Closing the Case on Self-Fulfilling Beliefs

Chad Marxen

Abstract Two principles in epistemology are apparent examples of the close connection between rationality and truth. First, adding a disjunct to what's rational to believe yields a proposition that's also rational to believe. Second, what's likely if believed is rational to believe. While these principles are accepted by many, it turns out that they clash. In light of this clash, we must relinquish the second principle. Reflecting on its rationale, though, reveals that there are two distinct ways to understand the connection between rationality and truth. Rationality is fundamentally a guide to the belief-independent truth, rather than a guide to acquiring true beliefs. And this in turn has important implications for current discussions of permissivism, epistemic reasons, and epistemic consequentialism.

Keywords epistemology, epistemic consequentialism, self-fulfilling beliefs, the aim of belief, the truth connection, epistemic closure

1. The Clash

Rationality and truth are closely connected. Since we'll be focusing on this connection—which involves epistemic (rather than, for example, practical) rationality—the kind of rationality under consideration will always be the epistemic kind, unless noted otherwise. Now admittedly, a complete understanding of this connection is difficult to attain. But reflecting on a couple of epistemological principles—that are apparent examples of this connection—might significantly contribute to our understanding of it. To this end, here's the first principle:

> *Addition Closure*: if p is rationally permissible for you to believe, then (p or q) is rationally permissible for you to believe.[1]

Employing the familiar distinction between *ex ante* and *ex post* rationality [Goldman 1986]—that some proposition is rational to believe is a matter of *ex ante* rationality. The propositions that are ex ante rational to believe may differ from the propositions that are *ex post* rationally believed—the propositions that are believed *in a rational manner*. Whereas ex ante rationality is focused on *potential beliefs*, ex post rationality is focused on *actual beliefs*. Addition Closure solely concerns *ex*

[1] Hereafter, 'rational' will be used to discuss rational permissions rather than requirements, unless noted otherwise.

ante rationality. This distinction can also be expressed as the distinction between propositional justification and doxastic justification. If Addition Closure were expressed with this terminology, it would say that if p is propositionally justified for you to believe, then (p or q) is propositionally justified for you to believe.

Intuitively, if p is rational to believe and p obviously implies q, then q is also rational to believe. Addition Closure follows from this intuitive principle since p obviously implies (p or q). In a similar vein, Dylan Dodd [2012: 340. Italics mine] writes:

> Intuitively, if one has justification for believing *p* and is in a position to know that *p* entails *q*, then whatever provides one with justification for *p* provides justification for *q* too—after all, *p* entails *q*! . . . *virtually all epistemologists agree that something like this is correct* . . .[2]

Now here's the second principle:

Rationality Seeks True Beliefs (RSTB): if p is likely if you believe it, then p is rational for you to believe.[3]

'Likely' should be understood as likely on the agent's evidence. The fundamental motivation behind RSTB is that—as Laurence BonJour [1985: 7–8] expresses in the following—rationality (justification) is a guide to truth:

> [T]he goal of our distinctively cognitive endeavors is *truth* . . . bringing it about that our beliefs are epistemically justified will also tend to bring it about . . . that they are true . . . It is only if we have some reason for thinking that epistemic justification constitutes a

[2] For those committed to the view that (ex ante) rationality, justification, or warrant—is closed under Addition—see Feldman [1995], Okasha [1999], Pryor [2004], Wright [2004], Vogel [2008], Dodd [2012], Ichikawa and Jarvis [2013], and Whiting [2018]. Dretske [1971] and Rosenkranz [2007] accept the analogue of this view concerning knowledge or being in a position to know.

[3] A similar principle says that if p is likely if believed and p is at least as likely if believed as ~p is if believed, then p is rational to believe. But for our purposes we need not choose between these principles since our arguments apply to both. Thanks to an anonymous referee for mentioning this worry.

> path to truth that we as cognitive beings have any motive for preferring epistemically justified beliefs . . .[4]

This idea has been appreciated for quite some time; William James [1912: 116–7], for example, relies on it with his alpine climber case:

> [Belief] must continually outstrip scientific evidence . . . there is a certain class of truths of whose reality belief is a factor as well as a confessor; and that as regards this class of truths faith is . . . licit . . .
>
> Suppose . . . I am climbing in the Alps, and have had the ill-luck to work myself into a position from which the only escape is by a terrible leap. Being without similar experience, I have no evidence of my ability to perform it successfully; but hope and confidence in myself make me sure I shall not miss my aim, and nerve my feet to execute what without those subjective emotions would perhaps have been impossible . . . In this case (and it is one of an immense class) the part of wisdom clearly is to believe . . .[5]

James argues that it's antecedently rational for the climber to believe that he can successfully make the leap even though his evidence does not make this proposition likely.[6] Inspired by James, David Velleman [1989: 63] similarly writes:

> Does [rationality] require a rule that one shouldn't form a belief without prior evidence? I say no--at least, not if one has evidence that the belief would be true if one formed it.

---

[4] For similar thoughts, see: Cohen [1984], Schoenfield [2015], and Horowitz [2019]. Another motivation for RSTB is the idea that the aim of belief is truth. For discussions of this idea, see: Velleman [2000], and Wedgwood [2002].

[5] Setting aside James and a few who've discussed this case, I'm not aware of anyone—until Velleman [1989]—who relies on RSTB's motivation to then apply RSTB to this kind of case. But beginning with Velleman, this idea has gained serious momentum—as I mention in fn. 7 and discuss more fully in section 6.

For those who've discussed James' case and are sympathetic with his verdict, see Hocking [1912], Perry [1938], and Taylor [2002]. Although it isn't clear what Santayana's [2009] ultimate verdict is, he seemingly disagrees with James. While you might think Peirce does too, Hookway [2002] argues that Peirce plausibly sides with James. Finally, Moore [1907] and Doore [1983] don't take a stand on James' verdict even though they discuss his case.

[6] Although there's an alternative interpretation where James is only claiming that the given proposition is practically rational or ethical to believe, we'll follow Velleman.

> Why would rules designed to help one arrive at the truth forbid one to form a belief that would be true? . . . Here I am subscribing to a view put forward, most famously, by William James.

But while RSTB and Addition Closure are accepted by many, it turns out that there are powerful reasons to believe they're incompatible.[7] To illustrate this incompatibility, consider the following case:

> **Logic Class**. Friedrich is attempting to pass logic again. Given Friedrich's poor performance in his philosophy classes, the proposition that his philosophical abilities are *not* prodigious is rational for him to believe. He's narcissistic, though, so he irrationally believes that his philosophical abilities *are* prodigious. Applying the Addition inference rule to what he believes, he comes to believe that his philosophical abilities are prodigious or today will be terrible.
>
> While he has come to believe this proposition, he hasn't come to believe that today will be terrible. He knows that the chance that today will be terrible is minuscule since—in keeping with his self-absorbed ways—he meticulously records how his days go. While the chance is quite low, there have been some terrible days for him over the past few years. And almost always, the sole reason why these days are terrible is that he firmly believed that the day would be terrible: whenever he does, this belief makes him depressed, which in turn makes the day terrible. Thankfully, Friedrich is generally cheerful, and has strong reason to think that he will, as usual, avoid the pessimistic belief that today will be terrible.

[7] RSTB and similar principles frequently come up in discussions surrounding three epistemological topics: permissivism, epistemic reasons, and epistemic consequentialism. See section 6 for who—besides Velleman and James—commits themselves (in these discussions) to these principles, and how our response to the apparent incompatibility between RSTB and Addition Closure bears on these discussions.

Let T be the proposition that today will be terrible. While T is unlikely overall, T is likely if Friedrich believes it. So if RSTB holds, then T is rational for Friedrich to believe. Let P be the proposition that Friedrich's philosophical abilities are prodigious. According to Addition Closure, if T is rational for Friedrich to believe, then (T or P) is rational for him to believe too. So if RSTB and Addition Closure are both true, then (T or P) is rational for him to believe.

But is (T or P) rational for Friedrich to believe? It is not. (T or P) is unlikely*, even if he believes it*. Given how infrequent his terrible days are, it's quite unlikely that T. And given his difficulties with philosophy, it's quite unlikely that P. So (T or P) is unlikely overall.

And Friedrich's believing (T or P) would not make it likely. To see this, notice that Friedrich actually does come to believe (T or P), but it remains unlikely. When he comes to believe it, he infers it from P, and does not come to believe T. Further, he does not acquire any reason for believing that he will come to believe T. So T remains unlikely. As T and P are each unlikely, even after Friedrich believes (T or P), it is unlikely if he believes it. And since this is the case, it's plausible that (T or P) is irrational for Friedrich to believe.[8] So RSTB and Addition Closure are incompatible.[9]

Not only is there a clash between these principles, but there's also a clash between their credal analogues. According to the credal analogue of RSTB, if T is likely if Friedrich has a high credence in it, then a high credence in T is rational for him to have. And according to the credal analogue of Addition Closure, if a high credence in T is rational for Friedrich to have, then a high credence in (T or P) is rational for him to have too.

[8] Presumably Velleman too would find it plausible that (T or P) is irrational for Friedrich to believe. As he [1989: 63] says, 'This purpose [having accurate beliefs] may well require a rule that one shouldn't retain a belief unless one has evidence of its truth'.

[9] For another way to see that there's a clash, notice that just as it would be *practically* irrational for Friedrich to bet on his *epistemically* irrational belief that P (since he has no good reason to believe that P is true), it would similarly be practically irrational for Friedrich to bet on his belief that (T or P) as he has no good reason to believe that (T or P) is true. See Marxen [2021a] for the implications of this connection between epistemic and practical rationality.

But we can extend **Logic Class** so that T is likely if Friedrich has a high credence in it. Additionally, we can suppose that he comes to have a high credence in (T or P) because of his irrationally high credence in P. And when he comes to have this high credence in (T or P), each of T and P remains unlikely. So while T is likely if Friedrich has a high credence in it, (T or P) is unlikely if Friedrich has a high credence in it. As (T or P) is unlikely, even if he has a high credence in it, a high credence in it is irrational for Friedrich to have. So the incompatibility between RSTB and Addition Closure extends to their credal analogues.

2. Do the Principles Really Clash?

Here's the puzzle: do we give up RSTB, Addition Closure, or the claim that these principles are incompatible? Since RSTB and Addition Closure are intuitive and accepted by many, it's worth exploring the solution which says that the two principles can be reconciled.[10] According to it, despite (T or P) being unlikely even if Friedrich believes it, it is somehow rational for him to believe. The reason: Friedrich has a certain other way to believe it, and if he were to believe it in this way, he would rationally believe it. Here's the way: Friedrich would first come to believe T, making T quite likely. And then, rationally believing T, he would use Addition to infer (T or P). If he were to acquire a belief that (T or P) in this way, his belief would be rational.

But why is it plausible that if Friedrich has a way to acquire a rational belief that (T or P), then it is rational for him to believe? As John Turri [2010: 319] says, 'It is standardly assumed that there is some deep and important relationship between propositional and doxastic justification.' This standard assumption might naturally lead to the following view: p is propositionally justified or rational to believe iff there's a way to acquire a doxastically justified or

[10] See Hetherington [1996] for a notion of evidence where T is likely overall simply because it's likely if believed. On this notion of evidence, T does not generate a clash between RSTB and Addition Closure. But we'll be working with a more standard notion of evidence where T is not likely overall. Additionally, it's noteworthy that an analogous version of the puzzle plausibly arises on the alternative notion of evidence. After all, it's intuitive that if T is evidentially supported, then (T or P) is too. But (T or P)—unlike T—is not likely if believed.

rational belief that p. Let's call this view *the Ante-Post Connection*.[11] According to the solution under consideration, if the Ante-Post Connection holds, then (T or P) is rational for Friedrich to believe.

While this solution may seem promising, it faces a couple of problems. First, it's unclear whether there's a plausible version of the Ante-Post Connection that supports the claim that (T or P) is rational for Friedrich to believe.[12] To this end, consider the following case from Turri [2010: 321–2]:

> [A] first-year logic student . . . hasn't yet proven either of DeMorgan's Laws . . . He currently possesses a means such that, were he to employ it, he would come to justifiedly believe that Law. That much seems right. Notice, however, that this implies that, even *before* he goes through the proof, he possesses a means such that, were he to employ it, he would justifiedly believe < I have proven one of DeMorgan's Laws >: he just needs to go through the proof, [and] reflect on what he just accomplished . . .

Let's call Turri's case **Potential Proof**, and let's call D the proposition that the logic student has *demonstrated* one of DeMorgan's Laws. As Turri says, the student has a way to acquire a justified or rational belief that D: he can go through the proof, reflect, and then believe D.

While Turri endorses a view similar to the Ante-Post Connection, he acknowledges that no plausible view implies that D is rational for the student to believe. For plausible versions of the Ante-Post Connection, the student's way to acquire a rational belief that D doesn't count. What's important about the student's way to acquire a rational belief that D is that he acquires new evidence for D during his process of acquiring a rational belief that D. Since his evidence changes during this process, what's rational to believe changes during this process too.

---

[11] For views in the neighborhood of the Ante-Post Connection, see Goldman [1979] and Turri [2010].

[12] In so far as the arguments hereafter cast doubt on the Ante-Post Connection, they only bear on one direction of it (if there's a way to acquire a rational belief that p, then p is rational to believe).

But notice that this evidential change in **Potential Proof** also occurs during Friedrich's possible way of acquiring a rational belief that (T or P). When Friedrich comes to believe T, he acquires new evidence for T. Given his strong background evidence that he will be depressed if he believes T, along with his new evidence that he believes T, T becomes quite likely. So Friedrich comes to have new evidence to believe (T or P). Given the evidential diagnosis of **Potential Proof**, it looks like Friedrich's way of acquiring a rational belief that (T or P) does not count as making it rational for him to believe that (T or P) in his present situation.

To avoid this first problem for the solution, you might think that there's a relevant difference between the student's way of acquiring a rational belief that D and Friedrich's way of acquiring a rational belief that (T or P). To illustrate this difference, notice that I have a way to acquire a rational belief that I will be in serious pain in a few seconds: I can firmly kick a nearby wall. But this proposition is not thereby rational for me to believe. What's significant about my way to acquire a rational belief in this proposition is that it involves an action that creates good evidence for the relevant proposition. Plausible versions of the Ante-Post Connection will only count *purely doxastic* ways of acquiring rational beliefs. While Friedrich's way of acquiring a rational belief that (T or P) is purely doxastic, the student's way of acquiring a rational belief that D is not; the student's way involves the action of proving something.

While this response might help with **Potential Proof**, there are cases where it will not help. For example, consider a case where Bill has good testimonial evidence for the proposition that climate change is real. Despite having good evidence for this proposition, he doesn't believe it. Further, Bill has good evidence that no one else in his immediate family believes this proposition. Everyone in his family has long held that climate change is a hoax, propagated by China to hurt U.S. manufacturing.

Clearly, the proposition that someone in your family believes that climate change is real is not rational for Bill to believe. Nevertheless, Bill does have a purely doxastic way of acquiring a rational belief in this proposition. Here's the way: he can believe that climate change is real and

then infer that someone in his immediate family believes that climate change is real. While he has this purely doxastic way of acquiring a rational belief that someone in his family believes that climate change is real, it's clear that this proposition is not rational for him to believe in his current situation. So the first problem for the solution under consideration remains: it does not look like a plausible version of the Ante-Post Connection supports the solution.[13]

The second problem for the solution is that an extended version of **Logic Class** reveals that the solution fails to address the *core* of the puzzle:

> **Silver Lining**. If today will be terrible, there's a silver lining for Friedrich. Whenever Friedrich has a terrible day, the following day is spectacular. Since he's experienced this sequence of events many times, it's rational for him to believe that if today will be terrible, then tomorrow will be spectacular. Nonetheless, he doesn't believe this conditional since it's too painful for him to reflect on his unpleasant memories. He also doesn't believe that tomorrow will be spectacular. But unlike the proposition that today will be terrible, the proposition that tomorrow will be spectacular is unlikely even if Friedrich believes it. Friedrich knows, though, that if he comes to believe that tomorrow will be spectacular, then the proposition that today will be terrible will no longer be likely if he believes it. Believing that tomorrow will be spectacular would fill him with optimism, obstructing the self-fulfilling power of believing that today will be terrible.

Let S be the proposition that tomorrow will be spectacular. It's plausible that S is rational for Friedrich to believe, if T is rational for him to believe. To see that this is true, let's assume that T is rational for him to believe. According to **Silver Lining**, it is also rational for him to believe that if T, then S. Given Friedrich's ability to employ modus ponens, S is rational for him to believe too.

[13] Of course, if you just come to believe that someone in your family believes that climate change is real, it will not be true. So RSTB does not imply that this proposition is rational for you to believe.

But is S rational for him to believe? It is not. S is unlikely, even if Friedrich believes it. Further, Friedrich does not have a way to acquire a rational belief that S. Let's consider these points in turn. As Friedrich has good reason to believe that tomorrow will be a typical day, S is unlikely. Further, Friedrich knows that the belief that S is not self-fulfilling. So S is unlikely, even if Friedrich believes it.

But perhaps Friedrich has a way to acquire a rational belief that S. If he does, there's only one way: coming to believe both T and also (if T, then S), and then inferring S. But here's the problem: if he were to believe S in this way, he wouldn't rationally believe it. Admittedly, Friedrich has a way to acquire a rational belief that T and a way to acquire a rational belief that if T, then S. But he doesn't have a way to acquire a rational belief that S.

But how could that be? The reason: Friedrich knows that the belief that S obstructs the self-fulfilling power of the belief that T. So as soon as he infers S, he doesn't rationally believe T. Admittedly, T is likely if Friedrich believes it by itself. But T is unlikely if Friedrich believes T while also believing S.[14] So Friedrich irrationally believes T in this situation where he also believes S. But since Friedrich rationally believing S depends on him rationally believing T, he irrationally believes S in this situation. Since this way of acquiring a belief that S does not result in a rational belief, Friedrich has no way to acquire a rational belief that S. So the Ante-Post Connection does not imply that S is rational for him to believe. Therefore, contra the solution under consideration, S is irrational for Friedrich to believe.

Silver Lining reveals that the solution fails to address the *core* of the puzzle: RSTB is incompatible with the rationalizing power of logical consequence: many propositions are rational

[14] For an analogous situation, consider the following: that you will make it to work on time is likely if you catch your bus, but that you will make it to work on time is unlikely, if you catch your bus and your bus happens to get a flat tire. Are these claims consistent? They are. The reason: it's unlikely that your bus will get a flat tire, if you catch it. And for **Silver Lining**, it's unlikely that Friedrich will believe S, if he believes T.

to believe because they are logical consequences of what's rational to believe. Addition Closure is just one example of the rationalizing power of logical consequence.

3. The Extent of the Clash

Even if the principles clash—you might wonder how significant this is—perhaps they rarely clash? But as we've already heard from James, propositions like T—propositions that are not likely overall, but likely if believed—are commonplace [Carr 2017; Marxen 2021b]. And as the following case will help show, the principles usually—if not always—clash concerning these propositions.

> **New Toothbrush**. Whitney just purchased a toothbrush. She typically uses new toothbrushes for a couple months. Given her history, she knows it's quite likely that she will use her new toothbrush again. But if she believes that she will never use her new toothbrush again, she will throw it into the dirty wastebasket, in which case she will never use it again. So it's quite likely that she will never use it again if she believes that she will never use it again.

Let N be the proposition that Whitney will *never* use her new toothbrush again. While N is unlikely overall, it is likely if she believes it. So if RSTB holds, N is rational to believe. But we can show—using the method we applied to **Logic Class**—that if Addition Closure holds, then N is not rational to believe.

We begin by finding one of Whitney's unjustified beliefs. If Whitney is like the rest of us, she has many unjustified beliefs about the things she loves (e.g. friends, family, beloved sports teams). So let's pick a typical unjustified belief of hers: the belief that her beloved team will win the Super Bowl. And let's call W the content of her belief.

Next, we consider (N or W), an Addition-consequence of N. (N or W) is chosen in such a way that it is unlikely, since N and W are each quite unlikely. Further, (N or W) is unlikely if she believes it. As Whitney's beliefs do not increase her team's chance of success, W is quite unlikely even if she believes (N or W). N is also quite unlikely if she believes (N or W). The

reason: N is unlikely overall, and coming to believe (N or W) doesn't make it likelier that she comes to believe N. As Whitney already believes W, it's unlikely that she will come to believe N, if she believes (N or W). Since N and W are each unlikely if she believes (N or W), (N or W) is unlikely, even if she believes it. Thus (N or W) is irrational to believe. So if Addition Closure holds, then N is irrational to believe. For analogous reasons, the principles typically—if not always—clash concerning such propositions.[15]

4. Can Addition Closure be Relinquished?

Since the principles ultimately and frequently clash, we need to give one up. While we could relinquish Addition Closure, we should be quite reluctant to give up such a modest closure principle. After all, as I'll argue, Addition Closure is quite intuitive and it is immune to problems that stronger closure principles face.[16] Let's first consider Addition Closure's intuitiveness. To this end, suppose it's rational to believe that it will rain tomorrow. It plausibly follows that it's rational to believe that it will rain or snow tomorrow. According to Addition Closure, this natural inference is a good one. Moreover, rejecting Addition Closure requires accepting abominable conjunctions like the following: while it is *not* rational to believe that it will rain or snow tomorrow, nonetheless, it *is* rational to believe that it will rain tomorrow.[17] Thus Addition Closure is intuitively plausible.[18]

Moreover, Addition Closure's credal analogue is plausible too. According to it, if a high credence in p is rational to have, then a high credence in (p or q) is rational to have. The credal analogue of Addition Closure shares the intuitive appeal of Addition Closure. To illustrate, if it's rational to have a high credence that it will rain tomorrow, then, surely, it's rational to have a

[15] This lesson concerning the extent of the clash between the principles generalizes to their credal analogues.
[16] For a list of those committed to Addition Closure, see fn. 2.
[17] This argument is inspired by DeRose's [1995] 'abominable conjunction' argument.
[18] While you could try to only accept the weaker principle that (p or q) is rational to believe if p is believed (or rationally believed), how could p be rational to believe if (p or q) is not rational to believe? I'm grateful to an anonymous referee for mentioning this concern.

high credence that it will rain or snow tomorrow. And given the striking similarity between Addition Closure and its credal analogue, it's hard to see how those committed to Addition Closure are not also committed to its credal analogue. Finally, it is an implication of the Bayesian view of credences. According to Bayesianism, if a particular credence in p is rational to have, then at least as high a credence in (p or q) is rational to have too.[19]

To further the case for Addition Closure, let's consider a couple of general objections to closure principles to see whether they cast doubt on Addition Closure's verdict in **Logic Class**.

Let's begin by considering *the Preface Objection* to multi-premise closure principles [Makinson 1965; Christensen 2004]. An instance of these principles is that if p is rational to believe and q is rational to believe, then (p and q) is also rational to believe. According to the objection, we should consider a case where a careful historian rationally believes each historical claim in her new tome. In the tome's preface, though, she humbly expresses her belief that historical errors will be found in the tome, as more historical information is discovered. According to the objection, the conjunction of the tome's historical claims is not rational for the historian to believe since she is rational to believe that historical errors will be discovered in it.

A certain kind of risk-aggregation is the source of the objection. While p can be rational to believe even if there's a chance it's false, the chance can't be *too* high. When risky (independent) propositions are conjoined, the conjunction is riskier than the riskiest conjunct. A conjunction can be too risky to (rationally) believe even if each conjunct isn't too risky to believe. While this type of risk-aggregation is a problem for multi-premise closure principles, it is not a problem for widely accepted single-premise closure principles like Addition Closure.

But there's a different problem for them: *the Long-Sequence Objection* [Lasonen-Aarnio 2008; Schechter 2013]. According to one of these principles, c is rationally believed if it's competently deduced from p (which is rationally believed throughout the deduction). Now

[19] For discussions of Bayesianism centered on ex ante rationality, see Schechter [2013], Gibbs [2019], and Silva and Oliveira [forthcoming].

suppose that there happens to be a very long sequence of competent single-premise deductions, starting with p and ending with c. If the sequence is long enough, it's quite likely that a mistake was made. So contra the given principle, c is not thereby rationally believed.

But notice that the objection does not seem to cast doubt on the claim that (T or P) is rational for Friedrich to believe if T is rational for him to believe. What's significant about cases involving long-sequence deductions is that before c is competently deduced from p, it is *not* obvious that if p, then c. But notice that regardless of whether Friedrich has performed any deductions, it's obvious that if T, then (T or P). Given this salient difference between **Logic Class** and cases involving long-sequence deductions, we can remain confident in Addition Closure's verdict in **Logic Class**.

5. Maybe Rationality Doesn't Seek True Beliefs?

Assuming we hold on to Addition Closure's verdict in **Logic Class**, we must relinquish RSTB.[20] But first, we should revisit its fundamental rationale, based on the idea that rationality is a guide to truth. As Jason Konek and Ben Levinstein [2019: 69–70] say, 'Norms of epistemic rationality . . . have their binding force in virtue of the following fact: they are [a] good means toward the end of securing accuracy.' So what's definitive of rationality is that you can expect to have accurate beliefs if you follow its rules. We can capture this theory of rationality with the following slogan: rationality is (fundamentally) a guide to *true beliefs*.[21] Rationality then guides us toward the propositions that are likely *if believed*. Under this assumption, RSTB is hard to resist.

To resist this argument, it can seem that our only option is to deny that rationality is a guide to truth. Fortunately, we don't need to do this. Another theory of rationality says that rationality is (fundamentally) a guide to *true propositions*; rationality guides us toward the propositions that are likely (independent of whether we believe them). Let's call the view that rationality is a guide to true propositions, and that being (likely to be) a true proposition is a

[20] For another problem for principles like RSTB, see Willard-Kyle [2020].
[21] See Goldman [2001] for a similar view.

fundamental explanatory property of rationality, *truth-aimed epistemology*. And let's call the view that rationality is a guide to true beliefs and that being (likely to be) a true proposition if believed is a fundamental explanatory property of rationality, *true-belief-aimed epistemology*.

The key to resisting this argument for RSTB concerns how we understand the idea that rationality is a guide to truth. If we understand it as a true-belief-aimed epistemologist would want to, the argument looks irresistible. But if we understand it as a truth-aimed epistemologist would want to, the argument is resistible. The reason: being likely is explanatorily fundamental for rationality, rather than being likely if believed. So the fact that a proposition is likely if believed is epistemically insignificant.[22]

But if this is the case, then what do we make of James' aforementioned alpine climber case? For starters, even though the proposition that the climber can successfully make the leap is not likely, it is still *practically* rational for her to believe. Moreover, the given proposition is *epistemically* rational for her to believe *once* she believes it. So the defender of truth-aimed epistemology can certainly grant that there is something to the Jamesian intuition here. Now one might want to go even further and claim that the given proposition is epistemically rational for her to believe *even before* she believes it. But as we saw when assessing the scope of the clash between the principles, James' verdict has a hefty price: Addition Closure must be rejected.

For a second concern, it might be argued that being likely to be false if believed is epistemically insignificant if being likely if believed is epistemically insignificant. And if this is the case, then what should we think about 'self-frustrating' cases where p is likely, but is unlikely if believed? On any view, once the given propositions are believed, they are then not rational to believe. But these cases raise the further question: are these propositions *antecedently* rational to believe? More than one answer is compatible with truth-aimed epistemology.

[22] See Antill [2020] for an analogous response to the aim-of-belief rationale for RSTB.

According to the first, the connection between ex ante rationality and ex post rationality makes these propositions antecedently not rational to believe. So while being likely is one fundamental explanatory property for ex ante rationality, there's also another one: being capable of being rationally believed. **Potential Proof** may falsify the claim that being capable of being rationally believed suffices for being rational to believe. But it doesn't challenge the claim that being capable of being rationally believed is necessary for being rational to believe. And since these propositions are not capable of being rationally believed, they are thereby not rational to believe.

According to the second, these propositions are antecedently rational to believe. After all, as the agent's evidence provides strong reason to believe these propositions, how could they not be rational to believe? True, they are not capable of being rationally believed. But this is only because believing them changes the evidence. And why should p being irrational to believe after the evidence changes make p antecedently irrational to believe? After all, as **Potential Proof** makes it clear, p being rational to believe after the evidence changes does not make p antecedently rational to believe.

As choosing between these ways of developing truth-aimed epistemology raises complicated questions, I will stay neutral here on which view is ultimately correct. But either of them seems attractive enough to provide a promising way to let go of RSTB. Since the objections to truth-aimed epistemology and the rationale for RSTB can be resisted, we can solve the puzzle by letting go of RSTB.[23]

6. Our Solution's Epistemological Consequences[24]

The first upshot concerns the permissivism discussion. Many authors have used RSTB (or a similar principle) to argue for (intrapersonal) permissivism: the theory that there's a

[23] Thanks to an anonymous referee for bringing up these last two concerns about truth-aimed epistemology.
[24] See Marxen [2021a] for a similar discussion.

(possible) body of evidence where an agent with this evidence is rational to believe p and simultaneously rational to believe ~p (or rational to suspend belief).[25] The argument appeals to a case like **Logic Class** where not only is T likely if believed, but ~T is also likely if believed. Assuming RSTB, each of these propositions is thereby *antecedently* rational to believe. So, permissivism is true. But since the key premise (RSTB) in the argument is false, the argument for permissivism is unsuccessful.

A second upshot concerns the discussion over whether epistemic reasons must be evidential. Many have argued in the negative, using a similar principle to RSTB: if T is likely if believed, then there's good epistemic reason to believe T.[26] They've then applied the principle to propositions like T—that are not likely overall, but are likely if believed—drawing the conclusion that there's good epistemic reason to believe T even though it's evidentially unsupported. But notice that T is plausibly rational to believe if there's good epistemic reason to believe it. Since we've seen that T is not rational to believe, this argument fails to show that epistemic reasons can be non-evidential.

A final upshot concerns epistemic consequentialism, which has been the subject of extensive recent discussion. Let's begin with Kristoffer Ahlstrom-Vij and Jeffrey Dunn's [2018: 2] rough characterization of epistemic consequentialism:

> [M]uch of contemporary epistemology has been marked . . . by a tacit endorsement of *epistemic consequentialism*, the idea that epistemic rightness—denoted by terms such as 'justification' or 'rationality'—is to be understood in terms of conduciveness to epistemic goods.

[25] For those who rely on RSTB (or a similar principle), see Velleman [1989], Reisner [2013, 2015], Kopec [2015], Peels [2015], Raleigh [2015, 2017], Drake [2017], and Dahlback [forthcoming]. According to Velleman, although it strongly seems to us that we have freedom, we misunderstand its nature: our freedom is epistemic (i.e. permissivism holds) rather than—as we assume—metaphysical. Joyce [2007]—appealing to Velleman—relies on RSTB to explain how it can be antecedently rational to believe an evidentially-unsupported proposition.

[26] For example, see Foley [1991, 1993], Reisner [2007, 2013], Talbot [2014], Sharadin [2016], and Drake (2017). See Hetherington [1996]—briefly discussed in fn. 10—for a response to Foley.

According to prominent versions of it, accuracy is the only (or one of the) fundamental epistemic good(s). Additionally, what's definitive of epistemic norms is that they can be expected to (best) lead to the acquisition of epistemic goodness (i.e. accuracy) when followed. For example, Hilary Greaves and David Wallace [2006: 610] say:

> [E]pistemic rationality consists in taking steps that can reasonably be expected to bring about epistemically good outcomes.

Now it's quite plausible that these versions imply RSTB (or a similar principle) and thus that propositions like T are rational to believe. After all, if what's definitive of epistemic norms is that they (expectedly) serve the aim of *acquiring accuracy*, how could they prohibit acquiring a true belief? Moreover, when defenders of these theories—as applied to credences (rather than beliefs)—have considered how confident it is rational to be when p is not likely overall, but p is likely if you're highly confident in it, they've argued that it's rational to be highly confident in p [Caie 2013; Greaves 2013; Joyce 2018; Pettigrew 2018]. Their reason (as you might expect): p is likely if you're highly confident in it. But since we've seen that it's not rational to believe (or be highly confident in) T, there's a serious problem for these prominent versions of epistemic consequentialism.

To wrap things up, I should note two points concerning the final upshot. First, there are versions of epistemic consequentialism that avoid this result, but Jennifer Carr and others have raised worries about whether these versions of epistemic consequentialism can be motivated.[27] Second, despite the initial counterintuitiveness of counting self-fulfilling beliefs as rational, most defenders of epistemic consequentialism remain committed to something like RSTB.[28] This paper makes their position more difficult.

---

[27] For more on these versions of epistemic consequentialism, see Carr [2017], and Konek and Levinstein [2019]. And for defenses of this problem, see: Talbot [2014], Carr [2017], Pettigrew [2018], and Marxen [2021b]

[28] See Caie [2013], Greaves [2013], Joyce [2018], and Pettigrew [2018].

7. Conclusion

While RSTB's fundamental motivation has been appreciated for a long time—recently leading many to accept RSTB—**Logic Class** reveals that it clashes with an extremely plausible closure principle. And as **New Toothbrush** helps show, the conflict arises in a host of everyday situations. As we've seen, the best way to solve the puzzle is to let go of RSTB and hold on to Addition Closure. This tells us something important about how to understand the intuitive idea that rationality is a guide to truth: rationality is a guide to *the belief-independent truth*. And this in turn has important implications for current discussions of permissivism, epistemic reasons, and epistemic consequentialism.

Acknowledgements Many thanks to Zach Barnett, Jamie Dreier, Louis Gularte, Adam Pautz, Richard Pettigrew, Gerard Rothfus, Wes Siscoe, and an anonymous referee for *Australasian Journal of Philosophy*. I'm especially grateful to David Christensen and Josh Schechter for their many valuable suggestions.

Institutional Affiliation Brown University

References

Ahlstrom-Vij, K. and J. Dunn 2018. Introduction: Epistemic Consequentialism, in *Epistemic Consequentialism*, eds J. Dunn and K. Ahlstrom-Vij, Oxford: Oxford University Press: 1–22.

Antill, G. 2020. Epistemic Freedom Revisited, *Synthese* 197: 793–815.

BonJour, L. 1985. *The Structure of Empirical Knowledge*, Cambridge: Harvard University Press.

Caie, M. 2013. Rational Probabilistic Incoherence, *Philosophical Review* 122/4: 527–75.

Carr, J. 2017. Epistemic Utility Theory and the Aim of Belief, *Philosophy and Phenomenological Research*, 95/3: 511–34.

Christensen, D. 2004. *Putting Logic in its Place: Formal Constraints on Rational Belief*, New York: Oxford University Press.

Cohen, S. 1984. Justification and Truth, *Philosophical Studies* 46/3: 279–95.

Dahlback, M. Forthcoming. Infinitely Permissive, *Erkenntnis*: 1–9.

DeRose, K. 1995. Solving the Skeptical Problem, *The Philosophical Review* 104/1: 1–52.

Dodd, D. 2012. Evidentialism and Skeptical Arguments, *Synthese* 189/2: 337–52.

Doore, G.L. 1983. William James and the Ethics of Belief, *Philosophy* 58/225: 353–64.

Drake, J. 2017. Doxastic Permissiveness and the Promise of Truth, *Synthese* 194: 4897–912.

Dretske, F. 1971. Conclusive Reasons, *Australasian Journal of Philosophy* 49: 1–22.

Feldman, R. 1995. In Defence of Closure, *The Philosophical Quarterly* 45/181: 487–94.

Foley, R. 1991. Evidence and Reasons for Belief, *Analysis* 51/2: 98–102.

Foley, R. 1993. *Working without a Net: A Study of Egocentric Epistemology*, New York: Oxford University Press.

Gibbs, C. 2019. Basing for the Bayesian, *Synthese* 196: 3815–40.

Goldman, A. 1979. What is Justified Belief?, in *Justification and Knowledge*, ed. G.S. Pappas, Dordrecht: D. Reidel Pub. Co.: 1–23.

Goldman, A. 1986. *Epistemology and Cognition*, Cambridge: Harvard University Press.

Goldman, A. 2001. The Unity of the Epistemic Virtues, in *Virtue Epistemology: Essays in Epistemic Virtue and Responsibility*, eds A. Fairweather and L. Zagzebski, New York: Oxford University Press: 30–48.

Greaves, H. 2013. Epistemic Decision Theory, *Mind* 122/488: 915–52.

Greaves, H. and D. Wallace 2006. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility, *Mind* 115/459: 607–32.

Hetherington, S. 1996. Foley's Evidence and His Epistemic Reasons, *Analysis* 56/2: 122–6.

Hocking, W.E. 1912. *The Meaning of God in Human Experience: A Philosophic Study of Religion*, New Haven: Yale University Press.

Hookway, C. 2002. *Truth, Rationality, and Pragmatism: Themes from Peirce*, New York: Oxford University Press.

Horowitz, S. 2019. The Truth Problem for Permissivism, *The Journal of Philosophy* 116/5: 237–62.

Ichikawa, J.J. and B.W. Jarvis 2013. *The Rules of Thought*, Oxford: Oxford University Press.

James, W. 1912 (2010). *The Will to Believe: And Other Essays in Popular Philosophy*, Auckland: The Floating Press.

Joyce, J.M. 2007. Are Newcomb Problems Really Decisions?, *Synthese* 156/3: 537–62.

Joyce, J.M. 2018. Accuracy, Ratification, and the Scope of Epistemic Consequentialism, in *Epistemic Consequentialism*, eds J. Dunn and K. Ahlstrom-Vij, Oxford: Oxford University Press: 240–68.

Konek, J. and B. Levinstein 2019. The Foundations of Epistemic Decision Theory, *Mind* 128/509: 69–107.

Kopec, M. 2015. A Counterexample to the Uniqueness Thesis, *Philosophia* 43: 403–9.

Lasonen-Aarnio, M. 2008. Single Premise Deduction and Risk, *Philosophical Studies* 141: 157–73.

Makinson, D.C. 1965. The Paradox of the Preface, *Analysis* 25: 205–7.

Marxen, C. 2021a. On the Epistemic Rationality and Significance of Self-Fulfilling Beliefs, *Synthese*: 1–18.

Marxen, C. 2021b. Epistemic Utility Theory's Difficult Future, *Synthese*: 1–21.

Moore, G.E. 1907. Professor James' 'Pragmatism', *Proceedings of the Aristotelian Society* 8: 33–77.

Okasha, S. 1999. Epistemic Justification and Deductive Closure, *Crítica: Revista Hispanoamericana de Filosofía* 31/92: 37–51.

Peels, R. 2015. Believing at Will is Possible, *Australasian Journal of Philosophy* 93/3: 524–41.

Perry, R.B. 1938. *In the Spirit of William James*, New Haven: Yale University Press.

Pettigrew, R. 2018. Making Things Right: The True Consequences of Decision Theory in Epistemology, in *Epistemic Consequentialism*, eds J. Dunn and K. Ahlstrom-Vij, Oxford: Oxford University Press: 220–39.

Pryor, J. 2004. What's Wrong with Moore's Argument?, *Philosophical Issues* 14/1: 349–78.

Raleigh, T. 2015. An Argument for Permissivism from Safespots, in *Logic, Rationality, and Interaction*, eds W. van der Hoek, W.H. Holliday, and W-f Wang, Heidelberg: Springer: 308–15.

Raleigh, T. 2017. Another Argument Against Uniqueness, *The Philosophical Quarterly* 67/267: 327–46.

Reisner, A. 2007. Evidentialism and the Numbers Game, *Theoria* 73/4: 304–16.

Reisner, A. 2013. Leaps of Knowledge, in *The Aim of Belief*, ed. Timothy Chan, New York: Oxford University Press: 167–83.

Reisner, A. 2015. A Short Refutation of Strict Normative Evidentialism, *Inquiry* 58/5: 477–85.

Rosenkranz, S. 2007. Agnosticism as a Third Stance, *Mind* 116/461: 55–104.

Santayana, G., W.M. McClay, and J. Seaton 2009. *The Genteel Tradition in American Philosophy*: *And Character and Opinion in the United States*, New Haven: Yale University Press.

Schechter, J. 2013. Rational Self-Doubt and the Failure of Closure. *Philosophical Studies* 163: 429–52.

Schoenfield, M. 2015. Bridging Rationality and Accuracy, *The Journal of Philosophy* 112/2: 633–57.

Sharadin, N.P. 2016. Nothing but the Evidential Considerations?, *Australasian Journal of Philosophy* 94/2: 343–61.

Silva, P. and L.R.G. Oliveira Forthcoming. Propositional Justification and Doxastic Justification, in *Routledge Handbook of the Philosophy Evidence*, eds M. Lasonen-Aarnio and C.M. Littlejohn, Routledge: 1–17.

Talbot, B. 2014. Truth Promoting Non-Evidential Reasons for Belief, *Philosophical Studies* 168: 599–618.

Taylor, C. 2002. *Varieties of Religion Today: William James Revisited*, Cambridge: Harvard University Press.

Turri, J. 2010. On the Relationship between Propositional and Doxastic Justification, *Philosophy and Phenomenological Research* 80: 312–26.

Velleman, D. 1989. *Practical Reflection*, Princeton: Princeton University Press.

Velleman, D. 2000. *The Possibility of Practical Reason*, New York: Oxford University Press.

Vogel, J. 2008. Epistemic Bootstrapping, *The Journal of Philosophy* 105/9: 518–39.

Wedgwood, R. 2002. The Aim of Belief, *Noûs* 36: 267–97.

Whiting, D. 2018. Knowledge, Justification, and (a Sort of) Safe Belief, *Synthese* 197/8: 3593–609.

Willard-Kyle, C. 2020. Against the Doctrine of Infallibility, *The Philosophical Quarterly*: 1–22.

Wright, C. 2004. Warrant for Nothing (And Foundations for Free)?, *Aristotelian Society Supplementary* 78/1: 167–212.