Knowledge-First Evidentialism and the Dilemmas of Self-Impact

Eyal Tal
Brandeis University
Paul Silva Jr.
University of Cologne

When a belief is *self-fulfilling*, having it guarantees its truth. When a belief is *self-defeating*, having it guarantees its falsity. These are the cases of "self-impacting" beliefs to be examined below. Scenarios of self-defeating beliefs can yield apparently dilemmatic situations in which we seem to lack sufficient reason to have any belief whatsoever. Scenarios of self-fulfilling beliefs can yield apparently dilemmatic situations in which we seem to lack reason to have any one belief over another. Both scenarios have been used independently to challenge Evidentialism, on which what we may rationally believe is all and only what fits our current evidence. Here we tie the two scenarios together and explore what a knowledge-sensitive evidentialist approach to one implies for the other.[1]

Keywords: Anti-expertise; Instability; Knowledge-first; Epistemic paradox, Evidentialism

1. Introduction

Evidentialism is the law of the land in a multitude of philosophical debates. For many working in contemporary epistemology it is a truism that our evidence about a proposition, and only our evidence about that proposition, determines what is rational and irrational to believe. Indeed, it is initially hard to see what could possibly threaten this claim. However, some things do. There is a range of existing objections to simple evidentialist theses stemming from the problem of naked statistical evidence (Buchak 2014), pragmatic encroachment (Fantl and McGrath 2002), moral encroachment (Basu 2019), and enkratic requirements (Worsnip 2018; Lasonen-Aarnio 2020). In what follows we will consider a relatively neglected form of objection to evidentialist theses stemming from *self-defeating* and *self-fulfilling* beliefs.

Scenarios involving self-defeating and self-fulfilling beliefs arguably reveal two more discomforts for Evidentialism. Roughly put, in cases of self-defeating beliefs we know that *p is true iff we don't believe p*. The so-called paradox of *anti-expertise* (Sorensen 1987) or *belief instability* (Kroon 1990; Lee 1998), which employs cases of self-defeating beliefs, presents the unsuspecting evidentialist with a challenge. Upon believing p the agent would have excellent reason to believe that p is false, and upon disbelieving p or suspending judgment the agent would have excellent reason to believe that p is true. The upshot is that no belief state can be stably had, and thus none seem rational regardless of what our current evidence regarding p supports. Meanwhile, roughly put, in cases of self-fulfilling beliefs we know that *p is true iff we believe p* (Foley 1991; Raleigh 2017). Upon believing p the agent would have excellent reason to believe that p is true, and upon disbelieving p the agent would have excellent reason to believe that p is false. The upshot is that different doxastic states seem equally justified, regardless of what our current evidence regarding p supports. Thus, both kinds of scenarios seem to drive a wedge between what we should believe and what fits our current evidence, contra Evidentialism.

[1] Special thanks are owed to the Alexander von Humboldt Foundation's generous financial support for the Cologne Center for Contemporary Epistemology and the Kantian Tradition (CONCEPT).

Interestingly, little attention has been given to the fact that cases of self-defeating and self-fulfilling belief are closely connected. The connection lies in that both kinds of cases involve the agent's knowledge (or justified belief) about how adopting certain beliefs would impact their truth. And it is this unique knowledge that figures into how the agent should go about forming beliefs in both kinds of cases. In light of this key similarity, we may expect an approach to one kind of scenario to imply something by way of an approach to the other. For instance, if a view of self-defeating cases had it that we should form all and only those beliefs that fit our current evidence, knowledge be damned, then that would suggest the same in self-fulfilling cases. Or if a view of self-fulfilling cases had it that we can believe all and only that which we know would result in knowledge, then that would leave very few candidate solutions to cases of self defeat.

In this paper we explore one answer to the challenge of self-defeating beliefs, and show that it suggests a neat solution to the challenge of self-fulfilling beliefs. We begin (§2) with what we consider to be the most formidable challenges to Evidentialism that self-defeating and self-fulfilling scenarios have to offer. Those are versions of the ones mentioned earlier, in which we know *p iff we don't believe p*, or *p iff we believe p*. We then (§3) review a knowledge-sensitive answer to the former (Conee 1987, 1994**),** according to which expected knowledge (or expected lack of it) affects what doxastic attitudes we should have. On this line, neither belief nor disbelief is justified in cases of self-defeating beliefs, and instead we should suspend judgment. We continue (§4) by showing how this answer suggests a related one to the challenge from self-fulfilling beliefs. The basic idea is that our evidence determines what attitude we should initially have, but as soon as we adopt that attitude we should change it to a different (and stable) one. We then consider the costs of the position, and find that they are quite high (§5). Specifically, we argue that if one can bear the cost of requiring agents to adopt doxastic states that they must immediately abandon, one may as well stick to a much simpler form of Evidentialism. Lastly (§6), we evaluate nearby knowledge-first alternatives that have been defended in different contexts.

2. Self-impacting beliefs

Our beliefs do not exist in a vacuum. Having them affects what other things we believe and how we are inclined to act. Sometimes our beliefs affect the very issues that they are about. A belief that we will do poorly on an exam may make us study harder and do well. A belief that our favorite presidential candidate is a shoe-in may make us less inclined to vote and thus hurt the candidate's chances. An optimistic belief that we will beat cancer may improve our odds of doing just that.

Troubles start when we have good reason to think that a belief is self-impacting in this way. When that happens, the agent finds herself in one of two general kinds of cases. In the first case, the agent rationally believes that a belief of hers hurts the odds of its content being true. In the second, the agent rationally believes that a belief of hers improves the odds of its content being true. But in-between cases are possible as well. We might learn that if we have only middling confidence in p then p is more likely to come true, but if our confidence in p is too low then p would definitely not come true. Indeed, the options are many, and philosophers have used a number of them to generate challenging examples. We now discuss a few of these examples and the views that they challenge.

2a. Self-defeating beliefs

Roughly speaking, self-defeating beliefs involve those confident doxastic attitudes the having of which negatively impacts their likelihood of being true. In mild cases, the impact is not great. For instance, when we believe that our favorite candidate is a shoe-in, the odds that the candidate would win are only slightly reduced. These cases are fairly common and are not especially noteworthy. The cases at

the extreme are more interesting. There, the impact of a belief on its likelihood is radical. For instance, a belief that we have no beliefs ensures that its content is false.

Self-defeating beliefs have been employed primarily to generate puzzling situations, in which no belief state seems rational. Consider, for instance, Earl Conee's case in this context, as well as Roy Sorensen's similar one:

> A thirtieth century brain physiologist, T, knows that all of a person's N-fibres fire if, and only if, the person judges that not all his N-fibres ever fire when he considers the matter for the first time. Knowing this, T considers for what he knows to be the first time: (a) All of T's N-fibres will fire. (Conee 1982: 57)

> ...it might be observed that a particular student passes when and only when he studies and studies iff he is not sure whether he will pass. This description implies that he will pass iff he is not sure that he will pass. How should the student react to the observation? (Sorensen 1987: 305)

Thought experiments such as these are meant to give us propositions that are false if we believe them, and true if we believe their negation. The prima facie upshot appears to be that the agent can occupy no belief state rationally. Upon adopting the relevant belief, the agent would have sufficient reason to think that it is false, and in the absence of the belief, the agent would have sufficient reason to think that it is true.[2]

The apparent result is indeed worrying. It is a problem for those who subscribe to a thesis like Uniqueness (White 2005), and even the weaker thesis of Optimism (Turri 2012). According to Uniqueness, every situation makes available to us exactly one doxastic attitude toward a given proposition to occupy rationally. According to Optimism, every situation makes available to us at least one rational doxastic attitude toward a given proposition. But if there are cases in which we know that a proposition is true iff we disbelieve it, then Uniqueness and Optimism appear to fall.

Perhaps even more worrying is the fact that the relevance of one's evidence seems to completely drop out of the picture in such cases. What our evidence in fact supports, i.e., how probable our evidence renders the proposition, does not figure in one bit. In other words, self-defeating beliefs present us with a challenge to the Humean dictum that we should proportion our belief in any proposition to what our evidence supports.

2b. Self-fulfilling beliefs

Opposite to self-defeating beliefs, self-fulfilling beliefs involve those confident doxastic attitudes the having of which positively impacts their likelihood of being true. For example, believing that one is writing an influential paper can motivate one to work hard and indeed write an influential paper. A similar phenomenon can occur in connection with the Placebo Effect, where believing one will get well could improve the likelihood of getting well. In such self-fulfilling cases, one's act of believing p

[2] We will not quibble over the details of such cases. Sorensen (1987) has objected that one can never have sufficient evidence to believe (or know) a biconditional of the form p iff I don't believe it. But this strikes us as implausible in the present case and others as well. For defense of the possibility of such evidence see Conee (1987), Richter (1990: 150ff), Kroon (1990), and Silva (2018).

boosts the odds of p being true. But there are also more extreme self-fulfilling cases, in which a belief *guarantees* its truth: as when one believes that one has at least one belief.

Self-fulfilling beliefs have also been employed to generate puzzling situations, in which no belief state seems preferable to another. We find recent examples of such cases in Gregory Anthill and Morten Dahlback: [3]

> Suppose an eccentric billionaire comes to your door with the following credible offer: she will give you a million dollars if you believe that you will be a millionaire. Should you believe that you will be a millionaire? (Antill 2020: 793)

> A powerful demon wants to show off his powers, and you have been selected to be his audience. You know that the demon never lies. He informs you that he will do the following: first, he will ask you to form a belief as to whether a coin will land heads or tails and use his mind-reading powers to determine what you believe. He will then flip the coin, and use his magical powers to make it the case that the coin lands heads if you believe that it will land heads, and tails if you believe that it will land tails. What should you believe? (Dahlback forthcoming: 2)

Thought experiments such as these are meant to give us propositions that are true iff we believe them. Here, the prima facie upshot appears to be that the agent can occupy either of two opposing belief states rationally. Upon adopting the relevant belief, the agent would have sufficient reason to think it is true, and in the absence of the belief, the agent would have sufficient reason to think that it is false.

This upshot too should bother defenders of Uniqueness, and this has indeed been one of the goals in recent debates involving self-fulfilling beliefs (Raleigh 2017; Drake 2017). But it should also bother those who think that we should adjust our beliefs to match what our evidence supports. For in cases of self-fulfilling beliefs it is our evidence that adjusts itself to match whatever we decide to believe.

In what follows we leave Uniqueness behind and focus on what self-defeating and self-fulfilling beliefs mean for the significance of our evidence. Specifically, we look into one often-taken route that vindicates the occasional sidelining of our evidence. While this route has been popular in discussions of self-defeating beliefs, it suggests a clear way forward in cases of self-fulfilling beliefs.

3. Self-Defeat and Knowledge-First Evidentialism

To see the pressure that cases of self-defeat can put on Evidentialism let's start with a case of self-defeat that takes its lead from John Buridan (*Sophismata*, Sophism 13), who observed the self-defeating character of believing *that one doesn't believe this very sentence*. Here is our elaboration on his idea:

> Galileo considers the following self-referential sentence:
>
> > B: I do not believe this sentence.
>
> Having never considered B before, Galileo knows that he doesn't yet believe it, and even as he continues to reflect on B he still doesn't come to believe it for the following reason: Galileo

[3] Drake (2017) mentions a number of discussions of self-fulfilling beliefs, including ones by Berker (2013), Foley (1991), Firth (1981), Greaves (2013), Reisner (2018), and Sharadin (2016).

> knows that if he believes B, then B is false and so he'll have a false belief; and if he refrains from believing B by believing not-B, then B is true and so he'll again have a false belief. In short, he knows that B is true iff he does not believe B. Given that he knows that he does not believe B, his further knowledge of this biconditional ensures that Galileo has conclusive evidence in support of the truth of B.

While this case of self-defeat depends on self-reference, Conee (1987: 324) has pointed out that we needn't rely on the phenomenon of self-reference to generate the puzzle here. We'll continue with this self-referential case for illustrative purposes, but everything we have to say about it can be easily applied to the original cases discussed above.

There are two questions this kind of case raises. One question concerns the extent to which cases of self-defeat problematize the connection between evidence and rationality. The second concerns the rational response in these cases of self-defeat. Conee (1987; 1994) has, to our minds, provided a formidable answer to both questions.

3.1 Self-Defeat and Withholding

Conee has argued that when it comes to known self-defeating beliefs in a situation like Galileo's, *withholding belief* is what rationality requires. He writes:

> It is rational for Galileo to withhold judgement on [B] even though it is evident to him that [B] is thereby made true. This is rational because on Galileo's evidence it has the highest epistemic value. No epistemic good available to him is sacrificed by withholding, and only by withholding does he avoid all epistemic mistakes. Instrumental considerations aside, *there is nothing rational in accepting something that one knowingly thereby makes false and there is nothing irrational in not accepting something that one knowingly thereby makes true*. Such a proposition is evident, but rationally unacceptable. (Conee 1987: 325–26, our italics)

In these words Conee suggests two arguments in defense of the idea that withholding belief is the only rational attitude for Galileo to adopt towards B.[4] The first is an argument connected to the instrumental value of (dis)belief versus withholding. We will leave this argument for the reader to assess, as it raises complex issues concerning the relationship between rational attitudes and epistemic utility. Instead, we focus on Conee's second argument, which seems to appeal to the following principle:

> *Foreseen False Belief (FFB)*. If S knows that believing p will ensure that S's belief that p is false, then it is not rational for S to believe p.

Notice how FFB leads us to the conclusion that Galileo should withhold belief about whether B is true. For Galileo knows that were he to believe B then B would be false, and also knows that were he to believe ¬B then ¬B would be false. Accordingly, Galileo knows that he is incapable of truly believing B and incapable of truly believing ¬B. There is only one (outright) doxastic state towards B that is not eliminated by FFB, namely, withholding belief. And, as Conee argues, withholding seems like the rational state to be in with respect to B when belief and disbelief have been ruled out.

[4] We will not defend a particular understanding of withheld judgment, except as an outright doxastic attitude distinct from the lack of any doxastic attitude. See Boghossian (2008: 477).

3.2 Self-Defeat and Evidentialism

Now, cases of self-defeat and FFB threaten to generate both contradictions and dilemmas of rationality if we take on-board a simple evidentialist principle:

> *Simple Evidentialism (SE)*. It is rational for S to believe p on their total evidence E iff E strongly supports p.[5]

If SE is true, then *it is rational for Galileo to believe B*. But if FFB is true, then *it is not rational for Galileo to believe B*. This is a contradiction. It is also a dilemma of rationality in the sense that Galileo is rationally permitted to believe B per SE, but also rationally required not to believe B per FFB.[6]

A few comments about SE before proceeding. First, we do not distinguish between justification and rationality. For us 'rationality' refers to the intended normative component traditionally associated with knowledge. Second, the left-hand-side of SE limits it to cases where evidence is what makes one's belief rational. This makes SE consistent with non-evidentialist views that allow for rational beliefs that are not grounded in evidence. Third, while SE is insensitive to issues involving pragmatic encroachment, statistical evidence, and context sensitivity, we would not begrudge anyone who wanted to add such qualifications. The problems raised by self-defeating and self-fulfilling beliefs will impact qualified instances of SE that are sensitive to such issues. Fourth, we assume that E is a set of propositions and that E supports p iff E implies p or renders p probable.[7]

3.3 A Path to Knowledge-First Evidentialism

Later versions of Evidentialism advanced by Conee and Feldman appear to resemble SE. But before those, Conee did not seem to endorse SE. Drawing an important and often overlooked distinction between *evidential support* and *having sufficient reason to respond* in a given way, Conee writes:

> Galileo's rational course is to withhold judgement on B. To see that this is so, we should attend to a distinction. It is one thing for a person to have evidence that establishes the truth of a proposition. It is another thing, not quite equivalent, for the person to have sufficient epistemic grounds for acting so as to accept the proposition. (Conee 1987: 322)[8]

---

[5] In Feldman and Conee's (1985) classic defense of Evidentialism, they put forth a view on which justified beliefs are all and only beliefs that *fit* a person's total evidence. In (2004) and (2018), they take Evidentialism to be the view that "Believing is the justified attitude when the person's evidence on balance supports a proposition, disbelieving the justified attitude when the person's evidence on balance supports the negation of a proposition, and suspension of judgment is the justified attitude when the person's evidence on balance supports neither a proposition nor its negation." (2018: 75).

[6] This is admittedly not a dilemma of rationality in the usual sense where an agent is both forbidden and required to take an attitude towards p, but it is no less disturbing. We could get a dilemma of this latter sort by strengthening SE so that one is *rationally required* (not merely permitted) to believe p if their total evidence strongly supports it.

[7] The implication relation should be understood widely to involve material implication (if E then p), counterfactual implication (were E true p would be true), and strict implication (necessarily, if E then p). Other conditionals (e.g. indicatives, strict implication with different senses of 'necessity') that would allow for valid inferences from E to p should likewise be regarded as sufficient for evidence E to support p. We would not begrudge someone who wanted to include abductive or explanatory relations that legitimate inferences from E to p–though our inclination is to interpret the evidential support for p that is provided by its ability to explain some salient data set as a kind of probabilistic support. Nothing will turn on this issue in what follows.

[8] For similar observations see Foley (1991) and Silva (2018).

The basic idea here is that although typically one's evidence supports a proposition p if and only if one's evidence provides sufficient reason to believe p, this is not a strict necessity. There are cases where these two relations come apart. One kind of case where these two relations have been argued to come apart involves situations in which one utterly lacks the ability to believe p despite one's evidence supporting it (Lord 2018; Littlejohn 2017). For how could one have sufficient reason to do what one cannot do? But cases of self-defeat provide another kind of example. For how could it be rational for one to believe p when one knows that believing p would ensure that p is false? FFB seems to give us the intuitively correct verdict that it is indeed not rational to believe p in such a case.

If we grant that strong evidential support for p and having sufficient reason to believe p can come apart, a question arises: When does having evidence that supports p ensure that an agent has sufficient reason to believe p? In answer, Conee writes:

> [I] Our doxastic goals do not automatically give us epistemic reasons. Epistemic reasons are not up to us. Epistemic reasons to believe are constituted by the belief's capacity to make a contribution to some genuine epistemic accomplishment. Knowledge is the definitive epistemic accomplishment. Epistemic reasons to believe are reasons that pertain to gaining knowledge. (Conee 1994: 478)

This remark suggests that if evidential support for p is to provide a thinker with reason to believe p, it must be possible for believing p to result in some kind of epistemic achievement. If knowledge is, as Conee suggests, the relevant epistemic achievement, we can revise SE in a fairly straightforward manner:

> *First-Order Position to Know Evidentialism (FOP).* S's total evidence E makes it rational for S to believe p iff (i) E strongly supports p *and (ii) S is in a position to know that p.*

FOP, however, is likely not where Conee would have us land. For directly following [I] he writes:

> [II] ... it is impossible for us to have epistemic reason to believe something that we know would not be supported by the balance of our evidence when we would believe it. When believing would result in a loss of crucial evidence for the believed proposition, adopting the belief would not bring about knowledge of the proposition. Foreseeing this sort of loss excludes having an epistemic reason to believe when contemplating the proposition. Adopting a belief that continues to be accompanied by evidence which is adequate for knowing the belief always constitutes a contribution to a pursuit of knowledge of the truth of the proposition. Foreseeing this internal relation to a pursuit of knowledge always provides an epistemic reason for adopting such a belief, come what may. (Conee 1994: 478)

Conee's remark here suggests that what is relevant when it comes to having epistemic reasons for belief is not so much *being in a position to know* on the basis of one's evidence, but rather lacking sufficient reason to think that one's evidence does not "contribute to" one's knowing p. This represents a kind of higher-order constraint on rational belief.

Assuming that a belief is rational iff one has sufficient reason to hold it, and assuming talk of evidence *not contributing to knowing* can be put in terms of the evidence *not putting one in a position to know*, we can formulate a revision to SE as follows:

Higher-Order Position to Know Evidentialism (HOP). S's total evidence E makes it rational for S to believe p iff (i) E strongly supports p *and (ii*) it is not rational for S to believe that S is not in position to know p.*

Note two points about FOP and HOP. First, recall that the Foreseen False Belief principle, FFB, was intuitively central to the explanation of why it is not rational for Galileo to believe B. FOP and HOP support this because each individually entails FFB.[9]

Second, FOP and HOP are inequivalent conditions. To see how, take their distinguishing conditions:

(ii) S is in a position to know that p (PKp)

(ii*) It is not rational to believe that one is not in position to know p (¬R¬PKp)

HOP's ¬R¬PKp does not entail FOP's PKp. For suppose PKp is false because we are in a Gettier case, or because p is false. Even so, ¬R¬PKp may yet be true since we might have strong but misleading evidence in support of the claim that we are in a position to know p. In such a case FOP implies that it is not rational to believe p, while HOP implies that it is rational to believe p. The fact that HOP has this implication highlights its connection with traditional epistemology, which ordinarily allows for rational false beliefs and rational true beliefs that do not amount to knowledge due to Gettier conditions.[10]

4. Self-Fulfilling Belief and Knowledge-First Evidentialism

Now turn to a case of self-fulfilling belief, where the act of believing p ensures that p is true and the act of disbelieving p ensures that p is false. No matter which belief the agent takes towards p, the belief is guaranteed to be true. In some cases, an agent knows that their belief is self-fulfilling. The cases of self-fulfilling beliefs that we are concerned with are those in which agents know that their beliefs are self-fulfilling. But cases of known self-fulfilling beliefs come in two general varieties: cases where one *knows what one will believe* and cases where *one does not know what one will believe*. It is the latter kind that is of interest, as the former is not epistemically challenging. In the former case one's evidence would support p and it would therefore clearly be rational to believe p. But in the latter case one's evidence does not (or at least need not) support p, and yet as some argue, it still seems rational to believe p just in virtue of knowing that believing p ensures p's truth.[11]

We can now apply the knowledge-sensitive FOP and HOP to cases of self-fulfilling beliefs to learn how we should handle them. Here, and unlike in cases of self-defeating beliefs, the evidence matters a lot. When we know that *p iff we believe p*, our evidence could either support p, support ¬p, or neither. Let us initially consider the first two options.

---

[9] For if one knows that were one to believe p then p would be false, then one knows that (and hence it is true that) one is not in a position to know p on E. This precludes both (ii) and (ii*) from being satisfied. Thus both FOP and HOP entail that it is not rational to believe p on E. The fact that FFB was already plausible and used in diagnosing the irrationality of Galileo believing B lends further support to FOP and HOP.

[10] Whether or not PKp entails ¬R¬PKp is more controversial, as it is tied up with one's views on higher-order defeat. For discussion see Lasonen-Aarnio (2014) and Benton and Baker-Hytch (2015).

[11] See Raleigh (2017), Reisner (2018) and Dahlback (forthcoming).

When our evidence in a self-fulfilling case supports p, FOP and HOP would permit us to believe p and p only. This is because p would be the only proposition to satisfy FOP and HOP's shared condition (i). Believing ¬p would not be permitted since (i) would fail for ¬p, as the evidence would not support ¬p. Similarly, when our evidence in a self-fulfilling case supports ¬p, FOP and HOP would permit us to believe ¬p and ¬p only.[12]

Now consider the third self-fulfilling case in which our evidence fails to support p over ¬p and vice versa. To put things probabilistically, suppose $Pr(p|E) = Pr(\neg p|E) = .5$. Now, FOP and HOP would recommend neither belief in p nor belief in ¬p, since condition (i) would fail for both p and ¬p. But recall that HOP recommended neither belief in p nor belief in ¬p in cases of self-defeating beliefs, and there Conee maintained that withholding judgment was the only rationally permitted option. For, as Conee argued, withholding seems like the rational state to be in with respect to B when belief and disbelief have been ruled out. So by parity of reasoning, here too, withholding judgment would be the only rationally permitted option. But notice that upon withholding judgment it becomes obvious that p is false, since p is true iff the agent believes p. At that point the agent would have to upgrade her withholding to full disbelief where she could stably remain.[13] So in such cases of self-fulfilling belief, FOP and HOP have us sometimes performing two doxastic revisions. As we will see in the next section, this two-step revision process costs FOP and HOP a critical dialectical advantage over SE.

5. FOP, HOP and Simple Evidentialism: A Wash?

SE has it that we should follow our evidence through thick and thin. Initially, the puzzles of self-defeating and self-fulfilling beliefs appear too much for this view to handle. With self-defeating beliefs, intuition has it that we may not believe p when we know that doing so guarantees ¬p, even if our current evidence supports p. With self-fulfilling beliefs, intuition has it that it would be irrational to miss out on knowledge due to an obsession with following our current evidence.

However, trying to add a knowledge-sensitive constraint to the evidentialist view, as FOP and HOP did, came partly at the cost of allowing agents to adopt a view despite knowing that they would have to change it upon adopting it. This was the two-step process described in the previous section, in regard to the self-fulfilling case where one's evidence supports neither p nor ¬p. But notice that this two-step implication of FOP and HOP is just like the implication that was taken to be an objection to SE. There, SE faced a case of self-defeat where one's evidence supports p, but believing p ensures that p is false. In that case, SE recommended believing p (because the evidence supports it), but then ceasing to believe p (because believing p ensures that p is false). So if FOP and HOP can survive their implications involving two-step revisions, why can't SE?

Some might think that SE can't survive such implications because in some self-defeat cases SE requires *constant* belief revisions. When the evidence in such a case initially supports p, SE would say: at t_1

[12] Here we already see a difference between FOP/HOP and those who take cases of self-fulfilling beliefs to bolster Permissivism (like Raleigh (2017), Dahlback (forthcoming)). Not only should friends of Uniqueness like this result, but so should those who abhor the thought of our evidence falling out of the picture of what we should believe. However, one would be hard-pressed to sell FOP/HOP to those who want the evidence about p to always play a role, since both FOP and HOP go against this idea in cases of self-defeat. So it is ultimately of no use to motivate FOP/HOP using their sensitivity to the evidence in self-fulfilling cases. We discuss the role of evidence in FOP and HOP soon.

[13] This result too saves the advocates of FOP and HOP from having to deny Uniqueness, as well as from having to completely sideline the evidence in cases of self-fulfilling beliefs.

believe p because your evidence supports it, but then upon believing p at t_2, cease believing p because your evidence now supports that it is false, but then after ceasing to believe p at t_3 your evidence support p once again and you should form the belief in p once more—and the process continues indefinitely. Consequently, the agent is left with no stable doxastic position to occupy, which is counterintuitive.

Yet two things remain unclear about such an account. First, it is unclear that cases of self-defeat genuinely make SE require constant belief revision. Second, and in the event that SE ultimately does require constant belief revision, we suspect that FOP and HOP may be in the same boat. We discuss these points now in turn.

Why is it unclear that SE genuinely requires constant belief revision? Take the case of Galileo and B = Galileo does not believe B. It must be asked *when* Galileo does not believe B. After all, B has some hidden temporal parameter. It could be a definite one: B_1 = Galileo does not believe B_1 *at* t_1. But if Galileo is not located at t_1 when his evidence supports B_1, then it is rational for him to believe B_1 and he need not revise his belief in B_1 upon having it. Alternatively, the hidden parameter could be an indexical: B_n = Galileo does not *now* believe B_n. Here it can seem as if Galileo must embark on the indefinite revision process described above. But this is no more irrational than having to revise one's beliefs about the present temperature or the present weather. For example, 'it's *now* raining' and 'it's *now* 70 degrees' change their truth value with the ever changing weather,[14] but there is nothing irrational about constantly updating one's beliefs in response to changes in the weather. Perhaps it is *impractical* for Galileo to spend his cognitive life keeping track of B_n's changing truth value. But then this would be just one more case where the practical and the rational part company.

One might insist that there is still a way to press SE into requiring constant belief revision. Perhaps this could be achieved using a self-defeat case involving a future proposition that we know a powerful being will render true if and only if we disbelieve it. Even if this is right, we believe, it would not obviously put SE at a disadvantage compared to FOP and HOP. For it seems possible to generate parallel issues for FOP and HOP. To see this, consider the credence framework for doxastic attitudes. Using it, we could design a self-fulfilling case that raises a similar constant-attitude-revision challenge for FOP and HOP, by stipulating a self-fulfilling case in which the objective odds of p continuously change. For instance, we could have a case in which we know that high credence in p entails p, low credence in p entails ¬p, and anything in between sets the objective chances of p by .0000001 + our current credence. Then, if our initial evidence deems p and ¬p roughly equally likely, FOP and HOP would require us to start from a middling credence in p, after which we would have to constantly increase our credence in p until we reach a high enough credence. If we make the increase small enough, the agent would have to revise her degree of belief very many times before reaching a stable high one.

It is of course possible to resist this argument by resisting the need for credence revision upon knowledge of increased objective odds. It is also possible to resist the credence model for doxastic attitudes, or by arguing that a finite (even if long) belief-revision process and an infinite one are importantly different. But that would take an argument that, as far as we know, has not been put forth. So, as things stand, it is our impression that SE's problematic implications in tough cases of self-defeat

[14] Of course, they may not literally *change* their truth value with the changing weather. On one view it is only the propositions that one entertains with that single sentence type which change, and each distinct proposition may have a different truth value from the distinct propositions one entertained with that single sentence type.

should not be considered a deathblow. They seem to be on a par with the implications of FOP and HOP in the two-step revision approach to cases of self-fulfilling beliefs. Thus, it appears that FOP and HOP have yet to make sense of these puzzling scenarios.

6. Simplified Knowledge-First Evidentialism

Recall that FOP and HOP imply that some cases of self-fulfilling belief are cases where it is initially *irrational* to believe p and *irrational* to believe ¬p. These are self-fulfilling cases where (a) we know that believing p ensures that p is true while believing ¬p ensures that p is false, and (b) our total evidence E fails to support p over ¬p and vice versa (e.g., $Pr(p|E) = Pr(\neg p|E) = .5$). The second condition, (b), is what is responsible for the failure of condition (i) of FOP and HOP. But, as we and others have noted, it seems fully rational to believe p just in virtue of knowing that believing p ensures p's truth even if our evidence does not currently support p. How could we know that believing that p cannot be mistaken, and yet it be irrational for us to believe p in light of that fact? Moreover, we observed that by prohibiting belief in p in these cases of self-fulfillment FOP and HOP lose their edge over SE.

Could a knowledge-first evidentialist avoid these costs? It appears so. We could expunge the evidential support condition (i) from FOP and HOP:

> *FOP-Simplified* (FOPS). S's total evidence E makes it rational for S to believe p iff *E puts S in a position to know that p.*
>
> *HOP-Simplified* (HOPS). S's total evidence E makes it rational for S to believe p iff *E ensures that it is not rational for S to believe that S is not in position to know p.*

FOPS and HOPS have been defended by others in the knowledge-first literature.[15] However, what has gone unobserved is the fact that they neatly avoid counterexamples stemming from cases of self-defeating and self-fulfilling beliefs. For by not including an evidential support constraint—as condition (i) does—the threat posed by cases of self-fulfilling beliefs is removed. In self-fulfilling cases, the evidence puts us in a position to know that whatever we believe we will know. So our evidence puts us in a position to know p, but it doesn't put us in a position to know p *by supporting the truth of p.* Our total evidence does not support the truth of p in the sense that E implies or probabilifies p. No such evidential support relation obtains in the kinds of self-fulfilling cases at issue. Further, notice that expunging condition (i) doesn't impact the ability of FOPS and HOPS to deal with cases of self-defeat. For we have already seen that the right hand sides of FOPS and HOPS are not satisfied in cases of self-defeat. Thus, FOPS and HOPS are just as well placed as FOP and HOP to survive the problems associated with self-defeating beliefs.

7. Concluding Reflections

We cannot here weigh all the costs and benefits associated with all the evidentialist principles we've been considering. But there are some issues we would like to highlight.

---

[15] Williamson (2013) and Littlejohn (2017) defend the view that doxastically justified belief just is knowledge. This leaves open the question of what propositional justification amounts to. For such knowledge-first theorists, FOP is an exceedingly natural stance to take and it is motivated on independent grounds (cf. Silva 2018: 2926; Sylvan 2018; Lord 2018). See Rosenkranz (2018) for defense and discussion of a theory of justification that is very close to HOP.

Are FOPS and HOPS evidentialist principles? Some might worry that without an evidential support constraint like (i), FOPS and HOPS would not be evidentialist in an important sense. The thought is that on both principles, the degree to which our evidence supports the relevant proposition p can fall out of the picture, as the principles have it in self-defeat cases. But this worry is misplaced. For what makes these two principles deeply evidentialist is their commitment to the evidence's fundamental role in determining what we should believe. Both principles take one's evidence to be responsible for either putting one in a position to know, or for putting one in a position where one could rationally believe that one is in a position to know. The only difference is that while typical evidentialist principles take our evidence's say regarding p's likelihood to determine what we should believe, FOPS and HOPS take our evidence's say regarding our position to know p to determine that.

Moreover, notice that once one adopts a doxastic state permitted by FOPS and HOPS, one's run-of-the-mill evidence would be one's reason for retaining that doxastic state. At that point, one's evidence would indeed probabilify p or its negation—in accordance with one's doxastic state. So again evidence plays a key role in sustaining rationality and knowledge. First, our evidence about what we can and cannot know guides what belief we should form, and second, our evidence about what is probable guides what beliefs we may retain or abandon. Since both kinds of evidence involve p, it is still our evidence concerning p and only our evidence concerning p that determines what is rational and irrational to believe.

Admittedly, it is tempting to think that this move comes at the cost of saying that what puts us in a position to know is not evidence *about p*. Rather, it is evidence about how likely p is given our attitudes. But knowing that p would be true if one were to believe p *is* evidence about p. It is evidence about p in virtue of being evidence about p's truth value in nearby worlds. Of course, this is not evidence about whether p is actually true. But it is hard to see why this is a cost, and if it is a cost, why it is one worth fretting over. After all, much of our evidence is indirect. We are often in circumstances where we know that p is true if some condition C obtains, and we can have more or less evidence that condition C obtains without knowing whether it actually obtains. In self-fulfilling and self-defeating cases the condition C is just about our attitudes.

Is FOPS Plausible? Like FOP, FOPS forbids rational false beliefs and rational true beliefs that are gettiered. That's something that epistemologists have tended to treat as a criterion for an adequate theory of rational belief. So FOPS has familiar (and serious) costs.

Is HOPS Plausible? Above we noted that HOP has more in common with traditional epistemology than FOP. HOP allows for rational false beliefs and rational true beliefs that don't constitute knowledge. This is also true of HOPS. However, HOPS has a potential problem that FOP lacks. For HOPS threatens to under-intellectualize rationality. Agents who lack the concept 'knowledge' or the more complicated concept of 'being in a position to know,' could arguably *never* be in a position where it is rational for them to believe that they are not in a position to know. At least this holds so long as *being in a position* to know that one is not in a position to know requires one to be able to *somewhat easily* know that one is not in a position to know (Williamson 2000, p. 95). If that is correct, HOPS turns out to be a strangely permissive evidentialist theory of rationality. For example, very young children (e.g. 1-2 year olds) would know a lot. At age 1 children can associate words with objects and locate them upon request. This ability implies possessing concepts of the relevant objects as well as some propositional knowledge involving them. But at such a young age these children don't have the concept 'knowledge' and certainly don't have the concept of 'being in a position to know'. Therefore, $\neg R \neg PKp$ is always satisfied for them, i.e., for any claim p, it is not rational to believe that one is not

in position to know p simply because they lack the concept 'being in a position to know' p. So HOPS seems to imply that it is rational for children to believe *anything* whatsoever. That is, at the very least, a surprising implication.

References

Antill, G. (2020). "Epistemic freedom revisited". *Synthese* 197 (2): 793–815.
Basu, R. (2019). "What We Epistemically Owe To Each Other". *Philosophical Studies* 176 (4): 915–931.
Benton, M. and Baker-Hytch, M. (2015). "Defeatism Defeated". *Philosophical Perspectives* 29: 40–66.
Berker, S. (2013). "The Rejection of Epistemic Consequentialism". *Philosophical Issues* 23(1): 363–387.
Boghossian, P. (2008). "Epistemic Rules". *The Journal of Philosophy* 105(9): 472–500.
Buchak, L. (2014). "Belief, credence, and norms". *Philosophical Studies* 169(2): 285–311.
Buridan, J. (1982). John Buridan on Self-Reference: Chapter Eight of Buridan's 'Sophismata', G. E. Hughes (ed. & tr.), Cambridge: Cambridge University Press.
Conee, E. (1982), "Utilitarianism and Rationality". *Analysis* 42(1): 55–59.
Conee, E. & Feldman, R. (1985). "Evidentialism". *Philosophical Studies* 48: 15–34.
Conee, E. (1987). "Evident, but Rationally unacceptable". *Australasian Journal of Philosophy* 65: 316–326.
Conee, E. (1994). "Against an Epistemic Dilemma". *Australasian Journal of Philosophy* 72: 475–481.
Conee, E., & Feldman, R. (2004). "Afterword". In E. Conee & R. Feldman (Eds.) *Evidentialism: Essays in epistemology* (pp. 101–107). New York: Oxford University Press.
Dahlback, M. (forthcoming). "Infinitely Permissive". *Erkenntnis*.
Drake, J. (2017). "Doxastic Permissiveness and the Promise of Truth". *Synthese* 194 (12): 4897–4912.
Fantl, J. & McGrath. M. (2002). "Evidence, Pragmatics, and Justification". *Philosophical Review* 111 (1): 67-94.
Firth, R. (1981). "Epistemic Merit, Intrinsic and Instrumental". *Proceedings and Addresses of the American Philosophical Association* 55: 2–23.
Feldman, R. & Conee, E. (2018). "Between Belief and Disbelief". in *Believing in Accordance with the Evidence: New Essays on Evidentialism*, edited by K. McCain. Synthese Library: Studies in Epistemology, Logic, Methodology and Philosophy of Science 398: 71-89.
Foley, R. (1991): "Evidence and Reasons for Belief". *Analysis* 51(2): 98–102.
Greaves, H. (2013). "Epistemic Decision Theory". *Mind* 122(488): 915–952.
Kroon, F. (1990). "On a Moorean Solution to Instability Problems". *Australasian Journal of Philosophy* 68: 455–61.
Lasonen-Aarnio, M. (2014). "Higher-order evidence and the limits of defeat". *Philosophy and Phenomenological Research* 88(2): 314–345.
Lasonen-Aarnio, M. (2020). "Enkrasia or Evidentialism? Learning to Love Mismatch". *Philosophical Studies* 177 (3): 597–632.
Lee, B. (1998). "The Paradox of Belief Instability and a Revision Theory of Belief". *Pacific Philosophical Quarterly* 79: 314–328.
Littlejohn, C. (2017). 'How and Why Knowledge is First.' In A. Carter, E. Gordon & B. Jarvis (eds) *Knowledge First*. Oxford: Oxford University Press. 19–46.
Lord, E. (2018). *The Importance of Being Rational*. Oxford: Oxford University Press.
Raleigh, T. (2017). "Another Argument Against Uniqueness". *The Philosophical Quarterly* 67(267): 327–346.
Reisner, A. E. (2018). "Pragmatic Reasons For Belief". In *The Oxford Handbook of Reasons and Normativity*. Ed. Daniel Star. Oxford University Press.
Richter, R. (1990). "Ideal Rationality and Hand-Waving". *Australasian Journal of Philosophy* 68: 147–156.
Rosenkranz, S. (2018). "The Structure of Justification". *Mind* 127: 309–338.
Sharadin, N. P. (2016). "Nothing but the Evidential Considerations?". *Australasian Journal of Philosophy* 94(2): 343–361.
Silva, P. (2018). "Explaining Enkratic Asymmetries: Knowledge-First Style". *Philosophical Studies* 175 (11): 2907–2930.
Sorensen, R. (1987). "Anti-expertise, Instability, and Rational Choice". *Australasian Journal of Philosophy* 65: 301–15.
Sylvan, K. (2018). "Knowledge as a Non-Normative Relation". *Philosophy and Phenomenological Research* 97(1): 190–222.
Turri, J. (2012). "A Puzzle About Withholding". *Philosophical Quarterly* 247: 355–364.
White, Roger. (2005). "Epistemic Permissiveness". *Philosophical Perspectives* 19: 445–459.
Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.
Williamson, Timothy (2013). Knowledge First. In (eds M. Steup and J. Turri), *Contemporary Debates in Epistemology* (2nd ed.). Oxford: Blackwell: 1–9.
Worsnip, A. (2018). "The Conflict of Evidence and Coherence". *Philosophy and Phenomenological Research* 96(1): 3–44.