness to Uniqueness* would fail to vindicate any of the permisisvist's favorite examples. A central motivation for permissivism is to vindicate examples in which it appears that people are rationally disagreeing despite having the same evidence (e.g., the examples from Rosen). But the self-fulfilling example has nothing to do with permissivists' favorite examples, since these examples involve act-state independent doxastic attitudes. In sum, the self-fulfilling example seems very remote from the considerations that are most important to the debate about Uniqueness.

Conclusion

I have argued that, if we replace Uniqueness with Uniqueness*, then none of the five examples considered by Kopec and Titelbaum should lead us to reject impermissivism. Here is a quick summary of the responses I have given to each alleged counterexample: Impermissivists can respond to the jury example by granting that the agents in question are subideally rational, but denying that they are ideally rational. Impermissivists can respond to the materialism example by, again, granting that the agents in question are subideally rational, but denying that they are ideally rational. Impermissivists can respond to the community example by noting that the intuitive verdict doesn't support permissivism. Impermissivists can respond to the reasoning room example by rejecting the question-begging assumption that there are multiple ways of reasoning that are ideally rational. Finally, impermissivists can respond to the self-fulfilling example by replacing Uniqueness with Uniqueness*.[26]

---

[25] Dahlback, "Infinitely Permissive."

[26] Thanks, Cara Cummings, Chris Arledge, and the anonymous reviewers.

CAN KNOWLEDGE REALLY BE NON-FACTIVE?

Michael J. SHAFFER

ABSTRACT: This paper contains a critical examination of the prospects for analyses of knowledge that weaken the factivity condition so that knowledge implies only approximate truth.

KEYWORDS: knowledge, factivity, falsehoods, assertion, inconsistency, epistemic logic

1. Introduction

The decidedly timeworn and orthodox analysis of the nature of knowledge is that knowledge is justified true belief. So,

(JTB) S knows that p, if and only if,

(i) S believes that p,

(ii) S's belief that p is justified and

(iii) p is true.

Though this analysis is now discredited, (i), (ii) and (iii) were supposed to be individually necessary and jointly sufficient for knowledge. JTB was thus supposed to be a sort of decompositional and hence informative analysis of the nature of knowing. However, as we all now know, in 1963 Edmund Gettier (at least on the common interpretation) showed that the JTB account of knowledge is incorrect.[1] What Gettier specifically did was to present two cases where conditions (i)-(iii) were met but where our intuitions are supposed to be that the agent in question does not have knowledge.[2] In effect Gettier challenged the sufficiency of the JTB account on the basis of the following sort of case. Consider the case of Smith.[3] We are to suppose that Smith has strong evidence for the claim that Jones owns a Ford. This evidence

---

[1] Edmund Gettier, "Is Justified True Belief Knowledge?" *Analysis* 23 (1963), 121-123.

[2] It is worth mentioning that Gettier's case for the rejection of the JTB account *only* follows as a deductive consequence given the assumptions of epistemic closure and the idea that one can be justified in holding a false belief.

[3] Gettier "Is Justified True," 122.

includes that Jones has always owned a Ford and that Jones has just offered a ride to Smith while driving a Ford. Suppose also that Smith has a friend Brown and that Smith does not know where Brown currently is. So Smith formulates the following beliefs. Either Jones owns a Ford or Brown is in Boston. Either Jones owns a Ford or Brown is in Barcelona. Either Jones owns a Ford or Brown in Brest-Litovsk. All three are entailed by the claim that Jones owns a Ford. But, suppose that Jones does not in point of fact own a Ford, say he is presently driving a rental car. Moreover, by coincidence suppose that unknown to Smith Brown is actually in Barcelona. This means that Smith meets conditions (i)-(iii) of the JTB analysis, but intuitively we do not believe that Smith knows that either Jones owns a Ford or Brown is in Barcelona. What has happened is that Smith's belief has been caused in some inappropriate manner and the truth of his justified belief is, in some important sense of the term, a matter of luck. Of course, the intuition is supposed to be widely shared by philosophers and Gettier's cases have been widely taken to refute the JTB analysis of knowledge. Importantly, in light of this result, many practitioners of post-Gettier epistemology have then been concerned with the offering of an alternative analysis of knowledge, prominently including "fourth condition" analyses (JTB+ analyses) that are intentionally designed rule out Gettier cases as involving bona fide knowledge.[4]

There is, however, another line of thought that has arisen out of the challenge that Gettier's ruminations about the adequacy of the JTB analysis. Specifically and even more controversially, some thinkers have raised concerns about the necessity of condition (iii) of the JTB/JTB+ analyses despite wide-spread agreement that it should be part of a correct analysis of knowledge.[5] That is to say, these thinkers have raised concerns about the orthodox factivity condition on knowing.[6] Most

---

[4] See Peter Unger, "An Analysis of Factual Knowledge," *Journal of Philosophy* (1968): 157-170, George Pappas and Marshall Swain (eds.), *Essays on Knowledge and Justification* (Ithaca: Cornell University Press, 1978), Robert Shope, *The Analysis of Knowing* (Princeton: Princeton University Press, 1983), and Ram Neta, "Defeating the Dogma of Defeasibility," in *Williamson on Knowledge,* eds. Patrick Greenough and Duncan Pritchard (Oxford: Oxford University Press, 2009), 161-182 for a survey of the variety of post-Gettier accounts of knowledge.

[5] See, for example, Duncan Pritchard, *Knowledge* (London: Palgrave Macmillan, 2009), Robert Audi, *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (New York: Routledge, 1998), Laurence BonJour, *Epistemology: Classic Problems and Contemporary Responses* (New York: Rowman & Littlefield, 2002), Roderick Chisholm, *Theory of Knowledge* (Englewood Cliffs, NJ: Prentice Hall, 1989) and Jonathan Jenkins Ichikawa and Mathias Steup, "The Analysis of Knowledge" *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

[6] See, for example, Robert Ackermann, *Belief and Knowledge* (New York: Doubleday, 1972) and

prominently, Buckwalter and Turri have recently claimed that knowledge simply does not entail strict truth.[7] Notice that dropping the factivity condition amounts to the contention that epistemic agents can, at least sometimes, know propositions that are *false*. This line of thought has some independent support in the form of observations concerning the prevalence of approximations in human cognition. In fact, a number of philosophers and psychologists have compellingly argued that rational thinking and acting involves the use of all sorts of approximations, idealizations and/or inexact truths far more often than epistemologists have previously acknowledged.[8] That we are less than perfectly rational is, of course, not at all a new recognition and the work of the various defenders of the heuristics and biases tradition, the ecological rationality model and some more traditional views attests to this.[9] What is most relevant here is that this line of thinking strongly suggests that we sometimes base both practical and theoretical reasoning on propositions that are not-exactly-true and that we can be efficient problem solvers and deliberators in even though we do not reason in maximally accurate ways on the basis of exact truths. In other words, we often trade degrees of accuracy with respect to truth for things like efficiency, ease of use and generality without seemingly compromising rationality or success. So, there is nothing at all unusual about employing approximate, partial or inexact truths in our epistemic practices. This sort of sentiment is reflected in claims about epistemic states and their acceptance like this one: "…epistemic acceptability turns not on whether it [a proposition] is true, but on whether it is *true enough*–that is, on whether it is close

---

Allan Hazlett, "The Myth of Factive Verbs," *Philosophy and Phenomenological Research* 80 (2010): 497–522,

[7] Wesley Buckwalter and John Turri, "Knowledge and Truth: A Skeptical Challenge," *Pacific Philosophical Quarterly* (forthcoming).

[8] See Catherine Elgin, *Considered Judgment* (Princeton: Princeton University Press, 1996), Catherine Elgin, "True Enough," *Philosophical Issues* 14 (2004): 113-131, Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983), Elijah Millgram, *Hard Truths* (London: Wiley-Blackwell, 2009), Paul Teller, "Twilight of the Perfect Model," *Erkenntnis* 55 (2001): 393-415, Paul Teller, "The Finewright Theory," in *Nancy Cartwright's Philosophy of Science*, edited by Stephan Hartmann, Carl Hoefer and Luc Bovens, 91-116. London: Routledge, 2008), Mark Wilson, *Wandering Significance* (Oxford: Oxford University Press, 2006) and William Wimsatt, *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality* (Cambridge: Harvard University Press, 2007).

[9] See, for example, Christopher Cherniak, *Minimal Rationality* (MIT Press: Cambridge, 1986), Renée Elio (ed.), *Common Sense, Reasoning and Rationality* (Oxford: Oxrod University Press, 2002), Massimo Piattei-Palmarini, *Inevitable Illusions* (New York: Wiley, 1994) and Gerd Gigerenzer, *Adaptive Thinking* (Oxford: Oxford University Press, 2000).

enough to the truth."[10] However, on this basis, one might be tempted to draw the rather extreme and pessimistic skeptical conclusion that we do not really know very much at all, especially if one maintains strict factivity of knowledge proper.[11] Moreover, on this basis one might be tempted to re-orient epistemology on something other than knowledge. Alternatively, dropping the strict factivity condition might be one way to avoid this extreme skeptical conclusion, avoid giving up on the concept of knowledge and take seriously the observations about approximation. The rather radical alternative proposal under consideration here is then that these sorts of observations about approximation are best dealt with by revising the analysis of knowledge itself. The most obvious and reasonable proposal of this sort, as explicitly pursued by Buckwalter and Turri, involves replacing (iii) with a weakened necessary condition related to factivity but framed in terms of approximate truth.[12] In other words, the view in question here holds that we should adopt the view that *knowing implies approximate truth rather than strict truth*. Let us call this view the quasi-factivist account of knowledge.

2. Quasi-factive Knowledge and Approximate Truths

Quasi-factivism about knowledge is then the suggesting of something like the following alternative to the JTB/JTB+ analyses of knowledge:

(JATB+) S knows that p, if and only if,

---

[10] Catherine Elgin, *True Enough* (Cambridge: MIT Press, 2017), 16. Although Elgin does not endorse the view to be discussed here (i.e. that knowledge entails only approximate truth) she does suggests re-focusing epistemology on the concept of acceptance rather than on **belief.** See Elgin, *True Enough*, 3. The distinction between acceptance and belief is defended in L. Jonathan Cohen, *An Essay on Belief and Acceptance* (Oxford: Clarendon Press, 1992)**.**

[11] Peter Unger infamously defended this position in his book *Ignorance* (Oxford: Oxford University Press, 1975). The quasi-factivist view of knowledge is then pretty clearly intended to avoid having to draw this conclusion on the basis of the recognition of the prevalence of approximations in human cognition. Notice also that this line of argumentation very closely parallels the idealization argument against scientific realism as discussed, for example, in Nancy Cartwright, *How the Laws of Physics Lie* (Oxford: Oxford University Press, 1983), Lawrence Sklar, *Theory and Truth* (Oxford: Oxford University Press, 2000), and Michael Shaffer, *Counterfactuals and Scientific Realism* (New York: Palgrave MacMillan, 2012). More specifically, see Buckwalter and Turri, "Knowledge and Truth," for an attempt to avoid this skeptical conclusion.

[12] See Buckwlater and Turri, "Knowledge and Truth." They are explicit about adopting this view and claim that "We propose a fourth response: knowledge does not require truth (i.e. reject Line 1). Instead, false but approximately true propositions can be known. Call this *the approximation account* of knowledge. On this view, representations need not be true in order to count as knowledge (5)."

(i) S believes/accepts that p,

(ii) S's belief that p is justified,

(iii') p is approximately true and

(iv) S's justified belief that p meets the required "additional anti-luck condition(s)".

So, here (iv) (in an admittedly vague manner) is meant to indicate the idea that knowledge requires some sort of anti-luck condition necessary to secure reliability and stave off Gettier cases.[13] More importantly for the purposes at hand, (iii') is the condition that knowledge is "quasi-factive."[14] Accordingly, given JATB+ one can know at least some false propositions, specifically one can know approximate truths when all of the other conditions for knowing are met. This might seem to be a prima facie compelling view given how common approximation and hedging are in human cognition, but is it really defensible to suppose that one can really ever know a false but approximately true proposition? The answer defended here is an emphatic "no." This is the case because here are at least three serious problems with the quasi-factivist view of knowledge.

3. The Inconsistency and Explosion Objection

The first problem for the quasi-factivist about knowledge has to do with the following deeply troubling consideration related to inconsistency, closure and the logical principle known as ex contradictione (sequitur) quodlibet (ECQ). Suppose that the quasi-factivist about knowledge is right and one can know propositions that are approximately true. By definition all approximate truths are false and so the quasi-factivist about knowledge is thereby committed to the idea that some falsehoods can be known.[15] Consider then a claim p that is approximately true ATp

---

[13] See Peter Unger, "An Analysis of Factual Knowledge," *Journal of Philosophy* (1968): 157-170 and Alvin Goldman, "Williamson on Knowledge and Evidence," in *Williamson on Knowledge,* ed. Greenough and Pritchard, 73-91. There, concerning the JTB account, Goldman tells us, "It is obviously incomplete, and we have some fairly good ideas about how to repair it, or at least to improve upon it. That is not to say that there is unanimity among epistemologists. Nevertheless, some sort of additional conditions in an "anti-luck" vein are widely agreed to be necessary for a satisfactory account of knowing (75)."

[14] See Michael Shaffer, "Approximate Truth, Quasi-factivity and Evidence," *Acta Analytica* 30 (2015): 249-266.

[15] See Risto Hilpinen, "Approximate Truth and Truthlikeness," in *Formal Methods in the Methodology of the Empirical Sciences, ed. Marian* Przelecki, *et al.* (Dordrecht: Reidel, 1976), 19-42, Theo Kuipers, *What is Closer-to-the-truth? (*Amsterdam: Rodopi, 1978), Graham Oddie, *Likeness to Truth* (Dordrecht: Reidel, 1986), Graham Oddie, "Truthlikeness," *The Stanford*

and which is known, Ksp, by some epistemic agent S. For example, suppose that (in the quasi-factivist sense), after properly conducting a rigorous measurement using measurement process M, Joe knows that the value of some measured variable x is 5.6u but where the real value of x is 5.600000000001u.[16] That the value of x is 5.6u is then only approximately true. On this basis, the quasi-factivist view allows for the following state to obtain Ksp & ATp. But, suppose that S knows that p is approximately true KsATp and is aware of the fact that all approximate truths are false. Since ATp then implies ¬p and (as we shall see shortly) knowledge is closed under implication (even for quasi-factivists about knowledge), in such a situation Ksp and Ks¬p. Returning to our example let us suppose then that Joe knows that it is only approximately true that the value of x is 5.6u and that, on this basis and his understanding of the nature of approximate truth, Joe knows that it is false that the value of x is 5.6u. So, Joe knows that the value of x is 5.6u and he knows that it is not the case that the value of x is 5.6u. It should be obvious that this is deeply troubling. Such reflective knowledge of approximate truths entails knowledge of contradictions. However, this is not even yet as brutally damaging from an epistemic perspective as it in point of fact turns out to be, because we have not yet seen how this gives rise to the problem of epistemic explosion. Let us proceed by looking a bit more carefully at the logic of knowledge and the extent of this problem will become clear.

The standard axiomatization of the logic of knowledge is **KDT**.[17] The relevant axioms that constitute this system and its stronger and weaker relatives are as follows:

(**K**) Ks(p → q) → (Ksp → Ksq)

(**D**) Ksp → ¬Ks¬p

(**T**) Ksp → p

(**4**) Ksp → KsKsp

(**5**) ¬Ksp → Ks¬Ksp

Encyclopedia of Philosophy (Fall 2008 Edition), Edward N. Zalta (ed.), URL= <http://plato.stanford.edu/archives/fall2008/entries/truthlikeness/>.

[16] Here "u" is some arbitrary unit of measurement. See Michael Shaffer, "Rescuing the Assertability of Measurement Reports," *Acta Analytica* 34 (2019): 39-51 for some problems about these sorts of cases as they relate to approximation, assertion and knowledge

[17] See Vincent Hendricks and John Symons, "Epistemic Logic," *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2015/entries/logic-epistemic/>.

(.**2**) ¬Ks¬Ksp → Ks¬Ks¬p

(.**3**) Ks(Ksp → Ksq) ∨ Ks(Ksq → Ksp)

(.**4**) p → (¬Ks¬Ksp → Ksp)

What is of great importance here is that **K** in particular generalizes to the principle of closure for knowledge and even the logic of quasi-factive knowledge must obey **K**, though it explicitly does not obey **T**. The knowledge closure principle based on **K** is often then rendered simply as follows:

(SCK) If Ksp and Ks(p → q), then Ksq.

Now this version of the principle, the subjective closure of knowledge under material implication, says that if S knows p and also knows that p materially implies q, then S knows q. There is however a stronger version of the principle, objective closure of knowledge under material implication as follows:

(OCK) If Ksp and (p → q), then Ksq.

This version of the closure principle states that if S knows a proposition, then S knows all of the material implications of that proposition. In the argument that follows we need only assume SCK (grounded on the basis of axiom **K**) and stipulate that S is a reflective agent who knows the relevant implications built into the constructed (but totally generic) example introduced above. All of this guarantees that Ksp and Ks¬p implies Ks(p & ¬p). But, if this is true, then S knows *everything* due to the epistemic explosion problem that arises in virtue of ECQ and SCK. But, what exactly is ECQ? ECQ is a rather surprising but valid argument form in classical deductive logic. So it is a theorem of classical logic. Its proof is quite simple:

P1: (p & ¬p)   assumption

P2: p   conjunction elimination [P1]

P3: ¬p   conjunction elimination [P1]

P4: p ∨ q   disjunction introduction [P2]

P5: q   disjunctive syllogism [P3, P4]

∴ (p & ¬p) → q   conditional proof [P1-P5]

Of course, this generalizes for any q whatsoever and so every contradiction validly implies every proposition in the context of classical logic. So if (reflectively) Ksp & KsATp., ECQ is valid and knowledge in governed by SCK, then S would know everything. Based on the quasi-factivist account of knowledge, knowing an approximately true proposition in the reflective sense noted here logically entails knowing every proposition. But, this is absurd. Knowing a single approximate truth

should not entail omniscience. So, JATB+ cannot possibly be the correct analysis of knowledge.

4. The Safety Objection

The second, equally devastating, objection to JATB+ arises out of considerations having to do with the safety condition on knowledge. The safety condition for knowledge is a condition for knowing that has been most systematically defended by Williamson, Sosa and Pritchard.[18] This condition, among other things, is supposed to ground the difference between knowledge and lucky true belief (especially of the sort involved in Gettier cases) by introducing an element of reliability into the definition of knowledge that is lacking in the case of luckily true beliefs. In other words, it is supposed to do the work required of condition (iv) of the JTB+ analysis.[19] The safety condition can be understood simply as follows:

If S knows that p, then S could not easily have falsely believed that p.

This relatively non-technical gloss on safety and it can be made more precise in modal terms as follows:

(Safety) $(w_i \vDash K_s p) \rightarrow \neg[<w_i> \vDash (B_{s'} p \,\&\, \neg p)]$.

Here 'w_i' is world i, '$K_s p$' represents that S knows that p, '$<w_i>$' is the set of worlds sufficiently close to w_i, and '$B_s p$' represents that S believes that p. So understood, the safety condition is the claim that if S knows that p, then there are no worlds sufficiently similar to w_i (including w_i) where S′ (S's counterpart in those worlds) believes that p and p is false. This regimentation of the safety condition captures the core idea of that condition well and the contrapositive of safety is also interesting to note:

(Contrapositive Safety) $[<w_i> \vDash (B_{s'} p \,\&\, \neg p)] \rightarrow \neg(w_i \vDash K_s p)$.

This version of safety essentially is the assertion that if S could easily have falsely believed that p, then S does not know that p. More technically, it is the claim that if

---

[18] Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000), Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 141-54, Duncan Pritchard, "Anti-Luck Epistemology," *Synthese* 158 (2007): 277-98, Duncan Pritchard, "Knowledge, Luck, and Lotteries," in *New Waves in Epistemology*, eds. Vincent Hendricks and Duncan Pritchard (London: Palgrave Macmillan, 2008): 28-51, Duncan Pritchard, "Safety-Based Epistemology: Whither Now?" *Journal of Philosophical Research* 34 (2009): 33-45, and Duncan Pritchard, *Knowledge* (London: Palgrave Macmillan, 2009).

[19] See Michael Shaffer, "An Argument for the Safety Condition," *Logos & Episteme* 8 (2017): 517-520 for an argument that the safety condition is, in fact, implied by the JTB analysis.

in worlds sufficiently similar to w_i S believes that p and p is false, then S does not know that p at w_i. The easiest way then to see why safety should be regarded as a necessary condition for knowing is to understand what the denial of safety involves. It involves this:

(Unsafe Knowledge) ($w_i \vDash K_sp$) & [$<w_i> \vDash (B_sp \ \& \ \neg p)$].

Knowing *p* at a given world is compatible with falsely believing *p* in worlds close to that given world. But, it must be the case that $w_i \in <w_i>$. This because any account of closeness (i.e. any account of world similarity) must be reflexive and every world is *maximally* similar to itself. Thus, denying safety entails that one can know a claim at a world where that claim is false. But, this is absurd for all the familiar reasons. But notice now that, in virtue of JATB+, the quasi-factivist is automatically committed to the idea that there can be unsafe knowledge! In fact, according to the quasi-factivist one can have knowledge of p at a world where p is just false: $w_i \vDash K_sp$ & $\neg p$. The problem then is that in adopting (iii') the quasi-factivist thereby automatically rejects safety and thus deprives themselves of this sort of reliability condition that allows for the satisfaction of (iv). This then deprives them of an important resource that allows for dealing with Gettier problems and other problems related to epistemic luck. Perhaps there is some other sort of account of a condition that could do the work of (iv) that is compatible with the quasi-factivist view, but it is at best unclear what this might principle be or even what such a condition might look like.

5. The Moorean Objection

The third objection raised here against the quasi-factivist view of knowledge concerns explaining the infelicity of claims of the following sort: "I know that p, but p is false." Let us call these sorts of claims Moorean knowledge claims and they have the following form: K_sp & $\neg p$. Given JTB/JTB+ analyses of knowledge where the logic of knowledge is at least as strong as **KDT** the infelicity of Moorean knowledge claims is easily explained. According to **T** and its incorporation in JTB/JTB+ in the form of condition (iii), $K_sp \rightarrow p$ and so the utterance of a claim to the effect that "I know that p, but p is false" is just the utterance of a (very) thinly veiled contradiction. For example, the assertion of the claim that "I know that electrons are negatively charged, but it is false that electrons are negatively charged" does not seem to be a legitimate assertion and according to the factivist about knowledge it is illegitimate because it is contradictory.

For these sorts of reasons (and some other related ones) many variously motivated thinkers have defended the view that the proper norm for *assertion* is

knowledge.[20] Timothy Williamson in particular has strongly defended the knowledge norm for assertion on this basis.[21] In its most elemental form, this principle is the following claim:

> (KNA) one should assert a proposition only if it is known.

Now, the knowledge norm of assertion has been subjected to considerable criticism, but it has also been vigorously defended by some influential contemporary philosophers.[22] Williamson in particular defends the knowledge norm for assertion by appeal to its supposed explanatory power, especially with respect to the infelicity of Moorean belief claims of the form "p, but I do not believe that p." Specifically, he argues that the knowledge norm of assertion is the best explanation of the unassertability of such claims and this is, of course, closely related to the problem raised here about Moorean knowledge claims and the quasi-factivist view of knowledge. In mounting this defense Williamson claims that Moorean sentences are (1) unassertable and (2) that the best explanation of this fact is that knowledge is the proper norm of assertion.

So why is the alleged unassertability of Moorean claims supposed to support the knowledge norm of assertion, especially in the case of factive accounts of knowledge? This is supposed to be the case because if asserting that p is governed by the norm of knowledge and knowledge is factive, then it follows that one should assert p only if it is true. Thus, on this version of the knowledge norm view, to assert a Moorean sentence is to violate the knowledge norm of assertion and the infelicity of such assertions is just a consequence of any version of KNA involving factivity. Essentially, one ought not to assert that p when p is not believed or when it is false because then it cannot be known according to the JTB/JTB+ interpretations of the knowledge norm. There are of course, several extant weaker proposals than the knowledge norm, but the same point will hold for any account of the norm of assertion that is factive.[23] The problem then for the quasi-factivist about knowledge

---

[20] See Jessica Brown, "Fallibilism, and the Knowledge Norm for Assertion and Practical Reasoning," in *Assertion: New Philosophical Essays*, eds. Jessica Brown and Herman Cappelen (Oxford: Oxford University Press, 2011), 153-174.

[21] Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000).

[22] See, for example, Jonathan Hawthorne, *Knowledge and Lotteries* (Oxford: Oxford University Press, 2004), Jonathan Hawthorne and Jason Stanley, "Knowledge and Action," *The Journal of Philosophy* 105 (2008): 571-590, Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000), Timothy Williamson, "Contextualism, Subject-Sensitive Invariantism and Knowledge of Knowledge," *The Philosophical Quarterly* 55 (2005): 213-235.

[23] See Michael Shaffer, "Not-Exact-Truths, Pragmatic Encroachment and the Epistemic Norm of Practical Reasoning," *Logos & Episteme* 3 (2012): 239-259, Michael Shaffer "Moorean Sentences

who endorses KNA is *how to explain the infelicity of Moorean knowledge claims*. Presumably, the only option open to the non-factivists about knowledge with respect to this issue is to deny that (at least some) Moorean knowledge claims are, in fact, infelicitous. According to JATB+ knowledge only implies approximate truth and in cases where we have the assertion that S knows p and p is false due to its approximate truth, there would be no need for an explanation of the infelicity involved. In other words, if JATB+ is correct, there should be no appearance of such infelicity and so there is no need to explain such infelicities. But these claims do appear to be infelicitous and so it is simply a mystery why this is so if JATB+ is correct and one endorses KNA.

However, there is another related problem here about assertability that arises for quasi-factivists about knowledge who endorse KNA. Specifically, if JATB+ is correct and one endorses KNA, then it seems perfectly felicitous to assert contradictory pairs of claims. Given JATB+ let us suppose that Joe properly conducts a rigorous measurement using measurement process M and so, on the basis of (iii') in particular, knows that the value of some measured variable x is 5.61u but where the real value of x is 5.6u. Further suppose, that to be careful Joe rigorously measures the value of x again using M and obtains the result that the value of x is 5.59u and so also knows that the value of x is 5.59u. That the value of x is 5.61 and that the value of x is 5.59u are both only approximately true relative to the true value of 5.6u, but they are approximately true *to the very same degree*.[24] But, that the value of x is 5.59u implies that the value of x is not 5.61u and vice versa. Thus, unless the quasi-factivist has some appropriate response, it appears to be the case that given the JATB+ account of knowledge Joe can know and assert *both* that "I know that the value of x is 5.59u, but it is false that the value of x is 5.59u" *and* "I know that the value of x is 5.61u, but it is false that the value of x is 5.61u". The appearance of infelicity in such pairs of cases of assertions of Moorean sentences is, however, undeniable and so even if the quasi-factivist somehow denies the infelicity of individual Moorean knowledge claims this is not apparently an option with respect to pairs of cases like

---

and the Norm of Assertion," *Logos & Episteme* 3 (2012): 653-658, Ram Neta, "Treating Something as a Reason for Knowledge," *Nous* 43 (2009): 684-699, and Clayton Littlejohn, "Must We Act Only on What We Know," *The Journal of Philosophy* 106 (2009): 463-473 for discussions of this issue.

[24] Notice that this observation further exacerbates the inconsistency problems that the quasi-factivist faces that were raised earlier. Notice also that such pairs (or larger sets) of such claims do not even need to be of the same degree of approximate truth. According to JATB+ knowledge implies approximate truth simpliciter and so an appropriately poised epistemic agent would potentially know *every* approximation with respect to a strictly true claim all of which are mutually exclusive.

these. So, this constitutes another serious problem for the quasi-factivist view of knowledge.

6. Quasi-factivity, Knowledge-like States and Knowledge

So, the conclusion defended here is that the quasi-factivist view of knowledge is deeply problematic and considerations related to the prevalence of approximations in human cognition should not be used to motivate and justify the rejection of factivity. This is simply because the rejection of factivity entails a whole host of what appear to be catastrophic problems for the analysis of knowledge and for distinguishing knowledge from less valedictory sorts of belief. But, this leaves us with the worrisome possibility that the prevalence of approximations in human thinking implies an extreme form of skepticism. But, there is a middle ground here and the far less problematic position is that the prevalence of approximations in human cognition indicates *only* that there may well be all sorts of *knowledge-like states*, including quasi-factive and justified doxastic states, with all sorts of useful features of both the theoretical and practical sort. First and for the reasons articulated here, such quasi-factive states involving approximations should not be confused with knowledge. Second, we need to see how such states relate to the sorts of skeptical worries that are alleged to motivate quasi-factivism about knowledge. But, this second question requires us to have a theory of knowledge-like states that can be brought to bear given these skeptical worries. So, there appears to be a pressing need to explore the nature(s) of quasi-factive states that involve approximations and are knowledge-like.[25]

---

[25] This project is worked out in Michael Shaffer, *Quasi-factive Belief and Knowledge-like States* (Lanham: Lexington, forthcoming).

A FUNCTIONAL APPROACH TO CHARACTERIZE VALUES IN THE CONTEXT OF 'VALUES IN SCIENCE' DEBATES

Joby VARGHESE

ABSTRACT: This paper proposes a functional approach to characterize epistemic and non-epistemic values. The paper argues that epistemic values are *functionally homogeneous* since *(i) they act as criteria to evaluate the epistemic virtues a hypothesis ought to possess, and (ii) they validate scientific knowledge claims objectively.* Conversely, non-epistemic values are *functionally heterogeneous* since they may promote multiple and sometimes conflicting aims in different research contexts. An incentive of espousing the functional approach is that it helps us understand how values can operate in appropriate and inappropriate ways in scientific research and inappropriate influences can eventually be prevented. The idea is to argue that since non-epistemic values are functionally heterogeneous, they cannot provide objective reasons for the acceptance of a hypothesis. However, their involvement is necessary in certain research contexts and the problem is the involvement of these need not be always legitimate. By analyzing a case from chemical research, I demonstrate that how non-epistemic values might influence scientific research and, then, I go on to demonstrate that how a proper understanding of the functions of different kinds of values might promote the attainment of multiple goals of a particular research in a legitimate and socially relevant way.

KEYWORDS: science and values, functional approach, non-epistemic values, inductive risk

1. Introduction

At present, there are multiple debates about values and science. One such debate is whether values or norms should have any role in theory choice in science. The second debate is that, if we admit values, which values we should admit in the choice of a hypothesis? In other words, which values may influence science and how their influences can be evaluated.[1] This paper discusses the second debate in a detailed way.

---

[1] Matthew J. Brown, "Values in Science beyond Underdetermination and Inductive Risk, *Philosophy of Science* 80, 5 (2013): 829-839, Joby Varghese, "Influence and prioritization of non-

Joby Varghese

One way of drawing the dividing line between admissible values and inadmissible values in science is to characterize the admissible values as "epistemic" and the inadmissible values as "non-epistemic." But many have argued that it is difficult, perhaps impossible, to draw a principled line between epistemic and non-epistemic values.[2] This paper attempts to resolve the problem of epistemic/non-epistemic distinction by adopting a characterization grounded on the functions of various values. The paper also employs a functional approach to characterize epistemic and non-epistemic values because, at the end of the day, the ultimate goal of this distinction, from a pragmatic point of view, is to carefully avoid sneaking of illegitimate influences of values into scientific research which may bring unsolicited epistemic and non-epistemic ramifications.

This paper proposes and defends a characterization of both epistemic and non-epistemic values in terms of their functions. In section 2, I briefly analyze the questions concerning the nature of epistemic values. In section 3, I characterize epistemic values as functionally homogeneous. I demonstrate that the functional homogeneity of epistemic values can be depicted in twofold ways; *(i) they act as criteria to evaluate the epistemic virtues a hypothesis ought to possess, and (ii) they validate scientific knowledge claims objectively.* Section 4 discusses the characterization of non-epistemic values. These values are characterized as functionally heterogeneous because these values may promote multiple and even conflicting functions. I argue that a proper assessment of this heterogeneity is necessary to understand how values can operate inappropriate and appropriate ways. The reason for adopting the functional approach to characterize epistemic and non-epistemic values is that a large section of the ongoing debates on science and values is concerned with distinguishing the legitimate[3] and illegitimate influence[4] of non-

---

epistemic values in clinical trial designs: a study of Ebola ça Suffit trial," *Synthese* (2018): 1-17, Joby Varghese, "Philosophical import of non-epistemic values in clinical trials and data interpretation," *History and philosophy of the life sciences* 41, 2 (2019): 14, https://doi.org/10.1007/s40656-019-0251-4

[2] Phyllis Rooney, "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1 (Chicago: University of Chicago Press, 1992), 13-22, Phyllis Rooney, "The Borderlands Between Epistemic and Non-Epistemic Values," in *Current Controversies in Values and Science*, eds. Kevin C. Elliott and Daniel Steel (London: Routledge, 2017), 31-45.

[3] A legitimate influence is that which promotes the attainment of the goals of a scientific research in a scientifically better and socially relevant way.

[4] An illegitimate influence of values in scientific research might obstruct the attainment of the goals the research and it might also cause the production of a biased research outcome.

epistemic values when their influence is necessary[5] in certain scientific research contexts. In fact, the illegitimate influence of values always stems from the employment of certain values in different phases of scientific research when it is necessary.

Looming over the discussion about values is a discussion of what science is. We have two main theories: *the picture theory* which puts forth the view that science presents pictures of the world[6] and *the problem theory* which says that science is a problem-solving activity.[7] The picture theory has a low view of non-epistemic values, and consigns non-epistemic values to the dustbin labelled "logic of discovery." The problem theory is more accommodating towards non-epistemic values and rejects the distinction between the logic of discovery and the logic of justification. In sub-section 4.1, by analyzing a case from chemical research, I show that how non-epistemic values might influence scientific research. Then, I go on to demonstrate how a proper understanding of the functions of different values might promote the attainment of multiple goals of a particular research in a legitimate and socially relevant way. On top of that, an advantage of adopting functional approach is that the aims approach, which is generally used to distinguish the legitimate and illegitimate influence of non-epistemic values, can be further strengthened.[8]

2. Questions Regarding the Nature and Characterization of Values

The discussion on the concepts of epistemic and non-epistemic values and their distinction is a significant part of the present debates on science and values. In the subsequent sections, I will quickly brush through these concepts. The term

---

[5] Necessary involvement of values includes such scenarios where the problem of underdetermination and the problem of inductive risk arise. Policy related and socially relevant scientific research might also require necessary involvement of values in different stages of the research.

[6] Bas van Fraassen, *Scientific Representation: Paradoxes of Perspective* (Oxford: Oxford University Press, 2008), 269-288, Ludwig Wittgenstein, *Tractatus Logico Philosophicus: Logical-Philosophical Treatise* (The Edinburgh Press, 1922), 25-30, Bertrand Russell, *The Philosophy of Logical Atomism* (London: Routledge, 2009), 110-125.

[7] Kenneth F. Schaffner, *Discovery and Explanation in Biology and Medicine* (Chicago: University of Chicago Press, 1993), Larry Laudan, "Why Was the Logic of Discovery Abandoned?" in *Scientific Discovery, Logic and Rationality*, ed. Thomas Nickles (Dordrecht: Reidel, 1980), 173–183.

[8] Daniel J. Hicks, "A new direction for science and values," *Synthese* 191, 14 (2014): 3271-3295, Kristen Intemann, "Distinguishing between legitimate and illegitimate values in climate modeling," *European Journal for Philosophy of Science* 5, 2 (2015): 217-232, Varghese, "Influence and prioritization," 1-17.

"epistemic value" has been characterized in various ways according to the questions posed against such values. There are three general questions one might ask about epistemic values.

> (I) What does it mean to say X is an epistemic value, for example, what makes something an epistemic value versus some other kind of value?

> (II) Which particular values are epistemic values; for instance, are values such as simplicity, novelty, and ontological heterogeneity epistemic values?

> (III) What role should epistemic values play in guiding a theory or a hypothesis choice?

Before we critically examine these questions, a clarification on a general point is worth noting in the context of this paper. Although 'epistemic' and 'non-epistemic' are technical terms and technical terms must have their definitions stipulated, the current debates in values in science have been more concerned with the functions of these values rather than defining them. Moreover, an important debate on science and values is concerned with the question: how to distinguish the legitimate influence of values from possible illegitimate influences and prevent such illegitimate influences? So, this paper does not seek to furnish any kind of definition of these values; rather, the attempt is to highlight the functions of these values and characterize them by invoking the functional approach. Holding this view in mind, let us resume the discussion concerning the general questions regarding epistemic values.

Although the above-given questions are posed separately, they are interrelated. The relation is that they give rise to different points of agreement and disagreement among philosophers who have tried to address these issues. The first question is concerned with defining what an epistemic value is. While responding to the first question, some have argued that epistemic values are truth-conducive.[9] They propose the view that epistemic values should be acknowledged on the grounds that they are capable of promoting the attainment of truth. On the other hand, there are others who have argued that epistemic values must only advance certain scientific aims which might also include other cognitive goals besides truth. This has also led to different terminology being used in the literature, such as

---

[9] Ernan McMullin, "Values in Science," in *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association – Symposia and Invited Papers*, eds. Peter D. Asquith and Thomas Nickles (Chicago: University of Chicago Press, 1983), 3–28, Daniel Steel, "Epistemic Values and the Argument from Inductive Risk," *Philosophy of Science* 77, 1 (2010): 14-34.

cognitive values vs epistemic values,[10] which is a lengthy debate, which I am not going to take up in this paper.

The second question is an extension of the first question which is concerned with defining epistemic values. Moving on to the second question, we might find that even among those who agree with what it is to be an epistemic value, there are debates about which particular values might fit that criterion. Popper argues that a high degree of falsifiability must be considered as an epistemic value.[11] Bas van Fraassen suggests that simplicity or explanatory scope should not be termed as epistemic values and the only value he seems to be admitting is empirical adequacy.[12] Hilary Putnam, on the other hand, argues that the list of cognitive values ought to include instrumental efficacy.[13] Then, there are others who argue that a clear-cut epistemic non-epistemic distinction is not possible.[14] They advocate the view that values such as simplicity which often falls into the category of epistemic values may also incorporate non-epistemic concerns. On a similar vein, Longino also puts forth the claim that novelty, applicability, and ontological heterogeneity might very well operate in the same way as certain alternative constitutive values function.[15] Interestingly, the list of accepted epistemic or constitutive values might vary in accordance with the person's commitment to a particular school of thought.

---

[10] Larry Laudan, "The Epistemic, the Cognitive, and the Social," in *Science Values and Objectivity*, eds. Peter Machamer and Gereon Wolters (Pittsburgh: University of Pittsburgh Press, 2004), 14-23, Heather Douglas, *Science, Policy, and the Value-Free Ideal* (Pittsburgh: University of Pittsburgh Press, 2009), 87-114.

[11] Karl R. Popper, *The Logic of Scientific Discovery* (London and New York: Routledge, 1959), 57-73. Falsification, according to Popper, is an epistemic value and he argues that what is unfalsifiable is unscientific and what is falsifiable is scientific. A falsifiable theory or hypothesis is that which can be put to test by which the hypothesis could conceivably be refuted.

[12] Van Fraassen, *The Scientific Image* (Oxford: Oxford University Press, 1980), 41-68. According to van Fraassen, a theory is empirically adequate if the observable phenomena are "embedded" in the theory. All actual observable phenomena are relevant to a theory's empirical adequacy. In order to be empirically adequate, a theory should be able to accommodate more than just the phenomena which have been actually observed but the phenomena that will be observed.

[13] Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge University Press, 1981), 136-37. Putnam argues that cognitive values of coherence, simplicity, and instrumental efficacy are entirely arbitrary but a part of our idea of human cognitive flourishing, and hence is part of our idea of total human flourishing, of Eudemonia.

[14] Rooney, "On Values in Science," 13-22, Rooney, "The Borderlands," 31-45, Steel, "Epistemic Values," 14-34,

[15] Helen E. Longino, "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy," in *Feminism, Science, and the Philosophy of Science*, eds. Lynn Hankinson Nelson and Jack Nelson (Dordrecht: Kluwer, 1996), 39-58.

It is the third question I am more concerned with for the present purpose of the paper, i.e., the functions different values ought to play in the evaluation of a hypothesis. In what follows, I offer a detailed characterization of both epistemic and non-epistemic values.

3. The Functional Approach

The question regarding the appropriate and inappropriate influences of values is set in the background assumption that values perform diverse roles and their employment might promote multiple and sometimes conflicting research aims. Apparently, the origin of the illegitimate or inappropriate influence of values in science lies in the adoption of such values which are not consistent with the pre-specified primary or secondary aims of a particular scientific research. So, it is important to analyze how different values play different functions in science regardless of the different phases[16] of scientific research. Once we are in a position to understand how their functions vary according to the research context, it might be an undemanding task to prevent the illegitimate influence of certain values in certain research contexts. In the following sub-sections, I develop the Functional Approach Principle (FAP).

FAP states:

> Different values promote multiple and even conflicting goals in a particular research context. The involvement and the functions of both epistemic and non-epistemic values are legitimate in a scientific research context if and only if the functions each particular value performs during the evaluation of a scientific hypothesis are consistent with the pre-specified aims of the research.

The notion of consistency in the context of our discussion does not refer to logical consistency. On the other hand, it refers to the compatibility between some values and certain pre-specified aims of a particular scientific research. A value is consistent with the pre-specified aims of the research if and only if that value, in no way, obstructs the attainments of the aims of the research. Nonetheless, certain values are consistent with the pre-specified aims of the research to a certain extent and once their influences exceed their limit, they might obstruct the attainment of the pre-specified aims. In other words, if the influence of a particular value goes beyond its limited role, at that point of time that particular value becomes inconsistent with the pre-specified aims of the research and the influence of that value in that

---

[16] Different phases of scientific research are (i) the pre-epistemic phase, (ii) the epistemic phase, and (iii) the post epistemic phase. However, there are still debates going on regarding whether it is possible to make a clear cut distinction among these phases.

particular research context becomes illegitimate. However, the question which remains is that how we can know whether the values to be employed are consistent with the pre-specified aims of the research and, if at all they are consistent, in which cases they cease to be inconsistent. In order to address this question, I endorse the aims approach which, in simple terms, states that that the choice of values must be made in such way that the chosen values may *promote* the attainment of pre-specified research goals.[17] On the other hand, the functional approach states that the chosen values must be *consistent* with the pre-specified aims of the research. The starting point of the functional approach is the assumption that science often strives to achieve multiple goals and the aims approach also assumes the same. In the following sections, I offer a detailed account of homogeneous/heterogeneous characterization of values and defend the scope of this characterization in science and values debates. I also put forth a revised form of aims approach which, I argue, is a better account because of the adoption of the functional approach.

3.1 Epistemic Values - Values with Homogeneous Functions

I categorize epistemic values as functionally homogeneous since they perform two important functions during the evaluation of a scientific hypothesis. The homogeneous functions of these values can be portrayed in twofold ways. These two functions are:

    (i)   these act as criteria to evaluate the epistemic virtues[18] a hypothesis, a theory or a model ought to possess; and

    (ii)   these validate scientific knowledge claims objectively.

The functional homogeneity does not imply that every value which performs these two functions can be used as substitutes for other values which might also duly perform these functions. For instance, it is absurd to say that empirical adequacy which satisfies the two functions can be replaced for explanatory power which also satisfies the same functions. It is very intuitive that these values are qualitatively different even though they are functionally homogenous. The string that binds together these two functions in a homogeneous way is the act of scientific

---

[17] Kevin C. Elliott and Daniel J. McKaughan, "Nonepistemic Values and the Multiple Goals of Science," *Philosophy of Science* 81, 1 (2014): 1-21, Kristen Intemann and Inmaculada de Melo-Martín, "Social values and scientific evidence: the case of the HPV vaccines," *Biology & Philosophy* 25, 2 (2010): 203-213, Intemann, "Distinguishing," 217-232.

[18] Epistemic virtues are those qualities a hypothesis should possess in order to be certified as acceptable. For instance, epistemic significance, credibility, a high degree of confirmation etc. can be considered as epistemic virtues.

knowledge production and validation and, hence, the term "homogeneity" has been used in a less stringent sense. That is to say, "the functional homogeneity," I talk about in this context, is stipulated only to the functions of epistemic values, i.e., their functions as criteria to evaluate the epistemic virtues a hypothesis ought to possess, and validating knowledge claims objectively. In what follows, I make an attempt to synchronize the functions of epistemic values under the term homogeneity and then characterize these values from an epistemic standpoint. The idea is to argue that values which perform the above-given functions are integral parts of the production and confirmation of scientific knowledge because without the employment of relevant epistemic values the production and the assessment of scientific knowledge is not warranted.

The aims of modern science clearly indicate the fact that the aims are diverse in nature.[19] For instance, Ronald Giere and Bas van Fraassen argue that representations might very well be assessed in different ways.[20] It can be through the affairs that those representations bear to the world and sometimes it is in connection with the several uses to which they are employed. Since representations can be evaluated in different dimensions, it is plausible to think that the decisions regarding the acceptance of a hypothesis might also depend on various epistemic and pragmatic considerations. However, the point is that the functions epistemic values perform during the assessment of a hypothesis do not change according to the aims of the research. Their functions remain intact irrespective of the research contexts and aims. Martin Carrier argues:

> Epistemic values are employed in assessing how well hypotheses are confirmed by the available evidence. They are used for singling out acceptable hypotheses. Acceptance can either mean the belief that a hypothesis (or model or theory) is sufficiently confirmed or the recognition that the hypothesis is useful for building further theoretical considerations on it.[21]

The account of acceptance advocated by Carrier has two-fold implications, and in both cases, epistemic values play inevitable roles. The first notion of acceptance is in relation to the idea of a mental acceptance of a claim likely to be true which is to be

---

[19] Wendy Parker, "Confirmation and Adequacy-for-Purpose in Climate Modelling," *Aristotelian Society Supplementary Volume* 83, 1 (2009): 233–249, Philip Kitcher, *Science, Truth, and Democracy* (New York: Oxford University Press, 2003), 55-82, Intemann and de Melo-Martín, "Social values," 203-213, Varghese, "Influence and prioritization," 1-17, Elliott and McKaughan, "Nonepistemic values," 1-21.

[20] Ronald N. Giere, "How Models Are Used to Represent Reality," *Philosophy of Science* 71, 5 (2004): 742-752, Van Fraassen, *Scientific Representation*, 141-184.

[21] Martin Carrier, " Values and Objectivity in Science: Value-Ladenness, Pluralism and the Epistemic Attitude," *Science & Education* 22, 10 (2013): 2547-2568.

confirmed. The second notion of acceptance is allied with the usefulness of a hypothesis to present a new knowledge claim or to add something more to the existing knowledge. The thrust is that although epistemic values which are involved in these two cases of acceptance are different, their functions remain unchanged, i.e., they function as criteria for the evaluation and acceptance of a hypothesis.

Let us consider the first function of epistemic values, i.e., they act as criteria for evaluating the epistemic virtues such as epistemic importance or a high degree of confirmation of scientific knowledge claims. The epistemic importance often influences the choice and the pursuit of theories or models in scientific research, and this importance is often influenced by epistemic values. Consider the standard curve-fitting problem. For example, while fitting a curve to a data set, the scientists often choose between either a higher-order polynomial or a lower order polynomial. The former has the advantage of measuring the data more accurately but makes the curve less simple while the latter effects the curve simpler, albeit less accurate. If the epistemic merits of each polynomial are considered, each of them has a different epistemic advantage(s) over the other. One is more accurate but less simple, and the other is more simple but less accurate. Here, the epistemic importance of the polynomials can be measured on the basis of epistemic values such that the choice between these polynomials is made in accordance with their abilities to fulfil certain expectations in a particular research context. For instance, econometricians have a preference for solving curve-fitting problems using linear regression, thereby choosing simplicity over accuracy. On a similar vein, one may argue that epistemic values are the essential yardstick of epistemic importance.[22] The epistemic importance I endorse here has implications for some sense of objectivity. That is to say; it does not have anything to do with the psychological states of an individual scientist; rather, it concerns what appears to be significant to the scientific community in general.

Having said this, a minor yet a significant point, on curve fitting, is worth noting. The example of curve-fitting seems to show that epistemic values can conflict and in such cases, it is not clear whether one ought to prefer a higher-order or lower-order polynomial since one may be simpler but less accurate or vice versa. Obviously, the curve fitting problem shows that epistemic values can also conflict. However, the point is that accuracy can be trumped by simplicity in a particular research context because simplicity better fulfils certain expectations in that particular context where accuracy at least does not have the same weight as simplicity. The example also gives cues to answer the question that when it is legitimate to give less weight or priority to accuracy. A prudent answer would be

---

[22] Carrier, "Values and Objectivity," 2547-2568.

that it depends on the value's viability to satisfy certain pre-specified aims in that particular research context in a better way than other alternatives do.

At the beginning of this sub-section, I have underscored the claim that during the confirmation process, epistemic values play certain pivotal roles. Every scientific hypothesis needs to be confirmed to a certain degree in order for them to be treated as a part of the body of scientific knowledge. Epistemic values have a significant role to play while scientists try to confirm a hypothesis under study. If there are two or more empirically equivalent accounts, then the choice is often made in terms of a particular account's ability to attain certain epistemic aims. For instance, suppose the first account is built upon a large number of unrelated hypotheses while the second one appeals to a few overarching principles, although both of them are empirically equivalent.[23] However, the commitment to coherence or simplicity or broad scope favours the latter approach because, in such kind of evaluation of accounts assessed in light of any or some of these values, the data, obtained which act as evidence, favours the more unifying account even if both the approaches are empirically equivalent. In such cases, the scientific community often invokes the help of epistemic values for choosing between empirically indistinguishable alternatives. Carrier argues that when there is a tiebreaker situation between competing accounts that conform to the data to approximately the same degree, the scientists appeal to values that transcend the requirement of empirical adequacy.[24]

Secondly, epistemic values function as a constraint for validating scientific knowledge claims objectively. A significant criticism that value-laden account of science often confronts is that the involvement of non-epistemic values in scientific inquiry might destroy the scientific objectivity. However, philosophers like Longino and Douglas have argued that non-epistemic values – contextual values in Longino's terms – can legitimately influence scientific inquiries without undermining

---

[23] An example of empirically equivalent theories is Heisenberg's matrix mechanics and Schrödinger's wave mechanics in the 1920s. Both of them dealt with understanding quantum mechanics. Initially, scientists preferred wave mechanics and even now some do because the theory fitted better with tradition and it also used familiar mathematical tools and techniques. On the other hand, while those of matrix mechanics were in less common usage in physics, finally it provided a mental image which could be more easily visualized. Eventually, it was shown that both theories are empirically equivalent and people use whichever is the more convenient formalism for a problem. For instance, sometimes the emphasis is put more in the wave formulation because this is much easier for most quantum problems. Matrix mechanics is highlighted because it is simpler while dealing with the harmonic oscillator.

[24] Martin Carrier, "Underdetermination as an epistemological test tube: expounding hidden values of the scientific community," *Synthese* 180, 2 (2011): 189-204.

objectivity.[25] Ideally, epistemic values depict the features of knowledge we consider worthwhile irrespective of the particular contexts in which the knowledge is used. Some of the well-known epistemic values which are cherished in any scientific inquiry are empirical success, predictive accuracy, breadth of explanatory scope and unification, simplicity, and problem-solving effectiveness. These values could plausibly be considered as the features of a lot of scientific knowledge claims which are objective in nature. Values cannot function as a criterion of scientific knowledge claims if they are employed in isolation without any background assumptions or contexts. For instance, predictive accuracy cannot function as a mark of scientific knowledge in the absence of standards of acceptable empirical approximation.

In a nutshell, the functions, epistemic values perform during the evaluation of a hypothesis, are homogenous in nature. However, the claim that epistemic values are homogeneous is constrained only to the functions of these values to act as criteria to evaluate the epistemic virtues such as epistemic significance, credibility, and a high degree of confirmation and to validate scientific knowledge claims objectively.

3.2 Non-epistemic Values – Values with Heterogeneous Functions

In contrast with epistemic values, non-epistemic values are social, political, moral, commercial, religious or any other values which belong to various disciplines. These values are integral elements in forming the culture and customs of any society, and these values are held to be desirable by different social groups or communities.

In the previous sub-section, I characterized epistemic values as homogenous in terms of their functions in scientific inquiries. On a similar vein, I characterize non-epistemic values as heterogeneous with regard to the functions they perform. First of all, non-epistemic values are those values which do not perform the two important functions which epistemic values do during the production, evaluation, and confirmation of scientific knowledge. Further, non-epistemic values are functionally heterogeneous since they perform a variety of roles in order to promote the attainment of different objectives a particular discipline aspires to achieve. The expression "heterogeneous functions," in a naïve sense, means that although the values found in different disciplines such as politics, ethics, business, etc., fall under the category of non-epistemic values, there are qualitative differences among these values, their functions vary and sometimes these values even promote conflicting aims in different research contexts. In this sense, it is plausible to argue that non-epistemic values perform diverse functions. We are also justified in assuming that

---

[25] Heather Douglas, "The irreducible complexity of objectivity," *Synthese* 138, 3 (2004): 453-473, Helen E. Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry* (Princeton: Princeton University Press, 1990).

this functional heterogeneity is a significant characteristic of non-epistemic values in the context of values in science debates since most of the current controversies in science and values are centered on evaluating the functions of non-epistemic values when their influence is necessary in certain scientific research contexts.

The debates on science and values mainly focus on the repercussions of the influence of non-epistemic values in scientific research and assessing the influences. The following sub-sections demonstrate that although the influence of non-epistemic values is necessary in scientific inquiries, the influence of these values can turn out to be inconsistent because of their functional heterogeneity. The real problem is identifying which non-epistemic values should influence the research when multiple competing sets of values are engaged and when they influence, how we know whether their influence is legitimate or illegitimate.

The appropriate and inappropriate influences of non-epistemic values are parasitic on the employment of these values. A value is neither legitimate nor illegitimate without its employment in a particular context. Once a value is employed in order to perform certain functions in a specific research context, then with its relation to the pre-specified research aims and its consistency in promoting those aims, the legitimacy of the influence of that particular value can be assessed in that particular research context. The idea is to argue that since non-epistemic values are functionally heterogeneous, the choice of the values must be made in such a way that the chosen values may be consistent with the pre-specified aims of the research. In the following sub-section, I explore two important functions of non-epistemic values and characterize them as functionally heterogeneous.

3.2.1 Non-epistemic Values as Guiding Hands to Show Directions

The defenders of the traditional account of value-free ideal argue that, although non-epistemic values constitute a wide range of considerations such as political, religious, personal, commercial, social and ethical, they do not reliably promote any epistemic goal because these values seldom act as an indicator of a good scientific theory.

The admissible and inadmissible roles of non-epistemic values in science have been discussed extensively, and without copious dispute, many philosophers and scientists hold on to the view that the involvement of non-epistemic values is necessary in the pre and post-epistemic phase of scientific investigations. There are several accounts of pre-epistemic, post-epistemic, and epistemic phase distinction which are compatible with the proposal that is being taken for granted for the purpose of our current discussion.[26] Although this division varies widely, in general,

---

[26] Elizabeth Anderson, " Uses of Value Judgments in Science: A General Argument, with Lessons

the pre-epistemic phase is where the theories or hypothesis are formulated, and methodologies are selected, and the post-epistemic phase is where the accepted hypothesis or the outcomes of the study are put into use either for the production of further scientific knowledge or technology.[27] In the pre-epistemic phase, values can channel scientific research in particular directions. For instance, non-epistemic values play a crucial role when funding agencies encourage researchers to pursue socially relevant lines of investigation. Similarly, values can influence the way in which scientific research is applied in the realm of public policy or technology development, and it is a post epistemic scenario. In general, it is relatively uncontroversial to claim that non-epistemic values influence scientific research in the above-mentioned activities which occur in the pre or post epistemic phases. In such cases, these values act as guiding hands to show directions in scientific research and technology and this function of non-epistemic values is quite well accepted by both the critics and the defenders of the value-free ideal.

3.2.2 Non-epistemic Values as the Promoters of the Research Goals

Traditionally, the epistemic/non-epistemic distinction was thought to reflect a difference in the kind of aim about which the value is concerned, in other words, those that promote scientific or epistemic aims against those that deter or fail to promote such aims. One of the focal questions in values in science debates is whether non-epistemic values can legitimately influence the internal aspects of scientific research which include inferences from experimental data to theories or hypotheses. The defenders of the value-laden account of science answer positively and there become the debates more interesting. The critics of value-free account of science argue that the place of non-epistemic values cannot be undermined in certain scientific research because they play decisive roles even in the epistemic or internal phase of the scientific research.

4. Right Tool for the Job – The Functional Approach

Let us return to the discussion of heterogeneous functions of non-epistemic values and analyze why and how different competing sets of non-epistemic values need to be necessarily involved when the research enmeshes the problem of inductive risk. The analysis is based on the functional approach, which is primarily employed for distinguishing the legitimate and illegitimate influence of non-epistemic values

---

from a Case Study of Feminist Research on Divorce," *Hypatia* 19, 1 (2004): 1-24, Douglas, *Science, Policy*, 44-65, Brown, "Values in science," 829-839.

[27] Hicks, "A new direction," 3271-3295.

when their involvement is indispensable. I also argue that functional approach is the best candidate to address the problem of inductive risk because the approach suggests which set of non-epistemic values is consistent with the pre-specified objectives of the research and diagnoses which values are legitimate based on its consistency with the aims of the research. Moreover, a richer understanding of the heterogeneous functions of non-epistemic values might better capture the epistemic and non-epistemic repercussions of accepting or rejecting a hypothesis. In the following sections and sub-sections, I analyze a case study from chemical research which involves the problem of inductive risk. By exploring this case, I demonstrate how the legitimate influence of non-epistemic values during the research would promote the attainment of the research objectives in a better way. In order to assess the legitimacy of the chosen values, I employ the functional approach. Let us consider the problem of inductive risk.

4.1 The Problem of Inductive Risk

The critics of the value-free ideal have successfully argued that there are occasions even in the internal phase of scientific inquiry where the involvement of non-epistemic values is necessary. An instance of such a scenario is when scientists confront the problem of inductive risk.[28] Inductive risk is the possibility that one may make a mistake in rejecting or accepting a hypothesis that is under study. The problem of inductive risk is rampant in different phases of scientific inquiries. The case which I am going to analyze in this sub-section is concerned with the problem of inductive risk while making the decisions regarding the choice of methodology for conducting lab experiments.

Wilholt[29] draws attention towards a case from chemical research which involves the problem of inductive risk. The primary aim of this particular chemical research study was to find out whether there was an association between certain adverse health effects and their exposure to bisphenol A (BPA), an organic synthetic compound.

---

[28] Heather Douglas, " Inductive Risk and Values in Science," *Philosophy of science* 67, 4 (2000): 559-579, Carl G. Hempel, *Aspects of Scientific Explanation; And other Essays in the Philosophy of Science* (New York: The Free Press and London: Collier-Macmillan, 1965), 1-19, West C. Churchman, "Statistics, Pragmatics, Induction," *Philosophy of Science* 15, 3 (1948): 249-268, Richard Rudner, "The Scientist Qua Scientist Makes Value Judgments," *Philosophy of Science* 20, 1 (1953): 1-6.

[29] Torsten Wilholt, "Bias and values in scientific research," *Studies in History and Philosophy of Science Part A* 40, 1 (2009): 92-101.

Many researchers like vom Saal et al.[30] argued that the outcomes of various studies showed that the exposure to BPA could cause several serious health issues such as prostate and breast cancer, neurobehavioral problems, obesity, reproductive abnormalities, etc. in humans. However, scientists like Ryan et al. disagreed with these findings and argued that their study outcome showed '*no association between certain adverse health effects and the exposure to BPA.*'[31] The difference between the two research outcomes, one concluding the association between the adverse health effects and the exposure to BPA positively and the other one concluding the association negatively, was that the first one was conducted by non-profit organizations and the second one was conducted by those research institutes which were funded by the industries. There were 119 studies which were conducted by the government agencies and except 10 studies, all other studies showed that BPA is toxic and its use is harmful. On the other hand, 11 studies which were designed and conducted by the industries concluded that there was no association between adverse health effects and the exposure to BPA. It seems that the researches which were funded by the industries were carried away by structural bias[32] against finding positive effects of BPA. Vom Saal et al. point out what went wrong with all those researches which were conducted by the financial support of chemical companies:

> [F]or toxicological studies conducted without appropriate positive controls and that report only negative findings for a test chemical, interpretation of the negative results is not possible and violates basic rules governing experimental research design and analysis, specifically the need for a valid positive control[33] when test results for a drug or chemical with a known mode of action are uniformly negative.[34]

---

[30] Frederick S. vom Saal, Susan C. Nagel, Barry G. Timms, and Wade V. Welshons, "Implications for human health of the extensive bisphenol A literature showing adverse effects at low doses: a response to attempts to mislead the public," *Toxicology* 212, 2-3 (2005): 244-52.

[31] Bryce C. Ryan, Andrew K. Hotchkiss, Kevin M. Crofton, and L. Earl Gray Jr, "In Utero and Lactational Exposure to Bisphenol A, In Contrast to Ethinyl Estradiol, Does Not Alter Sexually Dimorphic Behavior, Puberty, Fertility, and Anatomy of Female LE Rats," *Toxicological Sciences* 114, 1 (2010): 133-148.

[32] Sheldon Krimsky, "Do Financial Conflicts of Interest Bias Research?: An Inquiry into the 'Funding Effect' Hypothesis," *Science, Technology, & Human Values* 38, 4 (2013): 566-587. Krimsky claims that "Structural bias," is an adoption of certain methods or norms which would distort (over- or underreport) the effects being studied.

[33] Positive control is a test scientists perform against something when they know what the effects of that will be. Negative control is a test scientists perform against something when they know that the test will have no effect.

[34] Frederick S. Vom Saal and Wade V. Welshons, "Large effects from small exposures. II. The importance of positive controls in low-dose research on bisphenol A," *Environmental*

There are two significant methodological errors that were committed by the researchers who concluded that there was no association between certain adverse health effects and the exposure to BPA. Firstly, they did not employ a positive control. Secondly, they used a particular strain of rats as model organisms for conducting the experiment. These rats which are known as Long Evans (LE) rats are not really sensitive to estrogen. Consequently, these particular test rats reduced the probability of finding the toxicity of BPA to a large degree. This instance from chemical research clearly shows that the problem of inductive risk might creep into any phases of scientific research which might eventually affect the study outcome and bring about severe epistemic and non-epistemic corollaries.

4.2 Saving the Inductive Risk Scenario

In the case of BPA, the aim of the research was to find out if there was an association between certain adverse health effects and the exposure to BPA. Test of BPA in the mice which were conducted by the government agencies clearly showed that there is an association. However, some industry-funded studies came up with a negative result because the industry-funded research chose such a model organism for the trial which was insensitive to estrogen. Furthermore, the doses fed to the models were also insufficient. These inappropriate methodological choices led to the production of dubious and inadequate data which eventually affected the outcome of the research in a negative way.

Karl Popper, in his famous work *The Logic of Scientific Discovery*, talks about the philosophical foundations of scientific methodology. Popper argues that science is not an inductivist venture, where truth is built up from the data that are consistent with a hypothesis. According to him, scientists must pursue to falsify a hypothesis and a distinguishing feature of any good scientific theory is that its hypotheses pass the test.[35] Hence, under the methodology put forth by Popper, one should look for such an instance which might falsify a hypothesis that is to be confirmed. A hypothesis can be confirmed only if there is no falsifying instance happens. Let us explore how we can adopt this methodology in the case of BPA and similar kinds of researches. Vom Saal and Welshons argue:

> [I]t is a common event in toxicological studies conducted by the chemical industry for purposes of reporting about chemical safety to regulatory agencies to provide only negative results from a study in which no positive control was included but

---

research 100, 1 (2006): 50-76.

[35] Popper, *The Logic*, 57-73

from which positive conclusions of safety of the test chemical are drawn.[36]

The point is that the scientific validity of any experiments with animal models might be questioned if the experiment does not take account of both negative control and positive control doses. In the case of BPA research, the structural bias instigated the researchers to choose such a methodology for the experiment so that they could find what they wished to find. Wilholt and Biddle[37] argue that it is a clear case of *preference bias* which occurs when the researchers are heavily influenced by their preferences and studies are conducted in such way as to amplify the probability of obtaining the preferred outcome. The problem of inductive risk involved in the choice of the methodologies, particularly the choice of an insensitive model organism, brought about the production and sharing of corrupted research outcome. While designing and conducting scientific research, there might be multiple (epistemic and non-epistemic) goals researchers may try to achieve and every aim of that research might also accompany certain epistemic as well as non-epistemic repercussions. Having set up this framework where scientific research is concerned with multiple aims, let us examine the possible aims of BPA research and the implications and consequences of the primary aim being overridden by other aims.

The primary aim of conducting BPA research was to find out if there was an association between adverse health effects and the exposure to BPA. Similarly, some other subordinate non-epistemic repercussions might follow if the research concludes that there is an association between adverse health effects and the exposure to BPA. For instance, if the research concludes that BPA is harmful, then it follows certain non-epistemic repercussions such as minimization of social cost both in finance and healthwise, severe restrictions in the use of BPA by industries, and probably a huge financial loss to the industries. On the other hand, if the research concludes that BPA is not harmful as Ryan et al.[38] concluded when it is actually harmful, this conclusion is also followed by some other non-epistemic repercussions such as an increase in the social cost, free use of BPA products, an increase in the profit margin of the industries which are involved in manufacturing BPA products, etc. In this scenario, it is plausible to assume that a form of structural bias really crept into the research methodology of industry-funded research and eventually ended up in the production of a biased research outcome.

---

[36] Vom Saal and Welshons, "Large effects," 50-76.

[37] Justin Biddle, "Institutionalizing Dissent: A Proposal for an Adversarial System of Pharmaceutical Research," *Kennedy Institute of Ethics Journal* 23, 4 (2013): 325-353, Wilholt, "Bias and values," 92-101.

[38] Ryan, Hotchkiss, Crofton, and Gray Jr., "In Utero," 133-148.

As I understand, the pursuit of profit is a motivating factor for the chemical and similar research industries to take up novel research endeavors to make the world better from a social utility point of view. Hence, according to the functional approach, the quest for profit which is a non-epistemic value is consistent with the aims of the research, and it is legitimate. However, this particular non-epistemic value ceases to be consistent if it goes beyond its presumed functions by corrupting the research and undermines all other competing non-epistemic values such as social or moral values. It is evident that the influence of a non-epistemic value – quest for more profit – illegitimately influenced the research design, especially the choice of animal model for the experiment. This influence is illegitimate because the primary aim of the research was undermined by certain non-epistemic factors such as industrial profit which also initiated more non-epistemic consequences such as adverse health effects and the increase in social costs. Hence, the influence of commercial values, in this case, is not consistent with the primary aim of the research and therefore, illegitimate.

I have already underscored the point that since non-epistemic values perform heterogeneous functions, they can influence scientific research both in legitimate and illegitimate ways. The consequences of the acceptance or rejection of a hypothesis can be seen in different ways. It is not the case that the knowledge one acquires through accepting or rejecting a hypothesis is not just limited to that particular research instance. It has large implications since such knowledge might be used for pursuing further related researches. Hence, it becomes the responsibility of the scientists that they may prevent the entry of any such illegitimate influence of values which might produce a corrupted research outcome.

Consider a hypothetical scenario in which non-epistemic values might legitimately influence the choice of more rational methodological decisions. In the case of BPA, non-epistemic values such as care for human and animal health, reduction of the financial burden to the society, etc. should have influenced the research while confirming the hypothesis because these values, although non-epistemic, are consistent with the primary aim of the research and hence, their involvement is legitimate. They are consistent with the primary aim of the research because had these values influenced the trial designs and the methodologies; the researchers would have alternatively designed the trials with appropriate positive control doses and alternate animal models. That is to say, when there is a problem of inductive risk, the employment of non-epistemic values should be made in such way that the chosen values may reduce the degree of inductive risk and also may promote the attainment of the pre-specified aims of the research. In the case of BPA, the functional approach clearly tips off why the involvement of care for human

health is legitimate, and the quest for profit is illegitimate. In this case, care for human and animal health would have prompted the researchers to take up a positive control test with an alternate model organism which, in turn, would have eventually led to the finding of the association between adverse health effects and exposure to the BPA. Hence, the employment of non-epistemic values of care for human and animal health is legitimate in this context. On the other hand, another non-epistemic value - the quest for profit – corrupted the research and impeded the attainment of the primary aim of the research because its involvement led to the production and distribution of distorted knowledge claims through the choice of inappropriate and biased methodologies and controls. Hence, the involvement of the quest for profit, in this context, is illegitimate.

Conclusion

This paper made an attempt to characterize epistemic and non-epistemic values in terms of their functions. Epistemic values are characterized as functionally homogeneous because every epistemic value fulfills two important functions which are necessary for knowledge production and evaluation. These are: (i) they act as criteria to evaluate the epistemic virtues a hypothesis ought to possess, and (ii) they validate scientific knowledge claims objectively. On the other hand, non-epistemic values are characterized as functionally heterogeneous because their functions vary. The rationale behind the legitimate as well as the illegitimate influence of non-epistemic values in scientific investigation is their functional heterogeneity. By critically examining BPA research, I argued that in certain research contexts, especially when scientists confront the problem of inductive risk, the involvement of non-epistemic values is necessary. However, I pointed out that necessary involvement does not necessarily imply legitimate involvement and hence, the illegitimate influence of non-epistemic values must be carefully eschewed from scientific inquiries.

To sum up, non-epistemic values such as moral, social, scientific, or commercial values can operate in many ways in different phases of scientific research, and because of their functional heterogeneity, their influence might turn out to be illegitimate sometimes. Therefore, distinguishing the differences among the functions the non-epistemic values perform during different phases of scientific research is necessary to assess whether the influence of these values is legitimate or illegitimate.[39]

ERRATUM NOTICE

Unfortunately, the name of one of the authors of the article „The Collapse Argument Reconsidered," published in *Logos & Episteme* 11, 4 (2020): 413-427, is misspelled as Morteza HAJHOSSEINI. The author's name is Morteza HAJIHOSSEINI. In addition, throughout the paper "An Argument for the Safety Condition," authored by Michael J. Shaffer and published in *Logos & Episteme* 8, 4 (2017): 517-520, the last name Nozick is misspelled as Nozik. We apologize for these errors.

NOTES ON THE CONTRIBUTORS

Christopher T. Buford is Lecturer at the University of Akron. His areas of specialization are epistemology, philosophy of mind, personal identity and his areas of competency are bioethics, logic, and history of modern philosophy. Contact: cb72@uakron.edu.

Filip Čukljević is Research Associate at the University of Belgrade. His primary research interest is in theory of knowledge, especially Wittgensteinian approaches to the problem of philosophical scepticism. He has also written on the Fregean philosophy of language, theories of truth and philosophy of psychoanalysis. His most recent publication is "The Problem of Cognitive Significance – a Solution and a Critique" (*Philosophy and Society*, 2018). He is currently working on an article about McDowell's and Brandom's views on observational knowledge. Contact: filipcukljevic@gmail.com.

Jonas Karge (Dresden University of Technology) is a PhD student and research assistant as part of the Computational Logic Group at the Institute of Artificial Intelligence. His research focuses on formal epistemology and its connections to decision theory as well as logic-based knowledge representation. In particular, he is interested in various forms of reasoning under uncertainty. Contact: jonas.karge@tu-dresden.de.

B.J.C. Madison is Associate Professor at the United Arab Emirates University. His main research interests lie in epistemology, and in related issues in the philosophy of religion and the philosophy of mind. The primary focus of his current research is on the internalism/externalism distinction in epistemology. Contact: brent.m@uaeu.ac.ae.

Ryan Ross is a PhD student at Johns Hopkins University. His primary interests are epistemology, philosophy of science, and ancient philosophy. His current research is on the epistemology of conspiracy theories. Contact: rross27@jhu.edu.

Michael J. Shaffer is a professor at Gustavus Adolphus College and an external member of the Munich Center for Mathematical Philosophy. He is also a fellow of the center for formal epistemology at Carnegie-Mellon University, a fellow of the Rotman Institute for Science and Values at the University of Western Ontario, a Lakatos fellow at the London School of Economics, a fellow of the University of Cologne's summer institute for epistemology and an NEH fellow at the University of Utah. His main areas of research interest are in epistemology, logic and the philosophy of science, and he has published more than fifty articles and book chapters on various topics in these areas. He is co-editor of *What Place for the A Priori?* (Open Court, 2011) and is the author of *Counterfactuals and Scientific Realism* (Palgrave-MacMillan, 2012), *Quasi-factive Belief and Knowledge-like States* (Lexington, forthcoming) and *The Experimental Turn and the Methods of Philosophy* (Routledge, forthcoming). Contact: shaffermj66@outlook.com.

Joby Varghese is an Assistant Professor of Philosophy at the Department of Humanities and Social Sciences at the Indian Institute of Technology Jammu, India. He received his PhD from the University of Hyderabad. His research interests include philosophy of science, epistemology, and bioethics. He has published papers on science and values in *Synthese*, and *History and Philosophy of the Life sciences*. Contact: jobypvk@gmail.com.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the Notes to Contributors.

Contact: logosandepisteme@yahoo.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*rtf) or Microsoft Word document (*docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author`s affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in their

manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.

11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice şi Sociale „Gh. Zane" Academia Română, Filiala Iaşi, Str. Teodor Codrescu, Nr. 2, 700481, Iaşi, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Copyright & Publishing Rights

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License (). Authors have the right to use, reuse and build upon their papers for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme are* also freely available on Philosophy Documentation Center web page.