

Reflection Deflated

Abstract. *Reflectivism* says you should be guided by your reflective beliefs about which doxastic attitudes are rational. You might for example revise a belief you take to be rationally prohibited, or adopt one you think is required. Against reflectivism I support *deflationism*. It says reflective beliefs should guide you only by providing ordinary evidence. If you think you are rationally required to believe it will rain, for example, that should guide you to believe it only if your being required to is evidence of rain. Despite endorsing deflationism, I argue it is more radical than it might seem. Ordinary standards of evidential support assign no special significance to first-personal normative premises about what I am required to believe. And this means deflationism leaves no room for me to adopt beliefs just because I take myself to be required to.

1. Introduction

Philosophers have narcissistic tendencies. Maybe this is true in multiple senses, but I mean it in an epistemic one: We are preoccupied with the rationality our own beliefs. Sure enough, we try to discover the truth and to persuade each other of it where we can. But if truth is too much to ask for, we hope at least to have good reasons for our beliefs. And when persuasion fails, we fall back on exhibiting to each other the coherence and systematicity of the views we have stitched together.

A little narcissism is probably forgivable, given what we do for a living. But we also have a tendency to hold up our own idiosyncrasies as a rational ideal, and build a preoccupation with one's rationality into what it is to be rational. Of course, the advocates of epistemic narcissism do not put it that way. They defend what they call *reflection* (or critical reasoning). As Declan Smithies explains it:

[T]he point of reflection is to bring your beliefs into line with higher-order reflections about which beliefs you have justification to hold. The reflective process has two stages: first, reflecting on which beliefs you have justification to hold; and second, revising your beliefs in light of these reflections.¹

Some authors take rationality to be impossible without reflection.² But most accept that some degree of rationality is attainable even if your attention remains directed at the world around you. If you know it rains when the barometer reads 'low', and can see that the barometer reads 'low', then you can, based solely on these worldly facts, adopt the belief that it will rain. But this kind of world-directed reasoning is still disparaged as exemplifying a kind of low-grade rationality. It is what we have in common with children and animals. It is "automatic" or "unconscious" or "offline."³

¹ Smithies 2019.

² E.g., Shoemaker 1996, pp. 32-33

³ E.g., Alston 1989, pp. 221-236; Burge 2013; McGinn 1982, pg. 20; and Smithies 2019.

We reflective agents in contrast have some appreciation of what rationality's requirements are, and of how well our current beliefs live up to them. And when we judge that our beliefs do not live up, we can modify them accordingly. This is what Smithies called the second stage of reflection. Tyler Burge puts it this way:

Essential to carrying out critical reasoning is using one's knowledge of what constitutes good reasons to guide one's actual first-order reasoning.⁴

[I]f, in critical reasoning, one correctly and with warrant judges that a lower-level state is (or is not) reasonable, then it rationally follows directly that one has reason to sustain (or change) the lower-level state.⁵

These authors make ambitious claims about the significance of reflection. While common, they are controversial.⁶ But what is less controversial is that, if one does happen to consider whether one's belief is rational, and one judges that it is not, then one ought to revise it, or will at least find oneself under some pressure to do so. As I will put it, the idea is that reflection should *motivate or guide* our attitudes.

This modest claim is what I call *reflectivisim*. It often turns up in discussions without the broader agenda, as a well-known truism cited in support of some other point.⁷ Here is Richard Moran discussing doxastic agency:

If a person is at all rational, his first-order beliefs will indeed be sensitive to his second-order beliefs about them, and will change accordingly. He may, for instance, discover some set of his beliefs is inconsistent, or suspect that a particular belief of his is the product of prejudice or carelessness, or, at the limit case, that it is just plain false. His first-order beliefs will then normally change in response to his interpretation of them.⁸

And here is Ralph Wedgwood, discussing the aim of belief:

[T]he concept "rational belief" is normative for the practice of theoretical reasoning. It is a constitutive feature of this concept that if one judges that a certain belief would not be a "rational" belief for one to hold, this judgment commits one to not holding that belief.⁹

Debates between internalists and externalists are rife with similar claims, and not just on the internalist side.¹⁰ Knowledge-first externalists think the norm for belief is knowledge,

⁴ Burge, 2013, pg. 74.

⁵ Burge 2013, pg 372. See also pg. 82.

⁶ For opposition, see Arpaly 2002, Broome 2013, and Kornblith 2012.

⁷ E.g., Elga 2005, pg. 116; Fogal 2020, pg. 1041; Scanlon 2007; and the authors cited below.

⁸ Moran 2001, pg. 55.

⁹ Wedgwood 2002, pg. 271.

¹⁰ Boghossian 2008, pg. 472; BonJour 1985, pp. 7-8; Dogramaci 2015; and White 2006, pg. 539. And see Goldman 1999 and Hughes forthcoming for review and critical discussion.

rather than some internalist requirement of rationality. But many still assign importance to an agent's reflections on whether her beliefs meet this externalist norm. For example, some say it is irrational to believe what you take yourself not to know.¹¹ Others identify justified belief with belief which for all you know is knowledge.¹²

But reflection has probably received the most sustained recent attention in connection with peer disagreement, epistemic akrasia, and higher-order evidence.¹³ Compare with these other authors what David Christensen says in his early work on the topic:

Intuitively, there seems to be a connection between what one is rational to believe, and what one is rational to believe one is rational to believe. ... Putting the thought in terms of justification, the idea is that (justified) higher-level doubts about the justification of one's belief that *p* can defeat one's justification for believing *p*.¹⁴

Even prominent critics of these views often grant what I call 'reflectivism', like Maria Lasonen-Aarnio when she says:

Consider the reason *I am required, given my epistemic reasons, to believe p*, or simply *I am rationally required to believe p*. These are conclusive reasons to believe *p*. Similarly, *I am forbidden, given my epistemic reasons, to believe p*, or simply *it is irrational for me to believe p*. These are conclusive reasons not to believe *p*.¹⁵

But I think we should ditch the whole idea of reflection. If you are wondering whether it will rain, you should take a look out there in the world, and consider what the barometer is reading, or what the clouds are like overhead. Of course you also are part of the world, and there is no barrier to your considering what is going on with you and your beliefs. But if you do, these will just be further items in the world to be weighed alongside the others.

This *deflationist* view fits the common slogan that higher-order evidence is just more evidence. But I am not sure everyone who likes the slogan understands it the same way.¹⁶ If higher-order evidence is just more evidence, that does not merely mean it can be outweighed by other considerations. It means it is relevant to belief only by the same general norms or standards that apply to any other evidence.¹⁷ I will argue this has more radical implications

¹¹ E.g., Littlejohn 2012, pp. 2-3 and Littlejohn and Dutant 2021.

¹² E.g., Carter and Goldstein 2021.

¹³ E.g., Christensen 2010a, 2010b, 2016; DiPaolo 2018; Feldman 2005; Littlejohn and Dutant 2021; Elga 2005; Greco 2014; Horowitz 2014; Kelly 2010; Lasonen-Aarnio 2014 and 2020; Smithies 2019, Chs. 9-10; Weatherson 2019, Worsnip 2018

¹⁴ Christensen 2010a. As I discuss below, Christensen has revised his view. But the new view still partly resembles reflectivism, and anyway these sentiments remain common (e.g., Elga 2005, Feldman 2005, Greco 2014 and 2019, Horowitz 2014, Salow 2019, Worsnip 2018).

¹⁵ Lasonen-Aarnio 2020. See also Arpaly 2002, Bergman 2005, Broome 2013, and Dorst 2019.

¹⁶ Cf. Lasonen-Aarnio 2014, 318.

¹⁷ Cf. Whiting 2019, p. 250.

than sometimes acknowledged. For it means that reflective beliefs and higher-order evidence cannot really guide us. If reflectively judge I am required to believe it will rain, that does not serve as a pointer or guide in settling on the belief, any more than a belief about someone else's requirements does. If it matters at all, it is only incidentally, by providing some indication of the weather.

2. The Argument for Deflationism

Reflectivists want our reflective beliefs to motivate or guide us. But this rough idea leaves a lot open. It might not mean that the overall most rational thing is always to follow reflective beliefs wherever they lead. If I reflectively believe my evidence supports the divinity of pickles, maybe what I really should do is give up this reflective belief, not follow through on it.¹⁸ This is one reason some reflectivists pitch their view in terms of higher-order evidence or justification rather than belief,¹⁹ while others speak not of the rationality but the reasonableness of being guided by reflective beliefs.²⁰ I hope mostly to stay neutral on these finer points, with a few exceptions I will mention. The basic reflectivist idea I oppose is just that higher-order reflections, however construed, differs from ordinary evidence in offering us a kind of advice or guidance.

We will see some common motivations for this idea later, but here is one for starters: Rationality is normative, and reflective beliefs are first-personal. So my reflective beliefs in effect concern what I myself ought to believe. That sets them apart from beliefs about barometers, or even other people. If I am deliberating about whether it will rain, what I take to be the barometer's reading is one factor to consider. But if I take myself to be required to believe it will rain, then as I see things the very deliberations I am now engaged in ought to terminate in belief. It might seem this should lead me to make sure they do.

This motivation draws on controversial but plausible connections between first-personal normative beliefs and motivation. Suppose my friend is accused of a crime, and I believe that I am required to publicly defend her innocence. Arguably that should motivate me to do it, or at least gives me some reason to. But even if normative beliefs should motivate actions, it is not so clear the same goes for belief. Suppose I also think I am required to *believe* my friend is innocent. Does this give me a reason to believe?²¹ Maybe in some sense it does, but it still strikes us as the wrong kind of reason to believe. It would be irrational to believe for a reason like that.

Reflectivists need beliefs about what is *rationally* required to be different, and it is easy to see why it might seem this way. Maybe you can be morally required to believe something that you have no evidence for. But believing something because you think it is rationally required is at least pretty close to believing it because you think your evidence supports it. (And if not, reflectivists are free to just say you should be guided by beliefs about evidence instead.) Of course, believing something on the basis that your evidence supports it is not quite the same as believing on the basis of the evidence itself. But if evidence supporting a proposition is

¹⁸ Cf. Arpaly 2002 and Smith 1994.

¹⁹ E.g., Christensen 2010a and 2010b, Feldman 2005, Littlejohn and Dutant 2021, and Smithies 2019.

²⁰ Lasonen-Aarnio 2014 and 2020. For related views, see Worsnip 2018.

²¹ Cf. Stroud 2006 and Rioux forthcoming.

the right kind of reason to believe it, it might seem the fact that your evidence supports it is close enough. What other kind of reason is this fact supposed to be, a moral or prudential one?

But I think this talk of different kinds of reasons—epistemic and practical, or the right kind and the wrong kind—has the potential to mislead. The right and wrong kinds of reasons are not just two types of considerations bearing on the same deliberative question, like personal and professional reasons for moving to a new city. Instead, they are considerations relevant to distinct kinds of deliberation or reasoning. Consider an example:

Cash's Prize: Cash is offered a large cash prize for believing it will rain. Cash can press a button that will cause himself to believe it will rain.

Should Cash adopt the belief that it will rain? Will doing the rational thing mean believing this? Many of us are reluctant to give univocal answers to these questions. We want to respond with some sort of distinction, and that is where talk of different kinds of reasons comes in. On the one hand, we want to grant that the prize gives Cash a good reason to bring it about that he believes, by pressing the button. But on the other, we want to say that this does not make it a reason to believe, or at least not a reason of the right kind. But why not? Cash prizes are perfectly good reasons for doing most anything else. Why not believing?

At least part of the answer appeals to a plausible connection between reasons and reasoning (or deliberation). When Cash knows that there is a prize for belief, or even if he merely believes there is one, this might potentially lead to him believing it will rain. In principle, there are many ways this might happen. Cash might have a microchip in his brain that scans for beliefs to the effect that some attitude would be beneficial, and then produces the attitude. But if that is how Cash's knowledge of the prize causes the belief that it will rain, it will just be an incidental factor in the belief's etiology. For his knowledge of the prize to make the belief rational, it needs to *motivate* the belief, in the sense of providing a basis, or motivating reason.

One way for the prize to be Cash's motivating reason is the old-fashioned way. Rather than using the button, Cash might simply consider the question whether it will rain, and be persuaded that it will by the fact that he will get a prize for believing it. Put another way, Cash might *infer* that it will rain from the premise that he will get a prize for believing it. In doing so, he would reason in a way he might express by saying "I'll get a prize for believing it will rain, so (probably) it will rain."

But reasoning like that would be irrational. Why? Well, the fact that he will get the prize is not evidence that it will rain, at least not by ordinary standards. It does not deductively entail that it will rain, or inductively or abductively support it, or anything like that. That will make at least this route to the belief irrational, if we think ordinary evidential standards like these govern inferential reasoning.

Even pragmatists, who think it is unambiguously rational for Cash to believe it will rain, do not deny this much. For example, when Pascal advocated belief that God exists on the basis that it maximizes expected utility, he did not expect you to infer God's existence from this premise. Rather, he wanted to motivate you to produce the belief in yourself, for example by going to church. Likewise, modern day pragmatists like Susanna Rinard do not want Cash to achieve belief by considering whether it will rain, and inferring that it will from

a premise about the prize.²² They want him to consider whether to believe it will rain, and to decide to do so based on that premise. In doing this, Cash would reason in a way he might express by saying “I’ll get a prize for believing it will rain, so I’ll go ahead and do it.”²³

It is hard to find anything irrational in Cash’s deliberating like this, or even in following through, and pressing the button. But that does not mean pragmatists are right that it is rational for him to believe. It might just mean it is rational for Cash to cause himself to believe irrationally. Sure enough, the prize will be Cash’s reason for pressing, and his pressing will be part of the etiology of the belief. But that does not make it the reason for which Cash believes.²⁴

Let’s sum this up as a valid argument that a cash prize is the wrong kind of reason for belief. First, there is the premise even pragmatists should accept, that it is irrational to infer it will rain from a premise about the prize, and more generally that:

(EVIDENTIALISM) The apparent fact that r is a premise relevant to p in inferential reasoning only insofar as it is ordinary evidence regarding p .²⁵

Second, there is the premise pragmatists like Rinard dispute, that motivating reasons for belief must be premises in inferential reasoning. More generally,

(INFERENTIALISM) The apparent fact that r can be a reason for which one rationally holds a belief about p only insofar as r is a premise relevant to p in inferential reasoning.

Assuming the prize is not evidence of rain, it follows that it is not the right kind of reason to believe it will rain, the kind of reason that might rationally motivate Cash to believe it will rain. But these two premises alone do not mean that Cash cannot rationally believe that it will rain, just that he cannot do so for the reason that he will get the prize. A pragmatist might take a different tack from Rinard, and say that it does not matter whether the prize ends up qualifying as the motivating reason for his belief. Even if Cash ultimately believes for no reason at all, maybe there remains a sense in which his belief was *guided* by his knowledge of the prize. And that, this pragmatist says, is enough for the beliefs to be rational. To rule out a view like that, we need one further premise. It says:

(MOTIVATIONALISM) One’s belief that r can rationally guide one’s belief about p only insofar as r can be a reason for which one rationally holds a belief about p .

From these three premises, it follows that Cash’s belief that he’ll get a prize cannot rationally motivate or guide his belief that it will rain without being evidence.

²² Rinard 2017 and 2019.

²³ Cf. Whiting 2019, 259.

²⁴ Cf. Hieronymi 2005 and 2006, Kelly 2002, Shah 2006, Rinard 2017 and 2019, and Way 2016.

²⁵ Cf. Jim Pryor’s (2005) ‘Premise Principle’. My discussion of EVIDENTIALISM below is indebted to extensive discussion with him.

While some of the details of this argument are my own, I hope the main contours are familiar. My goal isn't to give a new argument against pragmatism, but to apply broadly evidentialist commitments consistently to reflective beliefs.²⁶

Consider an agent who judges that one of her existing beliefs is not really supported by her evidence. Or an agent who is persuaded that some doubts he harbors are irrational, and that really he ought to believe. How should these agents respond? The same three premises just considered imply:

(DEFLATIONISM) One's reflective belief that one is required to (not) believe p can rationally guide one's belief about p only insofar as the apparent fact that one is required to (not) believe p is ordinary evidence regarding p .

Deflationism is so called because it deflates the epistemic significance of reflective beliefs. When I deliberate about whether p , and consider the facts as I see them, deflationism says the apparent facts about what this one measly person here ought to believe are relevant to my deliberations in the same way as the apparent facts about what others ought to believe, or for that matter about shoes and ships and sealing wax.

This formulation of deflationism, and the premise EVIDENTIALISM, invoke a notion of 'ordinary' evidential standards. I will have more to say in Section 5, but as a first pass, think of an informal logic exam question that asks "Which of the following assumptions, if added to the existing premises, would support that p ?" A good answer might deductively entail that p , or provide inductive, abductive, or probabilistic support for it. This list of course might not be exhaustive. But illustrates a kind of answers that might be appropriate.

In contrast, these ordinary standards do not count the premise that I morally ought to believe p as supporting p . And arguably any alternative standard counting this premise as 'supporting' p would be pretty different from the ordinary ones. One reason is that ordinary standards of evidence arguably do not assign a special role to *normative* premises or evidence. For example, the validity of a syllogism does not turn on whether the major term is a normative or non-normative predicate. And something similar seems plausible for non-deductive inference, though I will not argue for it here. If normative beliefs have special motivational force, it is arguably not because normative premises amount to especially strong evidence.

A second reason, which will be more important, is that ordinary standards are *impersonal*. Roughly speaking, they do not carve out a distinctive significance for evidence about particular individuals, including oneself.

Let's say we *anonymize* an agent's beliefs and evidence when we replace any *de se* beliefs or evidence with corresponding third-person ones. For example, if my evidence includes that I am uncertain whether it will rain, and my beliefs include one that I am a lousy forecaster, then my anonymized beliefs and evidence include the evidence that DB is uncertain whether it will rain and the belief that DB is a lousy forecaster. If it helps, you might imagine the evidence of a guardian angel who, with no regard to herself, has followed me throughout my life. She knows everything I know, including about my mind, and she believes everything I believe. But all of the knowledge and beliefs about me are cast as if about some anonymous

²⁶ For related thoughts, see the excellent Whiting 2019, which is especially recommended for its discussion of reasons and motivation.

third party. Or instead, you might suppose there is no such thing as *de se* thought, or else that I am incapable of it. I might still know all the same facts about myself that I now know, but conceive of them just as facts about DB.

Here is the sense in which ordinary evidential standards are impersonal. Suppose that my background evidence is such that, when I learn that I allowed the murderer to return to the crime scene, I can deductively infer that the butler did it. If so, then if my background evidence were anonymized, I still can deductively infer the butler did it if I learn that DB allowed the murderer to return to the crime scene. For my original evidence must entail that I allowed the butler to return to the crime scene, and so when anonymized will entail that DB allowed it. Or suppose that when I learn the butler lied to me about his whereabouts, I can abductively infer that the butler did it. Then if my background evidence is anonymized, I can abductively infer the same thing if I learn the butler lied to DB about his whereabouts. Evidential standards are impersonal when they are like that.

Since ordinary standards governing deductive, inductive, and abductive inferences are impersonal, the apparent fact that I ought to believe *p* does not support that *p* by those standards, at least no more so than that DB ought to believe *p*. But Sections 3 and 4 will argue that reflectivists cannot accept this. They need reflective beliefs to motivate or guide us in a way that is not licensed by impersonal standards, and that means denying one or more premises in the argument for deflationism, as discussed in Sections 5-7.

3. Motivations for Reflectivism

When I tell people I endorse deflationism, I get two immediate objections. The first is that deflationism is obviously false, and the second is that everyone already accepts it. Most of this paper responds to the first objection. But before getting back to it in Sections 5-7, let me address the second.

It sure does not sound to me like everyone accepts deflationism. The authors I quoted in Section 1, for example, do not seem merely to be observing that premises about one's own rational requirements can amount to evidence. And while I won't bombard you with more quotations, reflectivist thinking seems to me widespread. For what it is worth, as a deflationist I routinely feel at odds with discussions of 'reflective beliefs', 'higher-order evidence', 'epistemic levels', 'metajustification', and the like.²⁷ If I manage to convince you of deflationism, then maybe you will, too.

Not to be too dramatic about it, but reflectivism also seems to me presupposed by a natural conception of the whole enterprise of epistemology. At least as it is often pitched to students, epistemology is the branch of philosophy for figuring out what to believe, just like ethics is the branch for figuring what to do. Such a 'meliorative project' was certainly an ambition of many historical figures. Today's epistemologists might be less intent on reforming your beliefs, but many still haven't given up on it. Yet under deflationism, there is no room for theoretical knowledge about epistemology as such to guide us in our other beliefs. Epistemological inquiry might help us figure out what to believe *about epistemology*, but not much else.

²⁷ Some exceptions are Owens 2017 and Whiting 2017 and 2019.

All that said, a contrary impulse towards deflationism also seems to me common, including among some of the authors I have cited as reflectivists.²⁸ It may be that on further thought, many of them would prefer deflationism after all. But I do think the anti-deflationist tendencies go beyond some stray remarks. Hostility to deflationism seems to me baked into the central motivations commonly cited for reflectivism.

This is perhaps clearest for ambitious reflectivists like Burge and Smithies, who claim that reflection is part of what makes us persons, and distinguishes us from less sophisticated agents.²⁹ The problem is that if the brutes can make ordinary deductive and inductive inferences about cabbages and kings, it is hard to see why they would struggle to apply the same impersonal rules or standards to premises about their own rational requirements. Perhaps this peculiar kind of mental block is possible, but it seems an unlikely candidate for a deep division between kinds of rational beings.

Reflectivists without this broader agenda are often driven by more piecemeal intuitions and arguments. But the piecemeal motivations usually end up opposing deflationism, too.

We already saw one such motivation, appealing to the normativity of rationality. If I think some belief is rationally required, maybe that is like thinking I *ought* to hold it. And maybe thinking that I ought to do something should motivate me to do it.³⁰ Sometimes this idea is reinforced by an analogy with practical akrasia.³¹ Take the smoker who judges that she ought to quit, but keeps smoking anyway. Her behavior seems irrational. Maybe that is because, in judging she ought to quit, she commits herself to quitting, in some sense requiring a corresponding motivation.³²

I said in Section 2 why I am not persuaded by this, and I will say some more in Section 6. Here I just want to emphasize that this motivation for reflectivism opposes deflationism, for two reasons. First, premises about what I *ought* to do are not stronger evidence than non-normative substitutes. Maybe the smoker's normative judgment commits her to quitting, but not because its content is evidence for something she didn't already know.³³ Second, premises about what *I* ought to do are not stronger evidence than substitute third-personal premises about what *DB* ought to do. Yet assuming these judgments should motivate at all, their first-personal character is as essential as their normative character. If I judge that *I* ought to see to it that you win our tennis match, that is supposed to motivate me to throw the match. Not so if I judge that *you* ought to see to it that you win our tennis match.

²⁸ E.g., Christensen 2016; Lasonen-Aarnio 2014, pp. 317-318; and Declan Smithies in correspondence.

²⁹ E.g., Alston 1989, Burge 2013, and Smithies 2019.

³⁰ Wedgwood 2002.

³¹ E.g., Greco 2014; Horowitz 2014; Smithies 2019, Ch. 9; Titelbaum 2015 and 2019, pg. 228.

³² For review, see Rosati 2016, Sec. 3.

³³ Cf. Harman 2015, 61.

Aside from finer points about normative belief and motivation, many reflectivists just find it intuitively incoherent to combine a belief that p with a reflective belief that one ought to doubt that p instead.³⁴ They accept something like:

(REFLECTION REQUIREMENT) Rationality requires that if you believe you are rationally required to hold some doxastic attitude, then you do hold the attitude.

The intuition is sometimes reinforced by appeal to Moore's paradox, with the suggestion that it is irrational to believe Moorean conjunctions like *It will rain, but I am required to doubt that it will rain*.³⁵

Reflectivism offers a natural explanation of these intuitions. Suppose I believed the second Moorean conjunct, which says I am required to doubt that it will rain. Under reflectivism, I will have a reason to doubt the first conjunct. So I cannot believe the Moorean conjunction without having a reason to doubt it.

It is hard to square all this with deflationism, however. I can rationally believe third-personal Moorean conjunctions like *It will rain, but DB is rationally required to doubt that it will rain*. Likewise, I might think DB is required to doubt it will rain without having a reason to do so myself. Maybe things are different if I think I am the one required to doubt, but not because this is stronger evidence about whether it will rain.

Does this mean deflationists should reject REFLECTION REQUIREMENT? I reject it, but that is not baked in to the view. For it means rejecting not only reflectivism, but also:

(LUMINOSITY) You can always know whether you are rationally required to hold a given doxastic attitude.

If you are in a position to know that p , then arguably it is irrational to believe that not- p . If so, then LUMINOSITY will guarantee the truth of REFLECTION REQUIREMENT in an entirely "bottom up" fashion. It will be irrational to believe a given attitude is required if it is not, just because your higher-order beliefs should track their first-order subject matter.³⁶

Reflectivism is consistent with LUMINOSITY, and may be a natural compliment to it. But it's a distinct thesis. It alleges a "downward" pressure that higher-order beliefs are supposed to exert on your first-order attitudes. If you believe an attitude is required, the reflectivist says that should guide you to adopt it. It is that claim, not REFLECTION REQUIREMENT alone, that I say is incompatible with deflationism.³⁷

At the same time, I think the difference between REFLECTION REQUIREMENT and reflectivism can be overstated. For one thing, deflationism still undermines one obvious motivation for LUMINOSITY.³⁸ Under reflectivism you *had better be* able to know what

³⁴ E.g., Smithies 2019, pg. 314 and Worsnip 2018.

³⁵ E.g., Greco 2014; Gibbons 2013, Ch. 9; Horowitz 2014, pg. 725; and Smithies 2019.

³⁶ Cf. Smithies 2019 and Titelbaum 2015.

³⁷ For some reflectivism-adjacent views that I think combine luminosity assumptions with further claims the deflationist can accept, see Briggs 2009, Egan and Elga 2005, and Titelbaum 2015.

³⁸ See, e.g., Burge 2013, Ch. 3 and some authors surveyed by Goldman 1999.

attitudes are required, or else you will be in a bad situation. You might end up with otherwise rational beliefs that you suspect of irrationality, for example, or irrational beliefs which you have no way of knowing are irrational. But by deflationist lights, there is nothing so bad about these situations. Falsely suspecting a belief is irrational won't undermine its rationality. And realizing a belief is irrational wouldn't have helped you get rid of it anyway.

For another thing, supporters of LUMINOSITY should admit that false reflective beliefs are possible. A generally rational agent might consider her evidence, for example, and misjudge that it supports rain. Maybe her reflective belief must be less than ideally rational, but it is certainly realistic. So we are still left with the question how this agent should proceed, given the reflective belief she already holds. And so the detour through LUMINOSITY merely delays the central question at issue between reflectivists and deflationists.³⁹

There is one more motivation for reflectivism, stemming from examples of higher-order evidence. These are up next.

4. Higher-Order Evidence and Deflationism

Deflationism says reflective beliefs should not motivate or guide you, any more than moral beliefs should. But suppose I think one can never be morally required to believe a falsehood, and an oracle tells me I am morally required to believe my friend is innocent. I might infer she is innocent. Does this show that moral beliefs can guide me after all? Not really. For my being morally required just amounts to evidence of my friend's innocence. This evidence concerns what I ought to believe, but only incidentally.

The same goes for reflective beliefs. Like anything else, apparent facts about what you are rationally required to believe can be evidence. And to that extent, they can motivate belief. This is important because rational requirements to believe routinely have significant evidential weight. In a familiar slogan, evidence of evidence is evidence.⁴⁰ But we must distinguish this boring truth from the interesting falsehoods reflectivists endorse. For the boring truth, consider:

Sherlock's Secret Evidence: While investigating a murder, Sherlock tells Watson that he has gathered such strong evidence that he, Sherlock, is rationally required to believe the butler did it. Watson accepts Sherlock's claim.

Sherlock's being required to believe the butler did it is evidence that the butler did it, just like a fingerprint match might be. While misleading evidence can sometimes require false beliefs, we hope this is uncommon. So when Watson accepts that Sherlock is required to believe the butler did it, Watson gets a reason to believe it himself. But that is just because he gets evidence supporting the butler's guilt the usual way, by providing deductive, inductive, or abductive support for it. Compare this to:

Sherlock's Admonition: While investigating a murder, Watson is confronted with inconclusive evidence that his good friend, the butler, did it.

³⁹ Cf. Fogal 2020, which combines reflectivism with LUMINOSITY. But see Worsnip 2018 for an alternative.

⁴⁰ Feldman 2006, pg. 223.

Watson rationally withholds belief, proposing to himself various alternative explanations of the evidence. But Sherlock admonishes Watson to stop rationalizing, telling him that rationality requires belief given Watson's own evidence. Watson accepts Sherlock's claim.

Here it is plausible that Watson should be motivated to believe the butler did it. But that much is compatible with deflationism. Most obviously, the deflationist could just say Watson gets the boring, deflationary kind of reason to believe that he got in Sherlock's Secret Evidence.

But reflectivists should not agree that this is the whole story. They should say that when Watson thinks *he* is required to believe, this has a special relevance to his deliberations. If he treats his apparent requirement as just one more bit of evidence, as he would the apparent requirements of other agents, then Watson is missing something. For here the person doing the deliberating is the very same person who is apparently required to resolve those deliberations by believing the butler did it. If Watson is rational, this will lead him to resolve those deliberations as he apparently is required to, by adopting the belief.

So while reflectivists and deflationists disagree about why Watson should believe that the butler did it, they might agree that he should. For a clearer test of who is right, we must consider examples where one's rational requirements are not adequate evidence as to the truth. I will propose some below, and a few more have been proposed by others.⁴¹ I think deflationism comes out looking good in these examples. But I agree there is another kind of example where this is less clear, namely:

Watson's Apparent Mistake: While investigating a murder together, Sherlock and Watson gather a large body of clues. Watson concludes via a long chain of deductive reasoning that the butler did it, and reports his conclusion to Sherlock without recounting the reasoning. Sherlock reminds Watson of his track record of committing fallacies in long chains of reasoning. Though Watson's reasoning was in fact flawless, he accepts that the objective chance of him reasoning flawlessly is low. So does a bystander, who was eavesdropping on their conversation.

There are three characters here: Sherlock, the bystander, and Watson. Who should believe what?

Start with Sherlock. His total evidence supports that the the butler did it.⁴² For the clues by themselves support this, and his additional evidence does not diminish this support. Of course Watson's conclusion will not provide further support for his guilt, given Watson's track record. But it will not provide evidence against it, either.

Next up is the bystander. Until the very end of our story, her evidence might support that the butler did it. For initially she knew Watson believed the butler did it, and she had no reason to doubt his reliability. If she had typical background evidence about people's general reliability, this might be enough to support that the butler did it, by ordinary evidential

⁴¹ Christensen 2016; Schoenfield 2015; Turri 2012; Worsnip 2018; and Horowitz's 2014 and Weatherson's 2019 discussions of Williamson 2011.

⁴² But see Weatherson 2019 for a view that rejects this, and some of what I say below.

standards. But after she hears about Watson's track record, her evidence will no longer support this. Without knowing the clues herself, we can suppose she has no other evidence supporting the butler's guilt, besides the fact that the apparently unreliable Watson thinks so.

What about Watson? This is where things go haywire. On the one hand, it seems that like the bystander, Watson should not believe the butler did it. On the other, it seems that like Sherlock, Watson still knows clues that support the butler's guilt. And if so, it can seem unavoidable that Watson's total evidence still supports that the butler did it, at least by ordinary standards.

Why is that unavoidable? One common observation is that if we suppose the clues deductively entail the butler's guilt, then Watson's total evidence will entail it.⁴³ But perhaps it could be denied that deductive entailment suffices for evidential support. This already would be a serious departure from an ordinary probabilistic conception of support. But there is an even deeper problem, stemming from the impersonality of ordinary evidential standards.

Notice Watson's third-personal evidence is the same as Sherlock's. Beyond the clues, both characters know that Watson inferred from these clues that the butler did it, and that Watson's track record is poor. Of course, Watson also knows that *he* is Watson, and thus that *his* track record is poor. But this will not make room for his total evidence to fail to support that the butler did it by ordinary standards, if these standards are impersonal. Suppose for instance that the best explanation of Sherlock's evidence is that Watson hit on the right answer by dumb luck. That also will be the best explanation of Watson's evidence, if abductive standards are invariant.

How should we respond? The *steadfastness strategy* says that Watson should simply believe the butler did it, like Sherlock should.⁴⁴ This response has the virtue of elegance, at least for deflationists who think higher-order evidence is just more evidence. And perhaps there is even something heroic about Watson resolutely sticking with his own best judgment.

But to many this response strains credibility. The problem is not just that deflationism advises Watson to assign too little weight to his track record evidence. And arguably, it is not even that it advises Watson to discount it entirely, and resolutely press on with his belief despite appreciating its relevance. Rather, the deflationist says Watson should treat the track record evidence the same way Sherlock should. But if Sherlock believed the butler did it based on the clues, he would rightly see Watson's track record as a non-sequitur having no bearing on the case. If Watson treats his own track record that way, that is not heroic steadfastness, but *obliviousness*. For it means failing to recognize that the track record speaks to the very reasoning that he himself just engaged in, and which convinced him the butler did it.

But deflationists are not automatically committed to steadfastness. They might instead adopt the *alienation strategy*.⁴⁵ It says Watson should suspend judgment like the bystander, and for similar reasons. Watson finds within himself the a strong conviction that the butler did it, and reasoning that strikes him as valid. But he should take a step back, and consider these

⁴³ E.g., Christensen 2010b, Whiting 2019, Schechter 2013, Smithies 2019.

⁴⁴ Cf. Kelly 2010 and Weatherson 2019.

⁴⁵ For discussion, see Christensen 2016 and White 2009.

psychological states from a neutral perspective like the bystander's. The fact that someone with his track record is in this state of mind is not much indication that the butler really did it. So if Watson's evidence just consists these psychological facts, then it does not really support that the butler did it, even by ordinary impersonal standards.

While I like the alienation strategy, I can see how it too might strain credibility.⁴⁶ Watson is not some idle bystander to the reasoning whose reliability has come into question. It is his own reasoning, proceeding from clues that he knows, to a conclusion which by that very reasoning he came to believe. The clues are thus Watson's basis for believing the butler did it, and he should revise the belief only if further higher-order evidence somehow defeats this basis, rather than some distinct psychological one. The problem is that it cannot do so by ordinary standards, if Watson's basis includes the same impersonal evidence available to Sherlock.

So deflationism faces a dilemma: obliviousness or alienation. And here the *reflection strategy* rides to the rescue.⁴⁷ It says that Watson's distinctively first-personal evidence makes all the difference. Unlike both Sherlock and the bystander, Watson's evidence includes the clues and also the first-personal evidence that *his* track record is poor. If this total evidence should have him reflectively believe (or at least suspect) that his belief in the butler's guilt is irrational, then maybe Watson should give it up. Thus the reflection strategy can potentially honor multiple, seemingly conflicting inclinations many of us have, by assigning non-evidential significance to first-personal reflective beliefs or evidence.

But I think the argument from Section 2 shows us that reflectivism has got to go, and with it the reflection strategy for handling higher-order evidence. In Sections 5-7, I defend the premises of this argument against different versions of reflectivism that might oppose them. But again, the upshot is not necessarily steadfastness about higher-order evidence. For deflationists can instead take the alienation strategy, and agree that there is something oblivious about steadfastness. The disagreement is about what that something is. Roughly speaking, the reflection strategy says the steadfast Watson's conception of himself is *too impersonal*. He fails to appreciate that it is he himself whose reliability or rationality is in doubt, and to respond differently than he would to third-personal doubts about someone else. The alienation strategy instead says Watson's thinking is *too first-personal*. He is stuck in his own head, unable to consider himself from a detached perspective like the bystander's. So in my book, these strategies are fundamentally opposed.

Not everybody sees it this way, however. Some authors are sympathetic to both strategies,⁴⁸ while others have opted for a middle ground approach where what matters are Watson's first-personal beliefs about his reliability, not about his rationality.⁴⁹ This approach takes a step away from conventional reflectivism, which stresses the normative character of

⁴⁶ Cf. Kelly 2010.

⁴⁷ E.g., Burge 2013, p. 82; Christensen 2010; DiPaolo 2018; and Littlejohn and Dutant 2021.

⁴⁸ Compare Christensen 2007, 197-7 with 2010a, or White 2009 with 2005, 2006, and 2013. And see Korsgaard's (1996, Sec. 3.2.1) classic characterization of reflection as involving normative judgments, but from an alienated perspective.

⁴⁹ Christensen 2016; Dogramaci 2015, Sec. 1.4; Weatherson 2019, Ch. 8; and perhaps Horowitz 2014. For Christensen, this is a change from earlier work.

reflective beliefs. But I think it is still not far enough. So long as first-personal beliefs about his reliability that motivate Watson to revise, this view still has Watson motivated by first-personal evidence in a way that is not licensed by ordinary evidential standards.

5. Transparency Reflectivism

Deflationists like me say reflection should not motivate changes in attitudes, except incidentally, by providing evidence. Take Watson, when Sherlock tells him in Sherlock's Admonition that he is required to believe the butler did it. Should this convince Watson the butler did it? The deflationist says it should only if his being required to believe it is sufficient evidence that the butler really did it. For only then would he be in a position to infer that the butler did it from the premise that he is required to believe it.

This is the idea behind EVIDENTIALISM, one of three premises in the argument for deflationism. It says roughly that evidence governs inference, or more specifically that premise *r* is relevant to *p* in inferential reasoning only insofar as *r* is ordinary evidence regarding *p*. It is time to examine this premise, and the kind of reflectivist view that denies it. Then in Sections 6 and 7 we will consider the other premises.

EVIDENTIALISM is plausible, but not trivial. It relates a kind of reasoning to a relationship between propositions. The kind of reasoning, *inferential reasoning*, paradigmatically includes inferences, by which one forms a belief in a conclusion based on a prior belief in a premise. If a witness tells Sherlock that she saw the butler fleeing the scene of the murder, perhaps Sherlock could infer that the butler did it. If so, he would reason in a way he might express by saying, "The witness says the butler fled the scene, so (probably) the butler did it." That is a paradigmatic example of inferential reasoning.

But conscious inferences are just one example of a broader phenomenon. You can be persuaded some conclusion is true based on premises or evidence, but without consciously rehearsing an argument. That counts as inferential reasoning too, as I intend the notion. Inferential reasoning also can yield doxastic attitudes other than belief. Suppose Sherlock goes on to learn the witness has a grudge against the butler. He might think "The witness has a motive to lie, so maybe it wasn't the butler after all." This reasoning might not count as an inference, since the attitude it produces is suspended judgment rather than belief. But it still has an obvious affinity with paradigmatic inferences. Borrowing a phrase from Pamela Hieronymi, it still involves treating the witness's grudge as *bearing on the question* whether the butler did it.⁵⁰ A premise can bear on a question more or less directly, and more or less decisively. And how you reason from the premise can follow suit.

Turn now to the relevant relationship between propositions, ordinary evidential support. It too can be understood inclusively, but not too inclusively. Familiar examples include deductive, inductive, and abductive support, but I cite these just for illustration. The important thing is that, just as a matter of terminology, 'ordinary standards' are impersonal. If I can inductively infer that I will be late for tomorrow's meeting from the premise that I have been late for previous meetings, then I can inductively infer that DB will be late for tomorrow's meeting from the premise DB has been late for previous meetings.

So EVIDENTIALISM says your inferential reasoning should follow the evidence, at least understood a certain way. You should not be persuaded that *p* by a premise unless the

⁵⁰ Hieronymi 2005.

premise is ordinary evidence supporting that p , the kind you might give as an answer on an informal logic exam.

Reflectivists could resist the argument for deflationism by denying this. They might say these impersonal standards are incomplete. Maybe we need a richer understanding of evidential support, or else a conception of inference that is not so closely tied to evidence. Either way, the inferences you rationally can draw are not bound by impersonal standards. When I learn that I am required to believe it will rain, that ought to persuade me that it will rain, in a way that is not reducible to ordinary induction, abduction, and the like. I do not need to analyze statistical evidence about how likely it is for me to be required to believe falsehoods, or anything like that. I can just directly conclude that it will rain, on the basis that I ought to.

A possible rationale for this view borrows a line from transparency accounts of self-knowledge.⁵¹ It says the question whether it will rain is not from my perspective independent of the question whether it is rational for me to believe it will rain, the way it is independent of questions about what is rational for other people.⁵² From my point of view, my being required to believe something bears more directly on its being the case. If I think that some character DB is required to believe it will rain, that should persuade me only if it is ordinary evidence—for example, if my background evidence supports that DB is unlikely to have misleading evidence. But if I think that I am the one who is required to believe it will rain, that should be enough to convince me, even if it is not sufficient evidence by the ordinary impersonal standards. Call this view, whether motivated this way or not, *transparency reflectivism*.

If any reflectivists are looking for advice, mine is that transparency reflectivism is the strongest version of their view. It fits best with common intuitions that Moorean beliefs are somehow incoherent, despite being logically consistent. And it arguably avoids making reflective beliefs out to offer the wrong kind of reason for belief.

But by allowing inferential reasoning to break from impersonal evidential standards, transparency reflectivism faces the problem of *chauvinism*.⁵³ If I am persuaded by my own rational requirements in a way that I am not persuaded by yours, I seem to be treating my requirements as a better indication of the truth. And that seems irrationally chauvinistic. Consider:

Mystery Detective's False Lemma: Watson is investigating a murder, and has compiled a long list of hypotheses, about each of which he rationally suspends judgment. In the morning, Sherlock tells Watson of a Mystery Detective, who has been investigating the same murder. This Mystery Detective allegedly believes a lemma that entails one of Watson's hypotheses. But while the lemma is well supported by the Mystery Detective's evidence, Sherlock says, the lemma is false. Later on in the afternoon, Sherlock tells

⁵¹ E.g., Byrne 2018, Fernandez 2013, and Moran 2001.

⁵² Cf. Burge 2013, Moran 2001, and Gibbons 2006 and 2013. And see Owens 2017, 1.2 for criticism.

⁵³ Transparency accounts of self-knowledge arguably face related problems. See Barnett 2016, Boyle 2011, and Valaris 2011.

Watson that he, Watson, is the Mystery Detective. And in the evening, Sherlock tells Watson that the hypothesis in question is that the butler did it.

I say Watson should suspend judgment on whether the butler did it, despite believing that he is required to believe it.

Surely suspending is what he should do in the morning, when he thinks the Mystery Detective is required to believe one or another of the hypotheses. Ordinarily an agent's being required to believe something is evidence supporting its truth, but in this case the Mystery Detective's belief is allegedly required only because of a false lemma. While the hypothesis might coincidentally be true, it might not.

And Watson should continue suspending on its truth in the afternoon, when he learns he is the Mystery Detective. It would be irrationally chauvinistic to say "Well, if *I* am the one who is required to believe the hypothesis, then it probably is true, whatever it is." That would amount to considering himself generally less likely than others to be misled.

The same goes in the evening, when Watson learns what the hypothesis in question is. He knew all along it was one of the ones on his list. If there is nothing special about this one in particular, he should not change his mind now. So on pain of chauvinism, Watson should not believe the butler did it, despite believing he is required to.

The basic idea behind the example is simple. If an agent's being required to believe something is not impersonal evidence it is true, then it should not be persuasive, even to the agent himself. This is consistent with the mundane observation that an agent's being required to believe something usually is evidence it is true. Transparency reflectivism goes beyond this mundane observation, and says you should sometimes be persuaded even without impersonal evidence. So it is bound to require reasoning chauvinistically sometimes.

So the problem for transparency reflectivism goes beyond any one counterexample.⁵⁴ Still, let me take a minute to defend this one against a few responses.

First, a reflectivist might deny that in the evening Watson should still reflectively believe that he is required to believe the butler did it. But why?

One answer assumes LUMINOSITY. It says Watson can know he is not really required to believe the butler did it, whatever Sherlock says. But even if so, that much goes for all cases of misleading higher-order evidence. And as with other cases, we still can ask how Watson should respond if he is misled by Sherlock's testimony, even if an idealized agent would not have been.

A better answer is that Watson's evening evidence includes not just the clues, but also Sherlock's testimony. Maybe Watson should think this new evidence is an opposing defeater for the lemma, and thus undermines the hypothesis that the butler did it. After all, Sherlock has told Watson that the lemma is false.

But I think this better answer still smuggles in unrealistic luminosity assumptions. Sherlock's testimony is hardly evidence opposing *all* propositions entailing the butler did it, or even all the ones Watson knows to entail this. For example, if Watson suspends judgment

⁵⁴ For others, see Christensen 2016; Schoenfield 2015; Turri 2012; Worsnip 2018; and Horowitz's 2014 and Weatherson's 2019 discussions of Williamson 2011.

on whether the butler did it with the revolver, Sherlock's testimony is not evidence that the butler did not do it with the revolver. Rather, Sherlock's testimony could give Watson evidence against a proposition entailing the butler did it only if Watson had introspective knowledge that the proposition is *the one that he believes* (or that his evidence supports). But this is evidence Watson cannot have without discovering the lemma—and which unbeknownst to him he never will have, since in fact there is no such lemma to discover.⁵⁵

A second response is that the example involves a 'positive' reflective belief that a belief is required. Perhaps transparency reflectivism could avoid it by sticking to 'negative' beliefs that a belief is prohibited, and endorse reasoning like: "I am required to doubt the butler did it, so maybe he didn't do it."

But chauvinism problems recur:

Mystery Detective's False Defeater: Watson is investigating a murder, and has compiled a long list of conclusions, each of which he rationally believes. In the morning, Sherlock tells Watson of a Mystery Detective, who has been investigating the same murder. This Mystery Detective allegedly believes an opposing defeater for one of Watson's conclusions. The defeater is well supported by the Mystery Detective's evidence, Sherlock says, but it is false. Later on in the afternoon, Sherlock tells Watson that he, Watson, is the Mystery Detective. And in the evening, Sherlock tells Watson that the conclusion in question is that the butler did it.

By evening, Watson should reflectively believe he is required to doubt the butler did it. But this should not persuade him that the butler might not have done it. In the morning, Watson should think that whichever conclusion the Mystery Detective is required to doubt is probably still true. And the same goes in the afternoon and evening. When Watson learns that he is the Mystery Detective, it would be chauvinistic to think "Since *I* am the one who is required not to believe the conclusion, maybe it is false, whatever it is." And he will have no further reason to change his mind when he learns what the conclusion happens to be.

6. Management Reflectivism

Transparency reflectivists deny that higher-order evidence is just more evidence. But they still think higher-order evidence is supposed to persuade. If you reflectively believe that you are required to believe *p*, this is supposed to serve as a premise allowing you to reason 'I am required to believe *p*, so *p*.' But maybe this is where transparency reflectivism went wrong. Could there be a different way for reflective beliefs to motivate, other than by persuasion? That is the idea behind *management reflectivism*.

The management reflectivist denies INFERENTIALISM, the premise that an apparent fact can rationally motivate you to believe only by serving as a premise in inferential reasoning. This is the same premise rejected by pragmatists, who say that when Cash comes to believe it will rain by pressing a button, the prize he was offered for believing can be his reason not

⁵⁵ Alex Worsnip presses the objection further: Even if Watson admits his current evidence supports the butler did it, he should think that inferring this conclusion from it would require discovering the lemma, and thus gaining new evidence undermining this support. And arguably this means Watson should not consider himself required to believe the butler did it. But I think this assumes (or assumes that Watson himself must assume) he cannot infer a conclusion from a lemma without knowing he believes the lemma.

just for pressing but also for believing. Like pragmatists, managements reflectivists deny that motivating beliefs need to provide premises. So when you reflectively believe you are required to believe it will rain, you do not need to *infer* that it will rain from this premise. Instead, it can just motivate you to see to it that you believe it will rain, in a way you might express by saying, “I am required to believe p, so I’ll go ahead and do it.” Put roughly, the management reflectivist wants you to adopt beliefs just because you you are apparently required to, not because your being required to persuades you they are true. Even more roughly, they think you should *manage* your beliefs by the standard of rationality.

Some reflectivists are probably best read as favoring management reflectivism.⁵⁶ And in any case, it is a natural fit for reflectivists impressed by the motivational role of normative beliefs. Even if beliefs about what I ought to do are not distinctively *persuasive*, they might still be distinctly *motivating*. Maybe if I am rational I will be motivated to do what I think I ought to, on pain of akrasia. When I think I ought to clear my mind of distractions, I will be motivated to clear my mind. And when I think I ought to believe my friend is innocent, I will be motivated to believe. These cases involve moral and prudential obligations, but something similar might go for rational obligations. If I think I am rationally required to believe, maybe that also should motivate me.

But are these sorts of motives good reasons for belief? Unlike moral or financial motives, reflective beliefs at least concern the epistemic status of the beliefs they are supposed to motivate. That does not automatically settle the question, however. Consider for comparison:

Telekinesis: Matilda knows a particular coin will soon be tossed, but she doesn’t know how it will land. Then she learns she has the power of telekinesis, and can affect the outcome just by forming a belief about it. If she believes it will land heads, that will cause it to land heads. And if tails, then tails.

When Matilda learns of her telekinesis, does that give her a reason to believe the coin will land heads? It is not evidence one way or the other. So Matilda should not infer that it will land heads from the premise that she has telekinesis. At the same time, it does mean any belief she forms will be true. That might give Matilda a reason to bring it about that she believes, if she can do so for example by pressing a button. But I think they still won’t be good reasons for believing the coin lands heads.

Maybe this could be disputed, or somehow written off as irrelevant to management reflectivism. Even so, two additional problems remain. The first concerns the psychological difficulty of believing for non-evidential reasons. Speaking for myself, I do not think I’d be able to believe the coin lands heads if I were in Matilda’s place, not without a belief-producing button.⁵⁷ I can press buttons at will, but I can’t believe at will, whether my motives are moral, financial, or epistemic. Maybe this is no problem if the paradigmatic reflective agent is, say, someone struggling vainly against the irresistible pull of irrational

⁵⁶ DiPaolo 2018 and White 2005, Sec. 6; 2006, pg. 539; 2014, 306-308; and maybe also Gibbons 2013, Schoenfield 2014, and Littlejohn and Dutant 2021. For criticism, see Owens 2017, Sec. 3.3 and Whiting 2019—though I think Whiting’s positive view still faces management reflectivism’s fetishism problem.

⁵⁷ Cf. Owens 2000, p. 30.

phobia or bias.⁵⁸ But if reflective belief revision is supposed to be something we can easily pull off, it cannot be too much like believing at will.

The second problem concerns the particular motives that management reflectivism demands. Matilda is after true belief, but the whole point of management reflectivism is to let you be motivated to believe something even when you are not persuaded of its truth. This raises a lot of tough questions about the normativity of rationality, and its connection to motivation. But I don't think they could shake out in a way that is congenial to management reflectivism. Maybe in some sense you ought to believe the rational thing, even when it is false. But it is in a subjective sense of 'ought'. It is not as if being rational is always objectively for the best. And so there is something *fetishistic* about being motivated to believe rationally, if it comes at the expense of truth.⁵⁹ Consider:

Belief Button: As in Mystery Detective's False Lemma, in the morning Sherlock tells Watson about a Mystery Detective who is required to believe a hypothesis because he rationally believes a false lemma. Also as before, Watson is told in the afternoon that he is the Mystery Detective, and in the evening that the hypothesis is that the butler did it. But this time, Watson has available all day a button that will cause the Mystery Detective to believe the hypothesis.

In this case or others like it, reflectivists want Watson to believe the butler did it, even if his being required to believe is not sufficient evidence by impersonal standards. We saw in Section 5 the trouble with saying Watson should infer this from the premise that he is required to believe it. But management reflectivists have another option. They can have Watson motivated to believe the butler did it, without being persuaded. The problem is, this means fetishizing rationality.

Consider Watson's situation in the morning. He thinks the Mystery Detective is rationally required to hold some belief or other. Maybe that is tantamount to thinking the Mystery Detective ought to hold it. But that does not mean Watson should be motivated to press the button. Given the circumstances, Watson should think there is a serious risk the belief would be false, even if rational. If Watson presses anyway, he would be fetishistically prioritizing the rationality of the Mystery Detective's beliefs over their truth. Maybe there is nothing stopping a rational agent from having fetishistic priorities like this. But it is hardly rationally mandatory.

Pressing is still fetishistic in the afternoon, when Watson learns he is the Mystery Detective. Just imagine him thinking, with a sigh of resignation, "As much as I hate risking a false belief, rationality requires that I hold this one, so I'd better do it." And it would be equally fetishistic in the evening. Unless he is persuaded the butler did it, pressing still would mean deliberately risking a false belief, just because it is rational. Maybe that is permissible, but again, it isn't required.

Belief Button involves the 'positive' reflective belief that Watson is required to believe the butler did it. But similar points go for 'negative' reflective beliefs, as in Mystery Detective's False Defeater. If Watson thinks the Mystery Detective is required to doubt

⁵⁸ E.g., Greco 2014 and White 2005, 456.

⁵⁹ Cf. Dogramaci 2017, pp. 64-65. The term is borrowed from Smith 1994.

some truth or other, it would be fetishistic to press a button causing him to do so. And it is no less fetishistic when Watson learns who the Mystery Detective is, and what the truth is.

7. Guidance Reflectivism

The preceding views say reflective beliefs motivate our first-order beliefs by providing reasons. *Guidance reflectivism* is different. It denies MOTIVATIONALISM, the premise that providing reasons is the only way for beliefs to rationally guide us. Denying this allows a different role for reflective beliefs. When I reflectively believe my reasons require believing p , that doesn't need to supply a premise supporting p or a motive for believing p . Instead, it can just facilitate my believing p based on some *other* reasons, most plausibly whatever reasons I take to require the belief in the first place.⁶⁰ In a slogan, reflective beliefs are reasoning-guiding, rather than reason-providing.

The general idea behind guidance reflectivism is popular, if controversial. On many theories of inference and the basing relation, inferring a conclusion from some premises involves a further *guiding belief* about the logical or evidential connection between them.⁶¹ Importantly, the role of this belief is not supposed to be supplying a further premise from which one infers the conclusion, on pain of vicious regress. If you need a further premise whenever you infer p from e , then the same would go for inferring p from e and the extra premise, and we are off to the races.⁶² Instead, the idea is that in addition to the motivating role of the premise beliefs, there is a distinct role for the guiding belief. It is to help you to believe based on the original premises, not to supply a new one.

But I think guiding beliefs are poorly suited to the jobs reflectivists need them for. Even if they have some role in inference and basing, it has little connection to phenomena like higher-order defeat, akrasia, and Moore-paradoxicality.

A first issue involves the content of guiding beliefs. What exactly do you need to believe in order to infer p from your evidence e ? Some proposals might let it be a fairly minimal belief that e supports p . Others might require greater specificity, like a belief that an agent should (or may) infer p from e via some specific rule r . But guidance reflectivists cannot be happy with any proposals along these lines. They need guiding beliefs to be something essentially first-personal, such as that *my evidence* supports p , or that *I* should infer p by rule r .

This distinction might not amount to much if it were impossible to be ignorant or mistaken about one's evidence. When one's evidence is e , one would then believe that one's evidence supports p just when one believes that e supports p . But even supporters of LUMINOSITY should accept it is possible to be mistaken about one's evidence, albeit irrationally. Suppose an agent knows both that e_1 supports that not- p , and that e_2 supports that p . While in fact her total evidence is e_2 , she thinks it is e_1 . What will she infer about p , if she responds rationally to these beliefs? Some fans of guiding beliefs might say she will infer p , since her evidence is e_2 , and she believes e_2 supports p . But this will result in epistemic akrasia, and in believing the Moore-paradoxical p , *but my evidence supports that not- p* . So a

⁶⁰ Cf. Whiting 2017, which opposes a hybrid of management reflectivism and guidance reflectivism. See also Owens 2017, pg. 25 for a related view.

⁶¹ See, e.g., Neta 2019. For criticism, see Boghossian 2008 and 2014, Broome 2019 and Siegel 2019.

⁶² Cf. Boghossian 2014, 7-8.

guidance reflectivist motivated by akrasia and Moore's paradox should have our agent believe not- p instead. But if the reflective belief is not supposed to supply a new premise, then on what basis is she supposed to believe? Our agent's reflective belief cannot guide her to believe that not- p based on e_1 , which she does not possess. And it cannot guide her to believe it based on e_2 , which in her view supports p .

This is not the only way for an agent's reflective beliefs to be inaccurate. Recall Sherlock's Admonition, where Watson reflectively believes on Sherlock's authority that the clues support the butler did it. Here Watson is misled not about what his clues are, but instead about what those clues support. Maybe Watson still should respond by believing the butler did it, but it is hard to see how he could be guided to believe this conclusion based on the clues themselves. Since the clues do not support the butler's guilt, that would require Watson to invent some fallacious line of reasoning to get from the clues to this conclusion. Surely even if he should believe the conclusion, that is not how he should go about it.

These problems both involve inaccurate reflective beliefs. But related problems arise for nonspecific reflective beliefs. Maybe if you stick to the old-fashioned method of introspection and *a priori* reasoning, you will always end up with reflective beliefs with enough specificity to provide useful guidance. But if we allow reflective beliefs to be based on *a posteriori* evidence, it seems inevitable that you might reflectively believe and even know that some belief is required, but lack the detailed knowledge necessary to believe it in the way that is required.

Consider for example long chains of reasoning, where a series of individual inferences brings one to a conclusion that cannot be inferred directly from the original premises. For example, Sherlock might believe the butler committed a murder via a long series of deductive steps from his clues, even though there is no single permissible step from the clues to that conclusion.

Some fans of guiding beliefs can tell a plausible story about such cases. They can just say that for each inferential step, there must be an appropriate guiding belief. Sherlock's *modus ponens* inferences might be guided by beliefs about *modus ponens*, for example, and his syllogistic inferences might be guided by beliefs about syllogisms. But this plausible story is not available to guidance reflectivists who are concerned with higher-order evidence, akrasia, and Moore-paradoxicality. Suppose Watson also knows the clues, and is informed by Sherlock (accurately, this time) that they support the butler's guilt. The guidance reflectivist had better say this belief can guide Watson to believe the butler did it, or else he will akratically fail to believe something that he knows his evidence supports. But how will it guide him? Just knowing that there is some series of steps available might not enable Watson to figure out what they are, anymore than knowing there is a proof of Fermat's Last Theorem from arithmetic axioms enables me to construct it myself.

Maybe guidance reflectivists could respond by distinguishing positive and negative reflective beliefs. If you 'positively' believe that some attitude is required by your evidence, that might not be enough to guide you to a well-founded belief. But if you 'negatively' believe that your evidence prohibits the attitude, that still might prevent well-founded belief. This concession might still leave enough resources for some higher-order undermining cases, like Watson's Apparent Mistake. But it struggles with others, like:

Watson and Lestrade's Disagreement: Watson is informed by Sherlock that his clues either strongly support the butler's guilt or strongly support his

innocence. In fact it is guilt, and after thinking it over Watson rationally believes both that the butler is guilty and that his own clues support guilt. He then encounters Lestrade, who takes the same clues to support the butler's innocence. Watson considers Lestrade his peer, and suspends judgment on what the clues support.

A steadfast about higher-order evidence might say Watson should still believe the butler did it. But everyone else should say he ought to suspend. Can guidance reflectivists? Watson not only lacks a belief that the clues support suspension, but knows that they do not. So guidance reflectivists cannot allow his suspended judgment to be well-founded.

8. Reflection, Alienation, and Self-Awareness

We began with a caricature of the excessively reflective agent as a narcissist, preoccupied with her own beliefs rather than the world around her. But deflationism does not just caution against excessive reflection. It says not to let reflective beliefs guide us at all, not as such. Does following this advice also open one up to caricature?

Perhaps the unreflective agent will seem lacking in a kind of self-awareness. If I treat the apparent facts about myself as ordinary, impersonal evidence, maybe that betrays an obliviousness to myself and my place in the world. There are many people out there whose track record or biases I might learn about. But only one of them is me, the same person whose reasoning has just now led me (say) from meteorological premises to the belief it will rain. Shouldn't it matter what I learn about this person in particular? If what I learn is that I have a lousy track record, then it seems oblivious to press on in my belief that it will rain, as if it were just anybody whose track record is lousy. Yet singling myself out for special consideration is not permitted by ordinary, impersonal evidential standards. I can be logically consistent and probabilistically coherent, infer the best explanations of my evidence, and so on—all without information about the reasoning I am performing informing that reasoning in any distinctive way.

This threat of obliviousness, as we saw in Section 4, is a central motivation for rejecting deflationism in favor of reflectivism. But as we also saw, the deflationist has a way out. The apparent need for special first-personal reflection on my beliefs and reasoning arises because of another special relationship I have to them, namely being their subject. There is no such threat facing a detached bystander, who believes it will rain based not on meteorological evidence, but the psychological evidence that DB believes it will rain. If the bystander knows my track record is lousy, this psychological evidence will not support that it will rain, by ordinary, impersonal standards. If things are different for me, it must be because my own relationship to my beliefs and reasoning is somehow different from a bystander's. To avoid obliviousness, then, maybe I do not really need to add some special consideration for evidence about myself, but rather to subtract this special relationship.

Now to some this might seem to involve an objectionable form of alienation. It gets my relation to my own mental life all wrong, it might be complained, making it out like I am merely aware of my own beliefs and reasoning as objects, when in fact I am their subject. I have a lot of sympathy for this complaint. But I also think we need to somehow come to terms with alienation, if not as a universal predicament, then at least a local and circumstantial one. It may be that in calling into question a given belief or train of reasoning, I need to take a step back, and consider it from an alienated perspective. Maybe this even

means that for the deflationist, self-awareness is only achievable in a piecemeal fashion, since in stepping back from one perspective, I inevitably adopt another one that I am not stepping back from. But that is preferable to the reflectivist's picture of self-awareness as involving a privileging of first-personal evidence. Any kind of self-awareness worth wanting must be compatible with awareness of oneself as just one person among others, whose epistemic situation is no more germane to the truth than anyone else's.⁶³

⁶³ This paper was formerly titled 'Higher-Order Evidence is the Wrong Kind of Reason', and before that 'The Agony of Higher-Order Defeat'. I am especially indebted to Paul Boghossian and Jim Pryor, and I have also benefitted, in many cases dramatically, from comments and discussions with David Christensen, Justin Clarke-Doane, Brendan De Kenessey, Sinan Dogramaci, Hartry Field, Nick Hughes, David Hunter, Zoe Johnson-King, Hilary Kornblith, John Morrison, Jennifer Nagel, Ram Neta, Elliot Paul, Zain Rasa, Gurpreet Rattan, Josh Schechter, Stephen Schiffer, Miriam Schoenfield, Declan Smithies, Sergio Tenenbaum, Brian Weatherson, Mason Westfall, Roger White, Alex Worsnip, and audiences at Toronto Metropolitan University, McGill University, New York University, Columbia University, Queen's University, and the University of Toronto.

Works Cited

- Alston, William (1989) *Epistemic Justification*. Ithaca: Cornell University Press.
- Arpaly, Nomy (2002) *Unprincipled Virtue: An Inquiry Into Moral Agency*. OUP.
- Barnett, David James (2016) 'Inferential Justification and the Transparency of Belief' *Noûs* 50(1): 184-212.
- Bergmann, Michael (2005) 'Defeaters and Higher-Level Requirements' *Philosophical Quarterly* 55(220): 419-436.
- Boghossian, Paul (2008) 'Epistemic Rules' *Journal of Philosophy* 105(9): 472-500.
- (2014) 'What is Inference?' *Philosophical Studies* 169(1): 1-18.
- BonJour, Laurence (1985) *The Structure of Empirical Knowledge*. Harvard University Press.
- Boyle, Matthew (2011) 'Transparent Self-Knowledge' *Aristotelian Society Supplementary Volume* 111(2pt2): 223-241.
- Briggs, Ray (2009) 'Distorted Reflection' *Philosophical Review* 118(1): 58-85.
- Broome, John (2013) *Rationality through Reasoning*. Wiley-Blackwell.
- (2019) 'A Linking Belief is Not Essential for Reasoning' in Brendan Balcerak Jackson and Magdalena Balcerak Jackson (eds.) *Reasoning: New Essays on Theoretical and Practical Thinking*, OUP.
- Burge, Tyler (2013) *Cognition Through Understanding*. OUP.
- Byrne, Alex (2018) *Transparency and Self-Knowledge*. OUP.
- Carter, Sam and Simon Goldstein (2021) 'The Normality of Error' *Philosophical Studies* 178: 2509-2533.
- Christensen, David (2007) 'Epistemology of Disagreement: The Good News' *Philosophical Review* 116(2): 187-217.
- (2010a) 'Rational Reflection' *Philosophical Perspectives* 24(1): 121-140.
- (2010b) 'Higher-Order Evidence' *Philosophy and Phenomenological Research* 81(1): 185-215.
- (2016) 'Disagreement, Drugs, etc.: from Accuracy to Akrasia' *Episteme* 13(4): 397-422.
- DiPaolo, Joshua (2018) 'Higher-Order Defeat is Object-Independent' *Pacific Philosophical Quarterly* 99(2): 248-269.
- Dogramaci, Sinan (2015) 'Communist Conventions for Deductive Reasoning' *Noûs* 49(4): 776-799.

- (2017) ‘Why Is a Valid Inference a Good Inference?’ *Philosophy and Phenomenological Research* 94(1): 61-96.
- Dorst, Kevin (2019) ‘Higher-Order Uncertainty’ in Mattias Skipper and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, OUP.
- Egan, Andy and Adam Elga (2005) ‘I Can’t Believe I’m Stupid’ *Philosophical Perspectives* 19(1): 77-93.
- Elga, Adam (2005) ‘On Overrating Oneself... And Knowing It’ *Philosophical Studies* 123(1-2): 115-124.
- Feldman, Richard (2005) ‘Respecting the Evidence’ *Philosophical Perspectives* 19: 95-119.
- (2006) ‘Epistemological Puzzles of Disagreement’ in Stephen Hetherington (ed.), *Epistemology Futures*. OUP.
- Fernandez, Jordi (2013) *Transparent Minds: A Study of Self-Knowledge*. OUP.
- Fogal, Daniel (2020) ‘Rational Requirements and the Primacy of Pressure’ *Mind* 129(516): 1033-1070.
- Gibbons, John (2006) ‘Access Externalism’ *Mind* 115(457): 19-39.
- (2013) *The Norm of Belief*. OUP.
- Goldman, Alvin (1999) ‘Internalism Exposed’ *Journal of Philosophy* 96(6): 271-293.
- Greco, Daniel (2014) ‘A Puzzle About Epistemic Akrasia’ *Philosophical Studies* 167(2): 201-219
- Hieronymi, Pamela (2005) ‘The Wrong Kind of Reason’ *Journal of Philosophy* 102: 437-457.
- (2006) ‘Controlling Attitudes’ *Pacific Philosophical Quarterly* 87(1): 45-74.
- Horowitz, Sophie (2014) ‘Epistemic Akrasia’ *Noûs* 48(4): 718-744.
- Kelly, Thomas (2002) ‘The Rationality of Belief and Other Propositional Attitudes’ *Philosophical Studies* 110(2): 163-196.
- (2010) ‘Peer Disagreement and Higher-Order Evidence’, in *Social Epistemology: Essential Readings*, ed. Alving Goldman and Denis Whitcomb, OUP.
- Kornblith, Hilary (2012) *On Reflection*. OUP.
- Korsgaard, Christine (1996) *The Sources of Normativity*. CUP.
- Lasonen-Aarnio, Maria (2014) ‘Higher-Order Evidence and the Limits of Defeat’ *Philosophy and Phenomenological Research* 88(2): 314-345.
- (2020) ‘Enkrasia or Evidentialism? Learning to Love Mismatch’ *Philosophical Studies* 177(3): 597-632.

- Littlejohn, Clayton (2012) *Justification and the Truth-Connection*. Cambridge University Press.
- Littlejohn, Clayton and Julien Dutant (2021) 'Defeaters as Indicators of Ignorance' in Mona Simion and Jessica Brown (eds.), *Reasons, Justification, and Defeat*. OUP.
- McGinn, Colin (1982) *The Character of Mind*. OUP.
- Moran, Richard (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press.
- Neta, Ram (2019) 'The Basing Relation' *Philosophical Review* 128(2): 179-217.
- Owens, David (2000) *Reason without Freedom*. Routledge: London and New York.
- (2017) *Normativity and Control*. OUP.
- Pryor, James (2005) 'There is Immediate Justification' in *Contemporary Debates in Epistemology*, ed. Matthias Steup and Ernest Sosa. Blackwell.
- Rinard, Susanna (2017) 'No Exception for Belief' *Philosophy and Phenomenological Research* 94(1): 121-143.
- (2019) 'Equal Treatment for Belief' *Philosophical Studies* 176(7): 1923-1950.
- Rioux, Catherine (forthcoming) 'On the Epistemic Costs of Friendship: Against the Encroachment View' *Episteme*.
- Rosati, Connie (2016) 'Moral Motivation' *Stanford Encyclopedia of Philosophy*, ed. Edward Zalta. <<https://plato.stanford.edu/archives/win2016/entries/moral-motivation/>>
- Scanlon, Thomas (2007) 'Structural Irrationality' in *Common Minds: Themes from the Philosophy of Philip Pettit*, ed. Geoffrey Brennan, Robert Goodin, Frank Jackson, and Michael Smith. OUP.
- Schechter, Joshua (2013) 'Rational Self-Doubt and the Failure of Closure' *Philosophical Studies* 163(2): 428-452.
- Schoenfield, Miriam (2014) 'Permission to Belief: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief' *Noûs* 48(2): 193-218.
- (2015) 'A Dilemma for Calibrationism' *Philosophy and Phenomenological Research* 91(2): 425-455.
- Shah, Nishi (2006) 'A New Argument for Evidentialism' *Philosophical Quarterly* 56(225): 481-498.
- Shah, Nishi and J. David Velleman (2005) 'Doxastic Deliberation' *Philosophical Review* 114(4): 497-534.
- Shoemaker, Sydney (1996) *The First-Person Perspective and Other Essays*. Cambridge University Press.

- Siegel, Susanna (2019) 'Inference Without Reckoning' in Brendan Balcerak Jackson and Magdalena Balcerak Jackson (eds.) *Reasoning: New Essays on Theoretical and Practical Thinking*, OUP.
- Smith, Michael (1994) *The Moral Problem*. Blackwell.
- Smithies, Declan (2019) *The Epistemic Role of Consciousness*. OUP.
- Stroud, Sarah (2006) 'Epistemic Partiality in Friendship' *Ethics* 116(3): 498-524.
- Titelbaum, Michael (2015) 'Rationality's Fixed Point (Or: In Defense of Right Reason)' *Oxford Studies in Epistemology* 253-294.
- (2019) 'Return to Reason' in Mattias Skipper and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, OUP.
- Turri, John (2012) 'A Puzzle about Withholding' *Philosophical Quarterly* 62(247): 355-364.
- Valaris, Markos (2011) 'Transparency as Inference: Reply to Alex Byrne' *Proceedings of the Aristotelian Society* 111(2pt2): 319-324.
- Way, Jonathan (2016) 'Two Arguments for Evidentialism' *Philosophical Quarterly* 66(265): 805-818.
- Weatherson, Brian (2019) *Normative Externalism*. OUP.
- Wedgwood, Ralph (2002) 'The Aim of Belief' *Philosophical Perspectives* 36(s16): 267-297.
- White, Roger (2005) 'Epistemic Permissiveness' *Philosophical Perspectives* 19(1): 445-459.
- (2006) 'Problems for Dogmatism' *Philosophical Studies* 131(3): 525-557.
- (2009) 'On Treating Oneself and Others as Thermometers' *Episteme* 6(3): 233-250.
- (2013) 'What is My Evidence that Here is a Hand?' in *Scepticism and Perceptual Justification*, ed. Dylan Dodd and Elia Zardini, OUP.
- Whiting, Daniel (2017) 'Against Second-Order Reasons' *Noûs* 51(2): 398-420.
- (2019) 'Wither Higher-Order Evidence?' in Mattias Skipper and Asbjørn Steglich-Petersen (eds.), *Higher-Order Evidence: New Essays*, OUP.
- Williamson, Timothy (2011) 'Improbable Knowing' in *Evidentialism and its Discontents*, ed. T. Dougherty. OUP.
- Worsnip, Alex (2018) 'The Conflict of Evidence and Coherence' *Philosophy and Phenomenological Research* 96(1): 3-44.