AGAINST THE DOCTRINE OF INFALLIBILITY

Christopher Willard-Kyle
Forthcoming in *Philosophical Quarterly*

Abstract: According to the doctrine of infallibility, one is permitted to believe *p* if one knows that necessarily, one would be right if one believed that *p*. This plausible principle—made famous in Descartes' *cogito*—is false. There are some self-fulfilling, higher-order propositions one can't be wrong about but shouldn't believe anyway: believing them would immediately make one's overall doxastic state worse.

Keywords: Infallibility; Veritism; Defeat; Self-Fulfilling Beliefs; Higher-Order Beliefs; *Cogito*

1. The Doctrine of Infallibility

A proposition is epistemically infallible for an agent just in case that it's impossible for that agent to falsely believe it:

> **Infallibility**: A proposition *p* is infallible for S iff it's impossible that S falsely believes that *p*.

This definition closely resembles the following:

> **Infallibility***: A proposition *p* is infallible for S iff it's necessary that if S believes that *p* then S truly believes that *p*.

If, necessarily, all beliefs are either just true or just false—as I believe they are—then infallibility and infallibility* are equivalent. I shall suppose that they are equivalent throughout the rest of the paper, although there is interesting territory to explore for those who believe that some propositions are neither or both true and false.

Since one can't be mistaken about one's infallible beliefs, it's tempting to think we should always believe them. When there's no risk of false belief, why not? At least, if we're in a position to recognize that a proposition is infallible for us, surely then we may believe it. As Alston says, 'one could hardly have a stronger (epistemic) justification for holding a certain

belief than the logical impossibility of the belief's being mistaken' (Alston 1971: 229). Consider, then, the doctrine of infallibility:

> **Doctrine of Infallibility**: If S knows <necessarily, if she herself were now to believe that *p* then she would truly believe that *p*> then it is thereby (rationally) permissible for S now to believe that *p*.

In slogan form: You're always permitted to believe (known) guaranteed truths.

I argue, however, that not all propositions known to be infallible may be believed. In fact, some propositions that are known to be infallible should be disbelieved!

The doctrine of infallibility emerges as a battleground between two otherwise attractive philosophical theses: one the one hand, *veritism*, the thesis that accuracy is the fundamental epistemic good, and on the other hand what I shall call *reflectivism*, the thesis that one's reflective attitudes about one's first-order beliefs make at least some difference to the epistemic quality of those first-order beliefs or one's belief system. Although the thesis of this paper is that the doctrine of infallibility is false, a recurring theme is that reflectivism is in tension with veritism.

2. Clarifying 'Infallibility'

Some clarifications are in order. First, infallibility is sometimes ascribed to agents in relation to a proposition or subject matter (e.g. 'The Pope is infallible about matters of faith when he speaks *ex cathedra*'). Thus, Alston writes: 'one can be said to be infallible *vis-à-vis* a certain subject matter provided one cannot be mistaken in any beliefs he forms concerning that subject matter' (Alston 1971: 229). Talking about infallible *agents* might suggest to the reader some kind of super-knower or an agent with extra-special epistemic access to some domain (like a Pope, prophet, or supercomputer).

Terminologically, I prefer to speak in the opposite way: Infallibility is a property that *propositions* have in relation to *agents*. This serves to emphasize that infallibility (as used in this paper) requires no special competence on the part of the agent for whom a proposition is infallible. All necessary mathematical truths are infallible for infants, for instance, though of course infants do not *know* that those truths are infallible for them.

Focusing on propositions rather than 'subject matters' also allows us to be more fine-grained in targeting particular beliefs. That I exist is infallible for me (in my sense), but it's at least not obvious that I couldn't be mistaken about the subject. Suppose I believe that I am a fiction, for instance, and that strictly speaking I don't exist.[1] Or that there are no selves at all. I'd then be mistaken about the question (or subject) of whether I exist even though, necessarily, if I believed <I exist> I'd be right. One can be mistaken about a subject matter even if one can't be mistaken in believing a particular proposition that is an answer to a question belonging to that subject matter.

My definition of 'infallibility' also differs from certain other uses in the literature. For instance, Jeremy Fantl and Matthew McGrath suggest that an agent knows (and perhaps also believes) that *p* infallibly when there is no epistemic chance for them that not-*p* (2009: 11). This definition is in some ways more and in some ways less restrictive than the one used here. It is less restrictive because it counts as infallible those propositions for which an agent has maximal justification: there's no *epistemic* chance that not-*p* for such agents. But it is often still possible (though not *epistemically* possible, perhaps) that agents could be wrong about a proposition for which they have perfect justification: they could have had different evidence, for instance. Fantl and McGrath's definition is, in other respects, *more* restrictive than mine because they leave

[1] See Lebens (2015) for an intriguing articulation of a view in this neighborhood.

open that a belief in a necessary proposition may be fallible (e.g. when a math student believes an axiom on the basis of testimony rather than by working out the proof [cf. 2009: 8]) even though, necessarily, if they believe the axiom, they will be right.

Nor does my use of 'infallibility' perfectly overlap with the notion employed in what Jessica Brown calls 'probability 1 infallibilism' (Brown 2013: 626), the sort of infallibility one has toward a proposition when it has probability 1 conditional on one's evidence.[2] A proposition could have less than probability 1 and yet still be infallible for me. Necessary truths that I can't conceptualize may not have probability 1 for me (because I may not have *any* evidence bearing on them) and yet be such that, necessarily, if I believe them, I will be right.

Second, what I mean by 'infallibility' is related to but distinct from what some authors mean by 'incorrigibility.' It's useful to briefly disentangle them. Frank Jackson discusses the view that 'it is logically impossible to be mistaken about certain of one's current mental states' under the guise of an 'incorrigibility thesis' (Jackson 1973: 51). This is an obviously related notion, but one that comes apart from my definition of infallibility. Nothing in my proposal centers on an agent's own mental states, and, as for Alston, there is be a difference between one's ability to be mistaken *about* a proposition (by believing *either* that it is true *or* that it is false) and one's inability to be mistaken in believing of a particular proposition *that it is true*.

Sydney Shoemaker's definition of incorrigibility is closer in that it is conditional upon a kind of affirmation of the truth of a proposition: 'If a person sincerely asserts such a statement it does not make sense to suppose, and nothing could be accepted as showing, that he is mistaken' (Shoemaker 1963: 215). Shoemakerian incorrigibility is a kind of public infallibility. If a person believes an infallible proposition, their (sincerely believed) assertions of the same will be

[2] Brown (2013) has Williamson's E=K thesis particularly in mind.

incorrigible in Shoemaker's sense—after all, no proposition that is true can be shown to be mistaken. But not all (Shoemakerian) incorrigible propositions are infallible, for a proposition might be incapable of being disproved but false, nonetheless.

There may, however, be a sense in which my infallibility is a kind of incorrigibility. Incorrigibility is (according to one way of thinking) the inability to be corrected. An agent might be uncorrectable with respect to a proposition because (a) they are unable to change their mind (or, at any rate, to have their mind persuasively changed from the outside) or (b) any change of mind would not constitute a correction, i.e. because the belief in question was already true. Any belief in an infallible proposition is incorrigible in the second sense.

3. On Behalf of the Doctrine of Infallibility

But before defending the surprising conclusion that the doctrine of infallibility is false, let's give the doctrine its due. The doctrine is deeply intuitive and for good reason. First, the doctrine has historical pedigree. It is to infallibility that Descartes appeals at the climax of the *cogito*:

> [Suppose] there is a deceiver of supreme power and cunning who is deliberately and constantly deceiving me. …[L]et him deceive me as much as he can, he will never bring it about that I am nothing so long as I think that I am something. …I must finally conclude that this proposition, I am, I exist, *is necessarily true whenever it is put forward by me*[3] or conceived in my mind. (Descartes 1984: 7)

It's the infallibility of <I exist> that makes the wheels of the *cogito* turn and—on at least one way of reading Descartes[4]—makes it the first item to fully survive the scrutiny of Descartes' method of doubt.

---

[3] I take *believing* a proposition to be at least one way of *putting it forward*.

[4] There's an important exegetical question about how Descartes' notion of infallibility employed in the *cogito* relates to the clarity and distinctness criterion that takes center stage in Meditation III. I am grateful to Christopher Frugé and Ram Neta for excellent discussion on alternative interpretations of Descartes, which unfortunately cannot be captured in a footnote.

The *cogito* illustrates the benefit of having a principle whereby guaranteed truth is *sufficient* all by itself to license belief. If the doctrine of infallibility were false, that would mean that Descartes had not yet done enough (at this point in the *Meditations*, at least) to show that we are permitted to believe that we exist! Descartes is often accused of making the standard for defeating scepticism too high: rarely has he been accused of making the standard too low.

Contemporary epistemologists continue to appeal to Cartesian infallibility at crucial junctures. Ernest Sosa, for instance, appeals to infallibility to ward off dream scepticism. Having argued that in dreams 'we do not really believe; we only make-believe' (Sosa 2007: 8), Sosa claims that if we really believe that we are dreaming we must not be dreaming: If we were dreaming, our 'belief' wouldn't really be a belief at all, but only a make-belief. We should affirm (and neither suspend nor deny) that <I am awake> 'since only about that option [i.e. affirming] is it obvious to me now that *if I take it I will be right*' (Sosa 2007: 19).[5]

A second reason for the doctrine of infallibility emerges from a particular picture of epistemic value. 'Believe truth! Shun error!' says Williams James (1907: 18). Epistemically, 'these are our first and great commandments' (17),[6] and—a more ambitiously reductive epistemologist might have added—the only ones.

James's aphorisms are suggestive of one of the main characters in our dialectic: epistemic veritism (cf. Goldman 1999: 5). According to veritism (as used here), the only fundamental epistemic value is accuracy.[7] Whatever else can be said for and against this view, it is attractively

[5] Cf. Wittgenstein: 'The argument "I may be dreaming" is senseless for this reason: if I am dreaming, this remark is being dreamed as well—and indeed it is also being dreamed that these words have any meaning' (1969: 383). An important difference is that, for Wittgenstein, the belief <I am dreaming> can only be true or senseless whereas for Sosa it can only be true or not a belief at all.

[6] Intriguingly, James himself fluctuates between *believing* the truth (1907: 18) and *knowing* the truth (17) as the positive epistemic commandment. In this paper, I shall represent James as endorsing the commandment to believe the truth, although this is a simplifying, historical fiction.

[7] Cf. Goldman: '[T]rue belief is the ultimate value in the epistemic sphere' (2001: 32).

simple. James's aphorisms are telic: believing truth and shunning falsehood are goals that epistemic agents ought to promote.[8] James notes that an agent might emphasize one of these goals more than the other. A cautious believer might be hesitant to believe when there is even a small chance of error (cf. Kelly 2014). Infallible beliefs, however, are such that there is some chance they will lead to true belief and no chance that they will lead to false belief. If the only—or, at any rate, the most fundamental—epistemic goods are believing truly and avoiding false belief, then there's seemingly always a good decision-theoretic reason to take a chance on infallible beliefs: There's an opportunity (indeed a certainty!) to gain an epistemic good without any risk of epistemic harm.

One could accept the doctrine of infallibility without being a veritist of this stripe, but veritists should be especially attracted to the doctrine of infallibility. Adding a true belief *always* improves accuracy, at least if one can do so without adding any false beliefs or removing true ones.[9] And there's no obvious reason why believing an infallible proposition would require one to also acquire false or abandon true beliefs. Indeed, the very plausibility of the doctrine of infallibility is likely to be seen as an argument for veritism. If guaranteed truth is by itself sufficient to license belief, a good explanation is that accuracy is what matters most in epistemology.

Third, the doctrine of infallibility, or at least a principle that entails it, explains the right verdict in certain tricky cases that reverse the normal causal direction of fit between mind and world. Note that the doctrine of infallibility is a weak version of a family of principles that

[8] For a non-telic, or at least anti-consequentialist, version of veritism that falls outside the sort targeted in this paper, see Sylvan (2018).

[9] Berker (2013a), building on Firth (1981), argues against veritism on the grounds that it permits wrongheaded tradeoffs, allowing agents to believe obvious falsehoods to gain true beliefs downstream: veritists ignore 'the epistemic separateness of propositions' (2013a: 365). Since the cases in this paper that might cause trouble for veritism don't depend on tradeoffs of the relevant sort, I will stay neutral on whether veritists are committed to permitting them.

permit agents to believe when they know that their belief would have some truth-oriented property or other if believed. In particular, the

> **Doctrine of Infallibility**: If S knows that <necessarily, if she herself were now to believe that *p* then she would truly believe that *p*> then it is thereby (rationally) permissible for S now to believe that *p*.

is entailed by the

> **Doctrine of Truth:** If S knows that <if she herself were now to believe that *p* then she would truly believe that *p*> then it is thereby (rationally) permissible for S now to believe that *p*.

For if S is permitted to believe *p* in virtue of knowing that she *wouldn't* be wrong about *p*, then she is surely permitted to believe *p* in virtue of knowing that she *couldn't* be wrong about *p*.

The doctrine of truth echoes Velleman's (1989a/2000) articulation of epistemic freedom, in which he argues that one is 'entitled to say,' and, Velleman suggests elsewhere, *believe*, what 'wouldn't be false if [one] said it' (Velleman 2000: 40).[10] It turns out that the doctrine of truth is extremely useful in explaining why we are permitted to believe certain propositions when the ordinary direction of fit between mind and world is reversed. Indeed, several authors including Velleman (2000: 40, 44), Reisner (2013: §2), Kopec (2015: 404), Raleigh (2017: 332–333), Drake (2017: 4901) and Dahlback (forthcoming) appeal, at least implicitly, to something like the doctrine of truth in order to make sense of such cases.

Here's a representative case from Dahlback (forthcoming). A friendly demon guarantees that the result of a coin flip will match your belief about whether it is heads or tails. Dahlback reasons that so long as we know that we are in such a situation, we are permitted either to believe that the coin will land heads or that it will land tails (and so not heads) since we'd know that our

[10] Elsewhere (2000: 40), Velleman suggests that the agent must have *evidence* that they wouldn't be wrong about *p*, bringing Velleman's principle even closer to my formulation of the doctrine of truth.

belief was correct. In this case, the doctrine of truth seems to yield the proper result. It does seem permissible to believe either that the coin will land heads or that it will land tails.

Importantly, this seems right even though the evidence favors neither the thesis that the coin will land heads nor that it will land tails: the favoring evidence runs out. By 'favoring evidence,' I mean evidence that favors one attitude or contiguous range of attitudes over its competitors. Normally when the evidence favors neither *p* nor not-*p* we are required to suspend judgment. But in this case, to suspend judgment would to 'willingly cast aside the promise of truth' (Drake 2017: 4902). One can't *follow* the evidence to the conclusion that the coin will land heads or tails—one simply believes and thereby makes oneself right. That the antecedent evidence favors neither thesis is important: It makes clear the role that the doctrine of truth plays (or seems to) in explaining the permissibility of believing either that the coin will land heads or that it will land tails.

As Dahlback notes, an important feature of the case is that the ordinary causal direction of fit between mind and world is reversed: the belief is a kind of self-fulfilling prophecy. We are guaranteed to be right not because our mind is tracking the world but because the world is tracking our mind. That the doctrine of truth can deliver the right verdict in such cases—when the favoring evidence seems to run out—is a strong point in its favor. Considered broadly, such cases fit neatly with the reductive gloss on James sketched earlier. True and false beliefs are what matter. So long as you knew you'd be right, who cares (our imagined Jamesian shrugs) whether how you got there was by following the evidence? After all, one's self-fulfilling beliefs and one's evidence remain 'subjunctively linked' insofar as the belief 'creates adequate evidence for it[self]' (Foley 1991: 102). And if the goal of following the evidence is to find the truth, then we may ask with Velleman (1989b: 63): 'Why would rules [to follow the evidence] designed to

help one arrive at the truth forbid one to form a belief that would be true?' What's important isn't how you get there but that you knew you'd be right at journey's end.

I've suggested that the ability of the doctrine of infallibility to explain why we are permitted to believe that the coin will land heads is a point in its favor. But some disagree: some think that we are *not* epistemically entitled to believe that the coin will land heads (although we may, for instance, be *pragmatically* entitled to *form* the belief that the coin will land heads). There's an important debate to be had here about the right and wrong kinds of reasons for belief (see Kavka 1983; Resiner 2009; Schroeder 2012) and whether we aim sometimes to have true beliefs or only to believe what is true (Antill 2020). Intuitions on this subject can be hard to leverage. For instance, Kopec (2015), Raleigh (2017), Drake (2017) and Dahlback (forthcoming) use cases of self-fulfilling belief to argue for a robust permissivism whereas Antill (2020) argues against interpretations like theirs (in part) precisely *because* they lead to robust forms of permissivism.[11]

Philosophically, there's much left to resolve. Dialectically, however, we may sidestep this issue. Those who insist that we don't properly respect our evidence (because we don't *follow* it) in cases of self-fulfilling beliefs will *already* be sceptical of the doctrines of truth and infallibility. For the doctrine of infallibility does not require that agents have beliefs that are arrived at by following the evidence: it allows beliefs that they merely know they are guaranteed not to be wrong about. And, in cases of self-fulfilling belief, these criteria can come apart.

I think, however, that the defender of the doctrine of infallibility is right to take our intuitive permission to believe either way in certain cases of known-to-be self-fulfilling beliefs as evidence for their view. One further point in favor of this interpretation, articulated in Raleigh

[11] For my preferred defense of permissivism, see Willard-Kyle (2017).

(2017: 338–339), is that the corresponding principle according to which we are *not* permitted to have beliefs that are guaranteed to be *false* is also highly plausible. For instance, no one should believe <*p*, but I don't believe that *p*> (cf. Raleigh 2017: 329; Reisner 2014: 482). One doesn't (necessarily) decide not to believe this proposition by weighing the evidence for it: one can simply decide to reject it on the grounds that one can only believe it falsely.

In any case, I will ultimately argue that the doctrine of infallibility is false because there are *particular* (known-to-be) self-fulfilling judgments that we are not epistemically entitled to make. But my argument won't depend on any qualms about the propriety of self-fulfilling beliefs in general.[12]

4. The Problem of Easy Downgrade

So, there's much to be said for the doctrine of infallibility. It anchors prominent anti-sceptical arguments, encourages us to take smart epistemic bets, and helps explain some otherwise tricky cases when the evidence follows our beliefs rather than our beliefs the evidence.

Nevertheless, the doctrine of infallibility is false. Although we usually make our overall epistemic state better by believing in such a way that we couldn't be wrong, we can also make our overall epistemic state worse.

The doctrine of infallibility is false because it makes it too easy to permissibly acquire defeaters for our beliefs (or at least too easy to rationally downgrade them). A belief is defeated by another belief, in the stipulative sense of this paper, when the second belief makes the first lose some positive epistemic status. Suppose I believe it is noon but then learn that the clock I

[12] See Berker (2013b: 376–377) for an argument against veritism that *does* rest on qualms about the propriety of self-fulfilling beliefs in general. Notably, however, Berker explicitly restricts his case to beliefs that are *not* known to be self-fulfilling.

had based my belief on runs an hour late. My belief that the clock is running an hour late defeats the justification for my belief that it is noon. Losing justification is one way of losing a positive epistemic status; so, the belief has been (in our sense) defeated. (Note that our sense of 'defeat' is intentionally broader than those that require loss of some *particular* epistemic quality like justification or knowledge.)

It's controversial just when defeat happens. To get a grip on the problem for the doctrine of infallibility, let's begin by considering a very permissive defeat principle. Although many (the author included) will find this first-pass defeat principle ultimately unconvincing, it will help us to identify a recipe for finding infallible propositions that ought not be believed. This recipe will give us a strategy for cooking up counterexamples to the doctrine of infallibility that can be adjusted for philosophical taste. Here is the first principle:

> **Easy Irrationality**: Necessarily, it is irrational for S to believe that *p* if S believes that it is irrational for S to believe that *p*.

The principle **Easy Irrationality** has some plausibility. Suppose S believes that it is irrational for her to believe some proposition, but she believes it anyway. The agent apparently displays a lack of appropriate epistemic reflection. She believes she has no good reason for believing *p* and believes it anyway. Something seems to have gone wrong epistemically.

Easy Irrationality is thus one avatar (though not the only one) of our second character: *reflectivism*. It's one way of expressing the intuition that one's reflective attitudes about one's first-order beliefs make at least some difference to the epistemic quality of those first-order beliefs.

Let's not worry too much about whether this principle survives scrutiny.[13] What I want to argue here is that *if* **Easy Irrationality** is true then there are certain infallible propositions that should not be believed.

Suppose an agent knows both *p* and **Easy Irrationality**. They consider whether they may believe, in addition, that it is irrational for them to believe that *p*. They ask themselves, 'Suppose I were to believe that it is irrational for me to believe that *p*. Would that belief be true?'

Absolutely! For according to **Easy Irrationality**, believing that it's irrational to believe that *p* is enough to make believing that *p* irrational. Simply having the belief makes it so. More than that, since **Easy Irrationality** is a necessary truth, it's impossible that the belief could be false. The agent knows that the belief <it is irrational for me to believe that *p*> is infallible for them. The agent knows that, necessarily, if she herself were to believe <it is irrational for me to believe that *p*> then she would truly believe <it is irrational for me to believe that *p*>.

But obviously, it's wrong to believe that a belief is irrational just because **Easy Irrationality** makes that higher-order belief infallible. This would lead an agent to have a worse set of beliefs overall if *p* was otherwise rational to believe (and if there wasn't independent reason to doubt that it was rational to believe that *p*). The first-order belief could become needlessly irrational.

This is most clear in the case in which, antecedent to forming the easily-irrationalizing belief, the agent had been in a position to know and rationally believe both <*p*> and <it is *rational* for me to believe that *p*>. The agent then has (at least) two choices:

A. Believe <it is rational for me to believe that *p*> and believe <*p*>.

B. Believe <It is irrational for me to believe *p*> and believe <*p*>.

[13] See Coates (2012) for a critique.

In this situation, the agent should clearly choose A over B. For only by choosing A will the agent emerge with two beliefs that are both true and rational. For if **Easy Irrationality** is true, then choosing option B will result in an irrational belief that *p*.

But even if the agent hadn't been in a position to know that their first-order belief was rational, it seems wrong to believe <it is irrational for me to believe that *p*> *merely* because one would be right. Doing so takes too cavalier an attitude toward the possibility of downgrading the rationality of one's first-order beliefs. One is making it irrational to believe something without any evidence that one's epistemic situation forces this undesirable outcome. Something seems to have gone wrong.

Maybe what's gone wrong is **Easy Irrationality**. After all, many epistemologists think defeat is hard to come by. Lasonen-Aarnio (2010), Coates (2012), Williamson (2014), and Weatherson (manuscript), for instance, all argue against **Easy Irrationality** or analogues of it. Nevertheless, thinking about **Easy Irrationality** was valuable, for it has given us a template for thinking about how certain higher-order beliefs could in principle be infallible. Consider the following infallibility recipe:

> **Infallibility Recipe**: Necessarily, if S believes that it is F for S to believe that *p* then it is F for S to believe that *p*.

Easy Irrationality is true just in case we can plug in 'irrational' as the ingredient for F. As noted, if defeat is hard then simply believing that a belief is irrational might not be enough to make the first-order belief irrational. But it still seems that negative higher-order epistemic appraisals make their corresponding first-order beliefs worse in some way, even if it doesn't always make them irrational. Our task is to find some F that captures whatever way it is that first-order beliefs become worse upon receiving negative higher-order appraisals.

There are four broad ways that one might argue that no negative, epistemic property satisfies the infallibility template. First, one might endorse *extreme level-splitting*,[14] the view that the quality of our belief systems is not at all impacted by the relationship between our lower- and higher-order beliefs.

Extreme level-splitting seems unduly strong. If we think it's a total disaster to believe that *p* but believe *p* anyway, surely that lowers the quality of our first-order belief or our belief system in *some* way. If extreme level-splitting is true, then we can completely ignore our higher-order beliefs when evaluating their first-order counterparts. That stretches credulity. Surely there are better ways to argue that nothing satisfies the infallibility template. In other words, endorsing extreme level-splitting violates the intuitive thesis we've called epistemic *reflectivism*: our reflective beliefs about first-order beliefs have at least *some* impact on the quality of our corresponding first-order beliefs or belief system.

I take the falsity of extreme level-splitting as a datum. But even those who embrace a degree of level-splitting don't explicitly endorse the extreme thesis that second-order beliefs have *no* effect on the quality of first-order beliefs of belief systems. Coates (2012) argues that agents can rationally believe that their belief that *p* is *irrational* while, nevertheless, *rationally* believing that *p*. This view doesn't entail extreme level-splitting though, since it's consistent with all this that believing that it's irrational to believe *p* makes one's belief that *p* worse in some way even if it doesn't make it flat-out irrational. Similar observations apply to Weatherson's claim that 'what we should believe can come apart from what we should believe that we should believe' (manuscript) and Williamson's (2014) view that one can know that *p* while knowing it is improbable that one knows that *p*.

[14] I borrow this term from Horowitz (2014), although my use is more restrictive.

These views are suggestive, however, of a second strategy for arguing that the infallibility template is never truly instantiated: One could argue that although negative higher order beliefs affect first-order beliefs in some way, no negative property is such that believing a belief has that property automatically makes the corresponding first-order belief bad *in that very way*. After all, we can be mistaken—even rationally mistaken—in our first-order beliefs. Why should our second-order beliefs be any different? On this picture, believing that one's beliefs are bad in an F-ish way does indeed make them worse (extreme level-splitting is false), but one's first-order beliefs may be made worse in a G-ish way rather than an F-ish way.

This strategy is consonant with the view of defeat articulated by Lasonen-Aarnio (2010). Lasonen-Aarnio argues persuasively that we shouldn't confuse our evidence being such that it's *unlikely* that we know that *p* with our actually not knowing that *p* (2010: 10). An agent might have evidence that their visual capacities are misfiring, but if their visual capacities are operating well, and if the agent bases their belief solely on their visual capacities, it might be that 'being stubborn pays off' (2). If this is right, agents who stubbornly believe what they are unlikely (on their evidence) to know may nonetheless emerge with full knowledge.

But Lasonen-Aarnio is equally emphatic that we can, nevertheless, genuinely criticize agents who believe against the evidence—they are being *unreasonable*. After all, they are risking a lot (epistemically-speaking) by forming a belief that, according to their evidence, is very unlikely to constitute knowledge. It's just that they might not be criticizable in the way that we first thought. They still get to count as knowers, but *unreasonable* knowers.

Indeed, if Lasonen-Aarnio could not explain why agents who believed (and thereby came to know) in the face of significant (but ignored) counter-evidence were criticizable in *some way or other*, that would speak against her account. It adds significantly to the plausibility of the

overall picture that she does not give the unreasonable knower uniformly positive marks. We need to be cautious about saying *how* higher-order criticisms negatively impact first-order attitudes—but that doesn't give us reason to doubt *that* they do so.

It's important that the most plausible views according to which defeat is difficult nevertheless preserve ways to criticize agents whose first-order beliefs are in tension with their higher-order ones. For it will allow us to fill in the infallibility recipe by going general. Our second strategy for avoiding the problem of easy defeat said that negative higher-order epistemic appraisals make their first-order counterparts worse in some way, just not automatically in the way we believed them to be worse. To counter this strategy, we can move to a principle that uses a sufficiently general negative, higher-order appraisal. Consider:

> **Easy Problems**: Necessarily, if S believes that it is problematic for S to believe that *p* then it is problematic for S to believe that *p*.[15]

'Problematic' is such a general, negative term that *any* way of making a first-order belief worse counts as problematic. So, it seems that **Easy Problems** is true even if more specific principles like **Easy Irrationality** are not.

Unless our third strategy for resisting the infallibility recipe succeeds. Our second strategy was to argue that although believing that one's beliefs are bad in an F-ish way makes them worse, it always makes them worse in a G-ish way and not an F-ish way. Our third strategy admits that there is some sufficiently general F such that believing a belief to be F makes it worse in an F-ish way. But it denies that the first-order belief must be bad enough to make it F outright. So, for instance, the defender of the third strategy insists that although believing that it is problematic to believe *p* entails that it is *more* problematic to believe that *p* than it might have been otherwise, it doesn't make believing *p* problematic outright. It treats 'problematic'—or any

[15] I'm grateful to Laura Callahan and Ernie Sosa for conversation that led to this version of the principle.

F that might otherwise satisfy the infallibility template—as a threshold term, such that being more F does not entail being F (just as being *taller* than something doesn't entail being *tall*).

It's unclear whether this strategy is properly motivated—it's far from obvious that 'problematic' is relevantly like 'tall.' But instead of pressing this line, we'll look for a term that avoids the objection altogether: an F such that being more F (than whatever) entails its being F outright.

We can do this by stipulating new terminology that does not operate with this sort of threshold. Let's introduce the term 'besmirched': An agent's belief is *besmirched* just in case a belief is (epistemically) problematic *to any degree*. 'Besmirched' maintains the generality of 'problematic' but, by stipulation, is not a threshold term.

One need not be terribly worried to discover that one has a besmirched belief—in certain epistemic circumstances, one probably *should* believe that one has besmirched beliefs (as may also be true for 'irrational' and 'problematic'). Nevertheless, there is something unfortunate about such beliefs. The best of the best beliefs are unaccompanied by *any* negative higher-order epistemic appraisals—even slight ones. And believing that a belief is besmirched is one way to have such a negative higher-order epistemic appraisal. Accordingly, the following thesis is true:

> **Easy Besmirchment**: Necessarily, if S believes that it is besmirched for S to believe that *p* then it is besmirched for S to believe that *p*.

The slightest stain besmirches: being more besmirched (than whatever) entails being besmirched outright. So, **Easy Besmirchment** escapes our third objection.

Suppose an agent knows that *p* and knows that **Easy Besmirchment** is true. They consider whether to believe, in addition, that it's besmirched for them to believe that *p*. They ask themselves, 'Suppose I were to believe that it is besmirched for me to believe that *p*. Would that belief be true?'

Absolutely! For according to **Easy Besmirchment**, believing that it's besmirched to believe that *p* is enough to make believing that *p* besmirched. Simply having the belief makes it so. And since **Easy Besmirchment** is a necessary truth, it's impossible that the belief could be false. The agent knows that, necessarily, if she herself were to believe <it is besmirched for me to believe that *p*> then she would believe so truly.

The doctrine of infallibility faces a problem: it insists that we are permitted to believe that our beliefs are besmirched just because believing it would make it so. But taking this path makes our total epistemic state worse even if we acquire true beliefs in the process. We're *not* rationally permitted to needlessly downgrade our beliefs in this way—not even lightly.

Let's summarize where we've come so far. Given that the quality of our first-order beliefs is at least somewhat impacted by our higher-order beliefs about them, it's hard to resist the conclusion that there are some infallible propositions such that (a) we can know that they are infallible for us, and yet (b) if we believe them, they will needlessly damage our (potential) first-order beliefs in a way that is not epistemically permissible.

One could, however, preserve the reflectivist intuition that the relationship between our beliefs and our higher-order assessments of those beliefs *matters* epistemically while denying that the way that it matters affects the quality of our first-order *beliefs* themselves. One could insist that what is impacted by negative, higher-order epistemic appraisals is not (necessarily) the corresponding first-order belief but a *network* of beliefs, including at least the first-order belief along with the negative, higher-order assessment.

This is, in many ways, an attractive position. What's bad about believing both <*p*> and <it's problematic for me to believe that *p*>? Part of the answer is that the beliefs do not mesh together as well as they might. There's tension. When the agent has very good reason to believe

both, the best choice might be to live with that tension, but there's tension just the same. The tension metaphor suggests that the problem is with how beliefs (or potential beliefs) fit together and not, in the first instance, with the beliefs themselves. This is at least suggestive of the view above that the thing damaged in easy downgrade cases is a belief system and not necessarily any particular belief. The damage done is wholly holistic.[16]

This seems to make trouble for the recipe for finding infallible propositions. Recall:

> **Infallibility Recipe**: Necessarily, if S believes that it is F for S to believe that *p* then it is F for S to believe that *p*.

The **Infallibility Recipe** prompts us to look for negative properties to ascribe to believing a single proposition. But if the sort of tension involved in the cases discussed so far is not a property of believing any single proposition but a property of a belief *network*, then the **Infallibility Recipe** has us looking in the wrong place.

Before jumping in to plug the hole, let's take a step back. We've already seen that we are not permitted to believe infallible propositions if so believing does needless doxastic damage to another (potential) belief. We are permitted—required, even—to forego the guarantee of truth when doing so protects the quality of our other beliefs in certain ways.

If the goal of safeguarding the epistemic quality of particular beliefs can require us not to believe certain infallible propositions, the goal of safeguarding the epistemic quality of our belief networks should serve just as well. Once we see that there are (at least) two potentially competitive goals in play—truth and maintaining other qualities (rationality, coherence, unproblematicness, etc.) of our beliefs or belief system—the idea that we can be automatically

[16] Matt McGrath and Ernie Sosa both helpfully pressed the importance of this alternative in conversation. This option evokes, for instance, the view expressed by Worsnip: 'Coherence requirements are widescope, and do not speak in favour of individual attitudes simpliciter but rather against particular combinations of attitudes' (2018: 36–37).

licensed to believe in virtue of being guaranteed to attain just one of them starts to sound suspicious.

But let's turn to a specific example. Suppose I form the following belief: <it is problematic for my belief system to include *p*>. On the holistic picture, so believing would *make* it so that there would be tension within my belief system if I also believed that *p*. That tension would not necessarily be a problem for my (potential) belief that *p* itself, but it *would* be problematic for my belief *network*. But of course, that is exactly what I believed in the first place: that it is problematic for my belief system to include *p*. It seems, then, that the following principle is true:

> **Easy Systemic Problems**: Necessarily, if S believes that it is problematic for S's belief system to include *p* then it is problematic for S's belief system to include *p*.

And, of course, **Easy Systemic Besmirchment** is at the ready if concerns about thresholds rise again.

Here, in short, is the problem: If reflectivism is true, then acquiring certain higher-order beliefs can make either our first-order beliefs or else our belief systems worse in some way (without making them false). So, we *shouldn't* form such beliefs when doing so can be easily avoided. But the doctrine of infallibility only cares about accuracy. It wrongly predicts that we *are* permitted to form certain problematic higher-order beliefs anyway, just because we're guaranteed to be right about them.

One could be complacent about this result. Beliefs and belief systems are easily besmirched. And we're far from ideal agents. It might turn out that the vast majority of our beliefs are already besmirched whether we believe that they are or not. And if so, why worry that the doctrine of infallibility permits us to believe that our beliefs are stained in precisely the way they already are?

But such complacency is unmerited. First, if the doctrine of infallibility were true, then even perfect knowers—oracles, supercomputers, gods—would be permitted to believe (truly, once believed) that their beliefs (or belief systems) were besmirched. But surely such powerful agents would not have antecedently besmirched beliefs, even if we mere mortals often do. They certainly shouldn't downgrade their beliefs so needlessly.

But second, when one believes, for example, that <it is problematic to believe that *p*> solely because that proposition is infallible, the proposition is not made permissible *thereby*. The doctrine of infallibility is supposed to be an explanatory thesis: S's knowing that <necessarily, if she herself were to believe that *p* then she would truly believe that *p*> explains why S is permitted to believe *p*. Guaranteed accuracy explains permissibility. But if the doctrine of infallibility only comes out true because we are permitted to believe some infallible propositions for reasons unrelated to their guaranteed truth—because we epistemically frail creatures have antecedently besmirched beliefs—then infallibility does not play the explanatory role that we first thought. The problem is not that we shouldn't believe our beliefs are besmirched (for all I've said, epistemic humility nearly always demands this of us!) but that we clearly shouldn't do so *just because* we'd be guaranteed to be right. Either way, the chain linking guaranteed truth with permission is severed.

5. Direction of Fit Solutions

The problem of easy downgrade shows that the doctrine of infallibility is flawed. Given how unassailable the doctrine appeared at the start, this itself is a significant conclusion. But those sympathetic to the doctrine may hope that its flaws can be mended—or at least safely ignored—in most contexts. In particular, it's noteworthy that the counterexample involves a reversal of the ordinary direction of fit between mind and world. Our aim in believing is (at least

in part) to form a representation in our mind that appropriately matches the outside world—to tailor our minds to fit the world. Ordinarily, the world cares little how our mind represents it. But not always. Sometimes the world tailors itself to fit our representation. And indeed, if any of the defeat principles proposed are true, the quality of our epistemic states depends in some part on how we represent those states to ourselves. Perhaps, then, even though the doctrine of infallibility is false, we can treat the doctrine as true when there's no reversal of the ordinary direction of fit between mind and world—when the way that the world is (in the domain of our proposed belief) does not depend on our beliefs themselves. We replace the original doctrine with this revised principle:

> Doctrine of Infallibility, Dependence Edition: If the truth-value of S's belief that *p* would (if formed) not depend on S's belief that *p* itself, *then* if S knows that <necessarily, if she herself were to believe that *p* then she would truly believe that *p*>, then it is thereby (rationally) permissible for S now to believe that *p*.

The ambition behind this strategy is a sensible one. It seems that there is something right about Descartes's appeal to infallibility in the *Meditations*, and given that the doctrine of infallibility is false, epistemologists should be eager to find a more restrictive principle that allows the Cartesian inference while avoiding the problem of easy downgrade. And since the problematic cases of easy downgrade *are* ones in which the downgrading beliefs explain their own truth, the revised principle is a tempting tactical retreat.

But the problems with the doctrine cannot be excised merely by restricting its domain to cases in which beliefs play no role in their own truth. For we seem to need the doctrine of infallibility (or something that entails it) when dealing with other cases with a reversal of dependence. The revised principle does not neatly divide the good cases from the bad.

Consider again the *cogito*.[17] Before Descartes concludes <I exist>, he concludes <I am thinking>. But <I am thinking> is the sort of proposition whose truth is (at least partially) grounded in the belief itself. If I believe <I am thinking>, my very belief grounds the truth of the believed proposition. Moreover, it is permissible to believe <I am thinking> even if <I am thinking> is the only thought one is having at the moment, so the belief could even be the full grounds for the truth of the proposition. The belief <I am thinking> *depends*—at least partially and potentially fully—on itself for its truth. But <I am thinking> is a paradigmatically good infallible proposition! The dependence edition of the doctrine of infallibility is thus overly restrictive: it does not succeed in neatly distinguishing the good infallible propositions from the bad.

Perhaps, the objector rejoins, this is because we were operating with too wide a notion of dependence. After all, in the case of <I am thinking>, the belief <I am thinking> *grounds* (or perhaps *constitutes*) the truth of the believed content. But grounding isn't the only kind of dependence out there. Perhaps *causal* dependence is the problematic kind:

> Doctrine of Infallibility, Causal Dependence Edition: If the truth-value of S's belief that *p* would (if formed) not causally depend on S's belief that *p* itself, *then* if S knows that <necessarily, if she herself were to believe that *p* then she would truly believe that *p*>, then it is thereby (rationally) permissible for S now to believe that *p*.

I think, however, that we should be suspicious of moves to restrict the epistemically relevant sort of dependence in this way. It's clear why *guaranteeing* the truth is epistemically relevant. But it's not clear why I should care, epistemically speaking, about whether the source of that guarantee is causal or constitutive. A guarantee is a guarantee. Neither kind is less likely to be true than the other.

---

[17] I'm grateful to Ezra Rubenstein for suggesting that I inquire into whether a restriction on beliefs that ground their own truth would solve the problem and to Ernie Sosa for pointing out the relevance of Descartes' <I am thinking>.

Moreover, some of the cases that motivate the doctrine of infallibility in the first place involve causal dependence on a belief. Recall the coin-flip case. In these cases, the agent knows that they will be right whether they believe that the coin will land heads or that the coin will land tails, and the fact that the agent knows this seems to license belief in either proposition. As noted earlier, the most straightforward explanation of this seems to be the doctrine of truth, which entails the doctrine of infallibility. But the coin-flip cases themselves have a reversed causal dependence! The agent's belief—through the demon's intervention—causes the world to match the belief. The correct judgment in the coin-flip case is not, of course, uncontestable.[18] But giving up the permissibility of believing either way in such cases does sap one of the main arguments in behalf of a doctrine of infallibility of its strength.

And even if one goes in for the version of the Doctrine of Infallibility that excludes cases in which the truth of one's belief causally depends on the belief itself, it's unclear that this avoids the problem. When I believe <it is besmirched for me to believe that *p*>, does this *cause* it to be besmirched for me to believe that *p*? Certainly, it's nothing like how the throw of a rock causes the breaking of a window. It's not a relation between events. Nor is it like the coin-flip case, in which a demon causally interferes in the world to guarantee a certain outcome. Rather, there is a conceptual link between my believing <it is besmirched for me to believe that *p*> and the belief's being true: it is a consequence of, among other things, the way 'besmirched' was defined.

I do not take myself to have shown that there is no possible variety of dependence that can divide the good infallible propositions from the bad—that can count as licit the inferences in the *cogito* and the coin-flip case while excluding the inferences central to the problem of easy

[18] See, for instance, discussion in Antill (2020)

defeat. But I hope I have cast doubt on the idea that the doctrine can be easily bandaged by invoking a simple distinction between beliefs that do and do not explain their own truth.

The objector hoped to show that there is something odd or deviant about beliefs that depend on themselves for their truth. It's not hard to enter this frame of mind. Since Anscombe (1957), philosophers have thought that one of the distinguishing features of belief as a mental state is that it has a world-to-mind direction of fit. That there is coffee in front of me (in the world) is a reason for me to *believe* (in my mind) that there is coffee in front of me, whereas my *desire* (in my mind) that there is coffee is a reason for me to make it true that there is a cup of coffee in front of me (in the world). Platt would later argue that 'beliefs should be changed to fit with the world, not vice versa' (1979: 257).[19] Certainly, it'd be very strange to form the belief that there is coffee in front of me and then, because I have that belief, to make a cup of coffee so as to make that belief true.

But I don't think we should extend our suspicion of this strange behavior to all cases in which the relationship between a belief and the truth of its content are entangled. Even in the coin-flip case, we can say that, were the result of the coin-flip different than I had believed—because, say, the demon had misread my mind—my belief would be *mistaken*. From my perspective at least (things may be different for the demon who *desires* that the world match my beliefs), the mistake is with my belief and not the world (cf. Anscombe 1957: 56). Similarly, Humberstone writes that even self-fulfilling prophecies 'involve beliefs with the same direction of fit as any other beliefs, being appraised for correctness ...in terms of how well their content matches how things are with their subject matter' (Humberstone 1992: 71). So, although self-fulfilling beliefs have a different *causal* direction of fit than typical beliefs, their metaphysical

[19] Cf. Williams: '[A] man's word, and his beliefs, should *reflect* things as they are' (1966: 20, emphasis mine).

role as representations of how the world is and their normative success conditions that depend on how the world is in fact remain unchanged. Once we see that beliefs whose truth-value is metaphysically entangled with being held can, nevertheless, bear the ordinary direction of fit between mind and world as other beliefs in *this* sense, the case to restrict the scope of the doctrine of infallibility to 'ordinary' cases loses some of its urgency.

6. The Disbelief Solution

Perhaps we've been focusing too much on aiming for the good outcome of a true belief and not enough on the bad outcome of believing a falsehood. With this thought in mind, we recall that the *cogito* has not just one but two things going for it. First, as we've noted, if one believes that <I exist> then one is guaranteed to be right. But equally, if one believes its negation, <I don't exist>, then one is guaranteed to be wrong (Sosa 2007: 18; Shah 2009: 189).

Moreover, this distinguishes the *cogito* from the easy downgrade propositions of this paper. Notice that <it is besmirched for me to believe that *p*> is *not* such that if I *disbelieve* it, I am guaranteed to be wrong.

This suggests a new doctrine:

> Doctrine of Infallibility, Disbelief Edition: If S knows that <necessarily, if she herself were to believe that *p* then she would truly believe that *p*> *and* S knows that <necessarily, if she herself were to believe that ¬*p*, then she would falsely believe that ¬*p*, then it is thereby (rationally) permissible for S now to believe that *p*.

This successfully saves the *cogito* without endorsing besmirching propositions. But much like the attempted fix by fit, it leaves unexplained why it's permissible to believe in cases when one knows one would be right either way. In Dahlback's case, I can believe that the coin will land heads—seemingly, just on the basis that I know I will be right—even though I will also be right

if I *disbelieve* that the coin will land heads. Once again, this revised principle cannot neatly divide all cases of good infallible propositions from the bad.[20]

7. Concluding Thoughts

We've encountered a puzzle. The doctrine of infallibility seems overwhelmingly plausible. It is the basis for the *cogito*, it makes sense of certain self-fulfilling prophecies, and it gives voice to the enticingly straightforward thought that accuracy is what matters most (epistemically) when deciding what to believe.

But the problem of easy defeat shows that the doctrine of infallibility is false. If reflectivism is true—if reflective attitudes about one's first-order beliefs make at least some difference to the epistemic quality of those first-order beliefs or one's belief system—then there are some guaranteed truths one should refrain from believing.

Does the falsity of the doctrine of infallibility lead to scepticism? Perhaps if one comes into the problem in a Cartesian mood. If one can't automatically trust even infallible propositions (the disillusioned Cartesian asks), what can we trust? But most epistemologists have (wisely) not demanded that some of our beliefs must be infallible to count as knowledge.

No, the primary puzzle is not, 'How can we really know that we exist or know that the coin will land heads if the doctrine of infallibility is false?' We were (rightly) pre-theoretically confident that these were good judgments, and we need not abandon them just because the principle that we thought explained their permissibility turned out to be false. Rather, the puzzle is how to separate the good infallible propositions from the bad. We're left to wonder: why wasn't the guarantee of truth good enough to license belief? What was so valuable, epistemically, that it was worth foregoing a guaranteed true belief?

---

[20] I'm grateful to Ernie Sosa for conversation on this and related points.

I see no easy answer to this question.

If this paper is right, we are left with a broken, false doctrine that had seemed foundational to Cartesian epistemology and left without an obvious way to repair it. Is there anything positive we have learned?

I conclude by suggesting two modest lessons. First, we learn that there's tension between veritism and reflectivism. Perhaps this shouldn't surprise us so much in the end. Veritism (as we are using the term) says that what matters most at bottom is accuracy. Reflectivism says that something else matters too: namely, how our first-order beliefs and higher-order beliefs fit together. Nevertheless, one might have thought that the reason it matters how our different levels of beliefs fit together is because such relationships can help us to become more accurate. But although this might be part of the story, it can't be the whole story. For believing easy-downgrade propositions on the basis of their infallibility guarantees accuracy while damaging reflective fit.

Second, we have found a surprising argument against Cartesian infallibilism. Cartesian infallibility (and the certainty it engenders) is often taken to be too stringent a requirement for either knowledge or proper belief. But if Cartesian infallibility seemed extreme, it at least also seemed like a natural stopping point. What more could one hope once infallibility had been achieved? What greater epistemic assurance? If infallibility seemed too stringent to be necessary for permissible belief, it at least seemed obviously sufficient for it.

But we've learned that infallibility isn't sufficient for permissible belief: sometimes, one shouldn't believe even when one knows one would be right. Fallibilists—who already believed that infallibility was not *necessary* for right belief—may feel justly emboldened knowing that it

isn't *sufficient* for permissible belief either. It isn't the natural cut-off point in epistemic normativity that we've been led to believe.

Infallibility isn't always worth having even in those rare cases when it is there to be had.[21]

References

Alston, W. (1971) 'Varieties of Privileged Access', *American Philosophical Quarterly* 8 (3): 223-41.

Anscombe, G.E.M. (1957) *Intention*, Oxford: Basil Blackwell.

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

Berker, S. (2013a) 'Epistemic Teleology and the Separateness of Propositions', *Philosophical Review* 122: 337–94.

Berker, S. (2013b) 'The Rejection of Epistemic Consequentialism', *Philosophical Issues* 23: 363–87.

Brown, J. (2013) 'Infallibilism, evidence and pragmatics', *Analysis* 73/4: 626–35.

Coates, A. (2012) 'Rational Epistemic Akrasia', *American Philosophical Quarterly* 49/2: 113–24.

Dahlback, M. (forthcoming) 'Infinitely Permissive', *Erkenntnis*.

Descartes, R. (1984) *The Philosophical Writings of Descartes, Vol. II*, Translated by J. Cottingham, R. Stoothoff, & D. Murchoch, Cambridge: Cambridge University Press.

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

Fantl, J. & M. McGrath (2009) *Knowledge in an Uncertain World*, Oxford: Oxford University Press.

Firth, R. (1981) 'Epistemic Merit, Intrinsic and Instrumental', *Proceedings and Addresses of the American Philosophical Association* 55/1: 5–23.

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

Goldman, A. (1999) *Knowledge in a Social World*, Oxford: Oxford University Press.

Goldman, A. (2001) 'The Unity of the Epistemic Virtues.' In A. Fairweather and L. Zabzebksi (eds.) *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, Oxford: Oxford University Press: 30–48.

Horowitz, S. (2014) 'Epistemic Akrasia', *Noûs* 48/4: 718–44.

Humberstone, I.L. (1992) 'Direction of Fit,' *Mind* 101/401: 59–83.

Jackson, F. (1973) 'Is There a Good Argument against the Incorrigibility Thesis?', *Australasian Journal of Philosophy* 51/1: 51–62.

James, W. (1907) *The Will to Believe and Other Essays in Popular Philosophy*, New York: Longmans Green and Co.

Kavka, G. (1983) 'The Toxin Puzzle', *Analysis* 43/1: 33–36.

[21] My thanks go to Charity Anderson, Robert Audi, D Black, Laura Callahan, Adam Carter, Charles Côté-Bourchard, Andy Egan, Megan Feeney, Will Fleisher, Carolina Flores, Danny Forman, Christopher Frugé, Adam Gibbons, Caley Howland, Chris Kelp, John Komdat, Ting-An Lin, Neil McDonnell, Matt McGrath, Ram Neta, Dee Payton, Julian Perlmutter, Pamela Robinson, Ezra Rubenstein, Susanna Schellenberg, Mona Simion, Kurt Sylvan, and Isaac Wilhelm for feedback on drafts of this paper. Finally, I cannot be wrong in giving special thanks to Ernie Sosa, whose work on infallibility piqued my interest in the topic, and conversation with whom has shaped each layer of this project.

Kelly, T. (2014) ‘Evidence Can Be Permissive’, in M. Steup, J. Turri, & E. Sosa (eds.) *Contemporary Debates in Epistemology*, 2nd edition, Oxford: Wiley-Blackwell: 298–313.
Kopec, M. (2015) ‘A Counterexample to the Uniqueness Thesis’, *Philosophia* 43: 403–9.
Lasonen-Aarnio, M. (2010) ‘Unreasonable Knowledge’, *Philosophical Perspectives* 24: 1–21.
Lebens, S. (2015) ‘God and his Imaginary Friends: A Hassidic Metaphysics’, *Religious Studies* 51/2: 183–204.
Lewis, D. (1996) ‘Elusive Knowledge’, *Australasian Journal of Philosophy* 74/4: 549–67.
Platts, M. (1979) *Ways of Meaning*, London: Routledge and Kegan Paul.
Raleigh, T. (2017) ‘Another Argument Against Uniqueness’, *Philosophical Quarterly* 67/267: 327–46.
Reisner, A. (2009) ‘The possibility of pragmatic reasons for belief and the wrong kind of reasons problem’, *Philosophical Studies* 145: 257–72.
Reisner, A. (2013) ‘Leaps of Knowledge,’ in T. Chan (ed.) *The Aim of Belief*, Oxford: Oxford University Press.
Reisner, A. (2014) ‘A Short Refutation of Strict Normative Evidentialism’, *Inquiry* 5: 1–9.
Roush, S. (2003) *Tracking Truth*, Oxford: Oxford University Press.
Schroeder, M. (2012) ‘The Ubiquity of State-Given Reasons’, *Ethics* 122/3: 457–88.
Shah, N. (2009) ‘The Normativity of Belief and Self-Fulfilling Normative Beliefs’, *Canadian Journal of Philosophy* 39/SI: 189–212.
Shoemaker, S. (1963) *Self-Knowledge and Self-Identity* Ithaca: Cornell University Press.
Sosa, E. (2007) *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, Oxford: Oxford University Press.
Sylvan, K. (2018) ‘Veritism Unswamped’, *Mind* 127/506: 381–435.
Velleman, J.D. (1989a) ‘Epistemic Freedom’, *Pacific Philosophical Quarterly* 70: 73–97.
Velleman, J.D. (1989b) *Practical Reflection*, Princeton: Princeton University Press.
Velleman, J.D. (1992) ‘The Guise of the Good’, *Noûs* 26/1: 3–26.
Velleman, J.D. (2000) ‘The Possibility of Practical Reason’, Oxford: Oxford University Press.
Weatherson, B. (manuscript) ‘Do Judgments Screen Evidence?’ <http://brian.weatherson.org/JSE.pdf> accessed 13 November 2020.
Willard-Kyle, C. (2017) ‘Do great minds really think alike?’, *Synthese* 194/3: 989–1026.
Williams, B.A.O. (1966) ‘Consistency and Realism’, *Proceedings of the Aristotelian Society, Supplementary Volumes*, 40: 1–22.
Williamson, T. (2000) *Knowledge and Its Limits*, Oxford: Oxford University Press.
Williamson, T. (2014) ‘Very Improbable Knowing’, *Erkenntnis* 79: 971–99.
Worsnip, A. (2018) ‘The Conflict of Evidence and Coherence’, *Philosophy and Phenomenological Research* 96/1: 3–44.

Rutgers University, USA