

# Reflection Deflated<sup>1</sup>

## 1. Introduction

Philosophers are narcissists. Or at least, we have tendencies towards narcissism in a special epistemic sense: 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 many philosophers take our idiosyncratic tendencies to reflect a rational ideal; they build a preoccupation with one's rationality into what it is to be rational.

Of course, advocates of epistemic narcissism do not put it that way. Instead, 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.<sup>2</sup>

Some authors take rationality to be impossible without reflection.<sup>3</sup> But most accept that some degree of rationality is attainable even if your attention remains directed at the world around you. If you know that it will rain if the barometer reads 'low', and can see that the barometer does read '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, which we share with children and animals.<sup>4</sup>

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

---

<sup>1</sup> I keep changing the title of this paper, and might do it again. It was formerly 'Higher-Order Evidence is the Wrong Kind of Reason', and before that 'The Agony if Higher-Order Defeat'. The central ideas grew out of dissertation research under Paul Boghossian and Jim Pryor, and the paper is hugely indebted to them. I have also benefitted 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, Elliot Paul, Zain Rasa, Gurpreet Rattan, Josh Schechter, Stephen Schiffer, Miriam Schoenfield, Declan Smithies, Sergio Tenenbaum, Brian Weatherston, Roger White, Alex Worsnip, and audiences at Ryerson University, McGill University, NYU, Queen's University, and the University of Toronto.

<sup>2</sup> Smithies 2019.

<sup>3</sup> E.g., Shoemaker 1996, pp. 32-33

<sup>4</sup> E.g., Alston 1989, pp. 221-236; Burge 2013; McGinn 1982, pg. 20; and Smithies 2019.

[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.<sup>5</sup>

These authors make ambitious claims about the significance of reflection. While common, they are controversial.<sup>6</sup> 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.<sup>7</sup> This modest claim often turns up in discussion with no broader reflectivist agenda, as a well-known truism cited in support of some other point. 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.<sup>8</sup>

And here is Ralph Wedgwood, while 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.<sup>9</sup>

Debates between internalists and externalists are rife with similar claims.<sup>10</sup> And it's not just on the internalist side. Knowledge-first externalists think the norm for belief is knowledge, rather than some internal requirement of rationality. But many still assign importance to an agent's views on whether her beliefs meet this externalist norm. For example, they might explain the rationality or blamelessness of a deceived agent by her inability to tell that she does not know the things she believes.<sup>11</sup> The idea, presumably, is that if she could tell she does not know, that would give her a clearer reason not to believe.

But moderate reflectivism has probably received the most sustained recent attention in connection with peer disagreement, epistemic akrasia, and higher-order evidence.<sup>12</sup> Just

---

<sup>5</sup> Burge 2013, pg 372. (??)

<sup>6</sup> For opposition, see Arpaly 2002, Broome 2013, and Kornblith 2012.

<sup>7</sup> E.g., Elga 2005, pg. 116; Scanlon 2007; and the authors quoted below.

<sup>8</sup> Moran 2001, pg. 55.

<sup>9</sup> Wedgwood 2002, pg. 271.

<sup>10</sup> Boghossian 2008, pg. 472; BonJour 1985, pp. 7-8; Dogramaci 2015; Smithies, and White 2006, pg. 539.

<sup>11</sup> Dutant and Littlejohn, Williamson, etc.

<sup>12</sup> E.g., Christensen 2010a, 2010b, 2016; Elga 2005; Greco 2014; Horowitz 2014; Kelly 2010; Lasonen-Aarnio 2014 and 2020; Smithies 2019, Chs. 9-10; Weatherson 2019, Worsnip 2018

compare with traditional reflectivism 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*.<sup>13</sup>

While all these authors put the point in different ways, they share a common idea. It is that one's reflective beliefs about an attitude can affect the rationality or justification of that attitude, or commit one to adopting or revising it, or give one a reason to do so. As I will put it, they think reflective beliefs should *guide* you.

This claim is all I mean by *reflectivism*. It is not confined to ambitious reflectivists like Burge and Smithies, or even moderate ones like Christensen. 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*.<sup>14</sup>

In contrast, I think the whole idea of reflective reasoning is a mistake. If I am wondering whether it will rain, I had better not sit around navel gazing. Instead I 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 I am also part of the world, and there is no barrier to my considering what is going on with me and my beliefs. But if I do, these will just be further considerations to be weighed alongside the others. Just because there is one measly person who seems required to believe it will rain, that hardly settles the matter.<sup>15</sup> And it won't settle the matter any more conclusively if the measly person is me.

Under this *deflationist* view, reflective beliefs only give you reasons for belief by providing ordinary evidence. For example, the fact that my evidence still supports that ribosomes synthesize proteins might itself support that Mrs. Peale was an effective biology teacher. And so a reflective belief about my evidence can give me reason for a belief about Mrs. Peale. If the fact that my evidence supports this also is evidence that ribosomes synthesize proteins, then my reflective belief can give a reason for this belief in the same way.

---

<sup>13</sup> Christensen 2010a. As I discuss below, Christensen has revised his view. But even the new view partly resembles reflectivism, and anyway these sentiments remain common (e.g., Elga 2005, Greco 2014, Horowitz 2014, Worsnip 2018).

<sup>14</sup> Lasonen-Aarnio 2020. Other critics who I think assume modest reflectivism include Arpaly 2002, Bergman 2005, Broome 2013, and Dorst 2019.

<sup>15</sup> Cf. Kelly

Now the suggestion that higher-order evidence is just more evidence is hardly new.<sup>16</sup> But I think its implications are more radical than usually appreciated. For it deflates the significance of reflective beliefs and higher-order evidence. It's not just that reflective beliefs or evidence are not always conclusive. It's that they cannot ever guide us the way reflectivists think. If I take myself to be 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. If it matters at all, it is only incidentally, by providing evidence about the weather.

## 2. The Argument for Deflationism

Reflectivists want our reflective beliefs to guide us. But this rough idea leaves a lot open. It might not mean 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, what I really should do is give up this reflective belief, not follow through.<sup>17</sup> Indeed, this is why some reflectivists pitch their view in terms of higher-order evidence or justification rather than belief.<sup>18</sup> 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, can guide our beliefs otherwise than by providing evidence.

We will see some motivations for this idea later, but here is one for starters: Rationality is normative, and reflective beliefs are first-personal. So reflective beliefs in effect concern what one *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 gives me an especially direct reason for making sure they do.

Reflectivism does not come so cheap, however. Morality is normative, too, but that doesn't mean I should believe what I think is morally required. Suppose my friend is accused of a crime, and I think I am morally required to believe she is innocent. Does this give me a reason to believe she is innocent?<sup>19</sup> Maybe in some sense it does, but it still strikes us as somehow 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 might seem obvious that they are. Maybe you can be morally required to believe something that you have no evidence for. But rational requirements and evidence go hand in hand. If you think you are rationally required to believe something, that is at least pretty close to taking your evidence or epistemic reasons to support it. Now maybe believing something on the basis that your evidence supports it is not quite the same as believing based on the evidence itself. But if evidence supporting a proposition is the right kind of reason to believe it, it might

---

<sup>16</sup> See especially Kelly, though he seems to mean that higher-order evidence is not automatically decisive, and must be balanced against first-order evidence.

<sup>17</sup> Cf. Arpaly and Williams

<sup>18</sup> Christensen, Smithies, etc.

<sup>19</sup> Cf. Stroud 2006 and Rioux MS.

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 conditional on his 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). Look at it this way. 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, irrelevant to its rationality. For his knowledge of the prize to make the belief rational, it needs to in some further way *guide* him. In what way? A natural answer is that it must be a motivating reason, a reason for which Cash believes it will rain.

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. They are not charitably interpreted as saying Cash ought to infer it will rain from the premise that he will get the prize. For comparison, 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 act so as to produce the belief, 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. 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.”

Surely there is nothing 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 a crucial part of the etiology of the belief. But that does not make it the reason for which Cash believes.<sup>20</sup>

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*.

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.<sup>21</sup>

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 guide his belief that it will rain without being evidence. While some of the details are my own, I hope the main contours are familiar. My goal isn’t to give a new argument against pragmatism, but to apply evidentialist commitments consistently to reflective beliefs.<sup>22</sup>

---

<sup>20</sup> Cf. Hieronymi 2006, Kelly 2002, Shah 2006, Rinard 2019, and Way 2016.

<sup>21</sup> See especially Jim Pryor’s (2005) discussion of the ‘Premise Principle’. I am indebted also to extensive discussion with him.

<sup>22</sup> See also Whiting 2017 for some closely related thoughts.

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 (or prohibited from) believing *p* can rationally guide one's belief about *p* only insofar as the apparent fact that one is required to (or prohibited from) believing *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 guy here ought to believe about *p* have no special significance. If they are relevant to my deliberation, it is in the same way as apparent facts about what other people ought to believe, or apparent facts about shoes and ships and sealing wax. They are all just apparent facts that might speak for or against the truth of *p* according to the same general evidential standards.

In contrast, I will claim reflectivists need reflective beliefs to provide reasons or guidance that goes beyond their ordinary evidential import. This is for two reasons. The first, which won't be my focus, is that many reflectivists think the *normativity* of rationality is central to the epistemic significance of reflective beliefs. I am supposed to have a special reason or motive to believe *p* when I think I *rationaly ought* to do so. In contrast, ordinary evidential standards 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 not because normative premises amount to especially strong evidence.

The second reason, which I will focus on, is that reflectivists need *first-personal* reflective beliefs to have a significance that does not extend to one's beliefs about other people. In contrast, ordinary evidential standards, like those governing deductive, inductive, and abductive inferences, do not carve out a distinctive role for first-personal premises or evidence. As I will put it, these and other familiar candidate standards of evidential support are invariant under *third-personalization*.<sup>23</sup>

To third-personalize an agent's beliefs and evidence, we replace any allegedly *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 a third-personalization of my beliefs and evidence includes 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 in third-personal terms. Or instead, you might suppose there is no such thing as *de se* thought,

---

<sup>23</sup> Some might suggest Sleeping Beauty and related cases favor evidential standards that are not invariant under third-personalization (e.g., Elga 2000). While I am skeptical, I won't pick a fight. My argument in Section 5 will focus on the particular standards for inferential reasoning required to vindicate reflectivism, rather than this general claim about third-personalization.

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 I claim ordinary evidential standards are invariant under third-personalization. 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 third-personalized, 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 third-personalized will entail that DB did so. 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 third-personalized, I can abductively infer the same thing if I learn the butler lied to DB about his whereabouts. Evidential standards are invariant when they are like that.

In Sections 3 and 4, I will argue that reflectivists need first-personal reflective beliefs to have a significance that does not survive third-personalization, and so is not afforded them by ordinary evidential standards. So reflectivists need to resist the argument for deflationism. But how? In Section 5, I discuss what I call *transparency reflectivism*, a version of reflectivism that denies EVIDENTIALISM. In Section 6, I turn to *management reflectivism*, which rejects INFERENCEALISM. And in Section 7, I consider *guidance reflectivism*, which instead denies MOTIVATIONALISM.

### 3. Motivations for Reflectivism

I have found that 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 most discussions I read about 'reflective beliefs', 'higher-order evidence', 'metajustification', and the like. 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 in epistemology. Today's epistemologists might be less intent on reforming your beliefs, but some still haven't given up on it. Yet under deflationism, there is no room for theoretical knowledge about epistemology to guide us in our other beliefs. Epistemological inquiry might help us figure out what to believe *about epistemology*, but not much else.

All that said, a contrary impulse towards deflationism also seems to me common, including among some of the authors I have cited as reflectivists.<sup>24</sup> It may be that on further thought, many of them would prefer deflationism after all. But I do think the inflationist 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.<sup>25</sup> If the brutes can make ordinary deductive and inductive inferences about cabbages and kings, why not their own rational requirements? Maybe a creature with general inferential capacities could face a peculiar mental block whenever premises about its own rational requirements come along. But that seems an unlikely candidate for a deep difference between kinds of rational beings.

Now moderate reflectivists do not share this broader agenda, and might be driven by 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.<sup>26</sup> Sometimes this idea is reinforced by an analogy with practical akrasia.<sup>27</sup> 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.<sup>28</sup>

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. Second, premises about what I ought to do are not stronger evidence than substitute third-personal premises about what DB ought to do. But when it comes to motivation, the first-personal character of such judgments 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.

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 *p* instead.<sup>29</sup> They accept something like:

---

<sup>24</sup> E.g., Christensen 2016; Lasonen-Aarnio 2014, pp. 317-318; and Smithies in correspondence

<sup>25</sup> E.g., Alsoton 1989, Burge 2013, and Smithies 2019.

<sup>26</sup> Wedgwood 2002.

<sup>27</sup> E.g., Greco 2014; Horowitz 2014; Smithies 2019, Ch. 9; Titelbaum 2015 and 2019, pg. 228.

<sup>28</sup> For review, see Rosati 2016, Sec. 3.

<sup>29</sup> E.g., Smithies 2019, pg. 314 and Worsnip 2018.

(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 it will rain.*<sup>30</sup>

Reflectivism offers a natural explanation of these intuitions. Suppose I believed the second Moorean conjunct, which says I am required to doubt 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 not to.

It is hard to square this with deflationism, however. I can rationally believe third-personal Moorean conjunctions like *It will rain, but DB is rationally required to doubt it will rain.* Likewise, I might think some character DB is required to doubt it will rain without having a reason to doubt it myself. Maybe things are different if I think I am the one required to doubt it, 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 for you to believe that not-*p*. If so, then LUMINOSITY will guarantee the truth of REFLECTION REQUIREMENT in an entirely "bottom up" fashion. Unless your first-order attitude is antecedently required, you will be guilty of irrationality by believing it to be required. So rationality will never permit you both to believe an attitude is required and not to hold it, just because your higher-order beliefs should track their first-order subject matter.

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 to be required, the reflectivist says that ought to motivate or otherwise guide you to adopting it. It is that claim, not REFLECTION REQUIREMENT alone, that I say is incompatible with deflationism.<sup>31</sup>

Even so, deflationism still undermines one motivation for LUMINOSITY. It might be claimed that you *had better be* able to know what attitudes are required, or else you will be in a bad situation.<sup>32</sup> 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 necessarily undermine its rationality. And realizing a belief is irrational won't help you get rid of it.

---

<sup>30</sup> E.g., Greco 2014; Horowitz 2014, pg. 725; and Smithies 2019.

<sup>31</sup> For some reflectivism-adjacent views that I think just combine luminosity assumptions with further claims the deflationist can accept (e.g., Briggs 2009, Egan and Elga 2005, Titelbaum 2015).

<sup>32</sup> E.g., Burge 2013.

There is one more motivation for reflectivism. It comes from reactions to particular examples, especially ones involving higher-order evidence. Here I think things are a mixed bag. Some common intuitions taken to favor reflectivism are probably consistent with deflationism, while others might not be. I turn to them next.

#### 4. Higher-Order Evidence and Deflationism

Along with cash prizes, moral reasons are paradigmatically the wrong kind of reason for belief. 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 reasons can be the right kind of reason 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. I could have inferred the same thing from the evidence that somebody else ought to believe my friend is innocent.

The same goes for reflective beliefs and higher-order evidence. Like anything else, apparent facts about what you are rationally required to believe can be evidence. And to that extent, they can be good reasons for belief. This is important because rational requirements to believe routinely have significant evidential weight. In a familiar slogan, evidence of evidence is evidence. But we must be careful to distinguish this boring truth with 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 people can be required to believe falsehoods when their evidence is misleading, 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, or something else of that kind.

Compare this to:

**Sherlock's Admonition:** While investigating a murder, Watson is confronted with inconclusive evidence that his good friend, the butler, did it. 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. The deflationist can just say Watson gets the boring, deflationary kind of reason to believe that he got in Sherlock's Secret Evidence.

But reflectivists might not agree that this is the whole story. They might think 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 might disagree about why Watson should believe that the butler did it, they might agree that he should. There are other cases, however, where it is less clear they can agree on even that much, such as:

**Watson's Apparent Mistake:** Sherlock and Watson gather a large body of clues regarding a murder. Watson concludes that the butler did it via a long chain of deductive reasoning, 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. The clues by themselves support this. And beyond the clues, Sherlock knows that Watson inferred from these clues that the butler did it, and that Watson's track record is poor. But that is no evidence against the butler having done it. A tendency to commit fallacies will make Watson less reliable at arriving at the truth, but it won't make him positively tend towards falsehoods.

Next up is the bystander. Until the very end of our story, her evidence might support that the butler did it. For she knew Watson believed the butler did it, and she had no reason to doubt his reliability. If she has typical background evidence about people's general reliability, this might be enough to support that the butler did it, by ordinary inductive or abductive standards.<sup>33</sup> But after she hears about Watson's track record, her evidence will not support that the butler did it. At that point, she should doubt that Watson's beliefs are a reliable indicator of the truth. And without knowing the clues herself, we can suppose she has no other evidence supporting the butler's guilt.

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 seems 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 still will entail it.<sup>34</sup> 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 invariance of ordinary evidential standards under third-personalization. Watson's third-personal evidence is the same as

---

<sup>33</sup> [something on reductionism about testimony?]

<sup>34</sup> Smithies, Christensen, etc.

Sherlock's. Beyond the clues themselves, 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 invariant under third-personalization. Suppose for instance that the simplest explanation of Sherlock's third-personal evidence is that Watson hit on the right answer by dumb luck. That also will be the simplest 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. This response certainly has the virtue of elegance, at least for deflationists who think higher-order evidence is just more evidence. But to others, it strains credibility.

But there is another deflationist strategy, the *alienation strategy*. It says Watson should suspend judgment like the bystander, and for similar reasons.<sup>35</sup> Watson finds within himself with the belief that the butler did it, and reasoning that struck him as sound. But he should take a step back, and consider these 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 third-personal standards.

But this strategy also might strain credibility. Watson is not some idle bystander to the reasoning whose reliability has come into question. It was his own reasoning, preceding 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 third-personal evidence available to Sherlock.

This is where the *reflection strategy* comes in. It can say 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 Watson should give it up. The reflection strategy thus has potential to honor multiple, seemingly conflicting inclinations many of us have, by assigning non-evidential significance to first-personal reflective beliefs or evidence.

This is roughly what led David Christensen to propose an immediate connection between our reflective beliefs and our first-order rationality.<sup>36</sup> But some recent discussions including Christensen's say what matters is really Watson's first-personal beliefs about his *reliability*, not about his rationality.<sup>37</sup> This takes a step away from conventional reflectivism, which stresses the normative content of reflective beliefs about rationality. But I think it is still not far enough. So long as it is supposed to be first-personal beliefs about his reliability that

---

<sup>35</sup> Christensen and White

<sup>36</sup> Cf. Dutant and Littlejohn (MS).

<sup>37</sup> Christensen 2016; Dogramaci 2015, Sec. 1.4; Weatherson 2019, Ch. 8; and perhaps Horowitz 2014.

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 think that reflective beliefs should guide you only 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 that evidence governs inference, in 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.<sup>38</sup> 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.

The relevant relationship between propositions, ordinary evidential support, also can be understood inclusively. Familiar examples include deductive, inductive, and abductive support, but I cite these just for illustration. The important thing is that ordinary standards of evidence are invariant under third-personalization. If I can inductively infer that I will be bored at tomorrow's meeting from the premise that I have been bored at previous meetings, then I can inductively infer that DB will be bored at tomorrow's meeting from the premise DB has been bored at previous meetings. I take it this goes for anything like conventional evidential support. If beliefs about me license inferences that beliefs about other people don't, it is not because the facts about me are stronger evidence by ordinary standards.

---

<sup>38</sup> cite

So EVIDENTIALISM says your inferential reasoning should follow the evidence. You should not be persuaded that  $p$  by a premise unless the premise is ordinary evidence supporting that  $p$ . And that goes for the premises supplied by reflective beliefs, too. The premise that you are required to believe that  $p$  should persuade you that  $p$  only if it is evidence for  $p$  under ordinary, third-personal standards.

Reflectivists could resist the argument for deflationism by denying this. They might say these ordinary third-personal standards are incomplete. When I learn that I am required to believe  $p$ , that ought to persuade me that  $p$ , 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. Instead, 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.<sup>39</sup> From my point of view, my being required to believe  $p$  bears more directly on its being the case that  $p$ . If I think that some character DB is required to believe that  $p$ , that should persuade me that  $p$  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 that  $p$ , that should be enough to convince me, even if it isn't sufficient evidence by the ordinary third-personal standards. Call this view (oxymoronically) *transparency reflectivism*, because of its similarity to transparency accounts of self-knowledge.<sup>40</sup>

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 transparency reflectivism faces the problem of *chauvinism*.<sup>41</sup> 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 has been investigating a murder, and has compiled a long list of hypotheses, about each of which he suspends judgment. In the morning, Sherlock tells Watson of a Mystery Detective, who has been investigating the same murder. This Mystery Detective, Sherlock says, believes a lemma that entails one of Watson's hypotheses. But while the lemma is well supported by the Mystery Detective's evidence, 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 hypothesis in question is that the butler did it.

In the morning, Watson should think the Mystery Detective is required to believe one or another of the hypotheses on the list. But that does not mean Watson should think the

---

<sup>39</sup> Cf. Gibbons 2006 and 2013.

<sup>40</sup> E.g., Byrne 2018, Fernandez 2013, and Moran 2001.

<sup>41</sup> Transparency accounts of self-knowledge arguably face related problems. See Barnett 2016, Boyle 2011, and Valaris 2011.

hypothesis is true. Usually an agent's being required to believe something is evidence it is true. But in this case the Mystery Detective's belief is required only because of a false lemma. While the hypothesis he is required to believe might coincidentally be true, it might not. So Watson should believe the Mystery Detective is required to believe the hypothesis in question, but suspend judgment on whether it is true.

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. Maybe that could be rational in special circumstances, if he has evidence supporting that he is somehow immune to misleading evidence. But absent something like that, Watson still should suspend on whether the hypothesis is true.

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 never be persuaded that the butler did it, since his being required to believe it is not ordinary, third-personal evidence.

Now certain features of this example are debatable. It might be argued that by the evening, Watson really should not believe that he is required to believe the butler did it.<sup>42</sup> (But is this reflective belief so irrational that I cannot even stipulate Watson holds it, and ask how he should proceed?) Or perhaps the example might just be misrepresentative in some way. Watson not only lacks evidence his situation is not misleading, but has evidence it is. Maybe transparency reflectivists could say that if so, the reasons usually provided by reflective beliefs are defeated.<sup>43</sup>

But the general lesson goes beyond any one example. We all agree that an agent's being rationally required to believe something is usually evidence it is true. Transparency reflectivism goes beyond this mundane observation, and says you should sometimes be persuaded that a conclusion is true by the premise that you are required to believe it, even when the premise is not sufficient evidence by third-personal standards. And if so, it is bound to require reasoning chauvinistically sometimes.

Transparency reflectivists might back off claims about 'positive' reflective beliefs that a belief is required. Maybe Watson should not reason "I am required to believe the butler did it, so he did it," unless his being required to believe is evidence. But that still leaves 'negative' reflective beliefs that a given belief is prohibited. Maybe Watson still should reason "I am required to doubt the butler did it, so maybe he didn't do it." This concession might mean licensing as rational some forms of akrasia and Moore-paradoxicality. But it might leave sufficient resources to handle higher-order undermining, as in Watson's Apparent Mistake.

Yet the chauvinism problems arises for negative reflective beliefs as well:

**Mystery Detective's False Defeater:** Watson has been investigating a murder, and has compiled a long list of conclusions, each of which he

---

<sup>42</sup> Thanks to Alex Worsnip for pressing this.

<sup>43</sup> Thanks to Jonathan Ichikawa for pressing this one.

believes. In the morning, Sherlock tells Watson of a Mystery Detective, who has been investigating the same murder. This Mystery Detective, Sherlock says, 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's evidence arguably supports that he has evidence opposing the butler's guilt. If so, he should reflectively believe he is required not to believe the butler did it. (And if not, we still can suppose Watson does believe this.) Should this persuade him the butler might not have done it? Usually when an agent has evidence opposing a conclusion, that fact is itself evidence opposing the conclusion. But here, the evidence only opposes Watson's conclusion by supporting some unspecified falsehood. So in the morning, Watson should think the Mystery Detective is required to doubt the relevant conclusion, but that might give Watson no reason to doubt it is true. 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." So Watson still should think the conclusion is probably true, despite believing he is required not to believe it. 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 just for pressing but also for believing. The pragmatist's idea is not that Cash should *infer* that it will rain from the premise that he will get a prize. Instead, he is supposed to 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." In doing so, he treats the prize as bearing not on the meteorological question whether it will rain, but on the practical question whether to bring it about that he believes something. According to pragmatist, that can be good enough for his belief to be rationally well-motivated.

Inferentialists like me disagree. Even if a prize can be a good reason to press a button causing you to believe it will rain, that does not make it a good reason to believe. To be a reason to believe—that is, to be able to serve as your motivating reason for believing, and not just for pressing—your being offered the prize would somehow need to be evidence that it will rain. And the same goes for reflective beliefs too. Suppose you are convinced that you are rationally required to believe it will rain. Even if that gives you reason to press a belief-causing button, that won't make it a reason to believe. To be a reason to believe, your being required to believe needs to bear on the question whether it will rain.

Like pragmatists, management reflectivists deny that motivating beliefs need to provide premises. And so reflective beliefs can just motivate you to see to it that you believe it will rain, like knowledge of the prize does for Cash. You might think “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 think you are 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.<sup>44</sup> And in any case, it is a natural fit for reflectivists moved 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?<sup>45</sup> 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 that 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 by pressing a button for example. Maybe Matilda’s reasons for doing so could even be counted as broadly epistemic, since they involve true belief rather than cash. But they still won’t be good reasons for believing the coin lands heads. Unless Matilda has evidence the coin lands heads, her belief will be irrational.

Maybe this could be disputed, or 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. 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 phobia or bias.<sup>46</sup> But if

---

<sup>44</sup> White 2006, pg. 539 and ???, and Dutant and Littlejohn.

<sup>45</sup> Berker

<sup>46</sup> White, Greco, maybe Gendler?

reflective belief revision is supposed to be something we can easily pull off, it cannot involve believing at will.

The second problem concerns the particular motives that management reflectivism demands. Matilda is after true belief, but management reflectivism gets nowhere if it has you motivated to believe the rational thing only when you are already convinced it is true.<sup>47</sup> It needs you to be motivated even when you are not antecedently convinced.

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 turns out to be false. But it is in a subjective sense of 'ought'. It is not as if being rational is always objectively for the best. If you unknowingly have gasoline in your glass, it might be rational to drink, and that might mean in some sense you ought to. But it doesn't mean that drinking is objectively for the best, as if the small price of drinking gasoline were worth paying for the greater good of rationality. "Better to die rationally than live irrationally," said nobody, ever. So if a friend saw you irrationally not drinking, and had the option to make you rationally drink by pressing a button, that need not motivate them to press. Pressing would mean *fetishizing* rationality, by prioritizing it above objectively good outcomes.<sup>48</sup> Being rational does not require us to fetishize rationality like this.

But fetishism is what management reflectivism demands, for instance in:

**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, whatever it is.

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. 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 will be motivated to press the button. Given the circumstances, Watson should think there is a serious risk the belief he produces will 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. Even if this fetishism is permissible, it is hardly required.

---

<sup>47</sup> Cf. Boghossian, BonJour, and White

<sup>48</sup> Cf. Dogramaci 2017, pp. 64-65. The term is borrowed from Williams, though I'm not sure his objections to fetishism is the same as mine.

Pressing is still fetishistic in the afternoon, when Watson learns he is the Mystery Detective. Just imagine him thinking “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. That might be 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:

**Suspension Button:** As in Mystery Detective’s False Lemma, Watson has a list of conclusions, each of which he believes. In the morning, Watson is told by Sherlock about a Mystery Detective whose evidence supports a false defeater for one of the conclusions. In the afternoon, Watson is told that he is the Mystery Detective, and in the evening that the conclusion in question is that the butler did it. This time, Watson has a button that will cause the Mystery Detective to suspend judgment about the conclusion.

Will Watson be motivated to press? Again, only if he fetishizes rationality. In the morning he thinks the Mystery Detective is required to doubt the conclusion. But Watson still should think it is probably true, whatever it is. If so, pressing is likely to rid the Mystery Detective of an irrational but true belief. Watson does not need to be fetishistically motivated to press in such circumstances. And it is hard to see why this would change in the afternoon, when Watson learns that pressing would likely rid himself of a true but irrational belief, or in the evening, when he learns which belief it 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 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.<sup>49</sup> In a slogan, the idea is that 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 having a further *guiding belief* about the logical or evidential connection between them.<sup>50</sup> Importantly, the role of the further 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. 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, rather than supplying a new premise.

---

<sup>49</sup> Cf. the ‘second-order reasons’ view criticized by Whiting 2017, which combines elements of management reflectivism and guidance reflectivism.

<sup>50</sup> See, e.g., Neta 2019, and Boghossian’s (2008) discussion of the ‘Intention View’. Also Broome and Siegel and Richard

But I think guiding beliefs are poorly suited to the jobs reflectivists have for them. 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 one 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 guidance reflectivist motivated by akrasia and Moore's paradox should have our agent believe not- $p$  instead. But if the reflective belief does not itself provide her premise, as guidance reflectivism claims, then on what basis is she supposed to believe it? 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. Remember Sherlock's Admonition, where Watson reflectively believes on Sherlock's authority that the clues support the butler did it, even though they really do not. Here Watson is misled not about what his clues are, but instead about what those clues support. Arguably, Watson still should respond by believing the butler did it. But according to guidance reflectivists, a reflective belief like Watson's need not provide some additional reason to believe the butler did it. Instead, it is supposed to enable him to believe the butler did it based on his preexisting reasons or evidence. The problem is that Watson's preexisting evidence does not really support that the butler did it. Watson cannot believe it based on this evidence without reasoning fallaciously. Maybe Watson could infer the butler did it from the higher-order premise that his clues support it, but he cannot infer this from the clues themselves.

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. Instead of thinking you are somehow or other required to hold a belief, you will always have some specific idea in mind of how to go about it. But if we allow reflective beliefs to be based on empirical 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 premises one started with. 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 that could take him directly 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 want to avoid 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 thinks 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 told 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 as reliable as he is, and suspends judgment on what the clues support.

A steadfast about alleged higher-order defeat 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 positive belief that his clues support suspension, but has a negative belief that they do not. So guidance reflectivists cannot allow his suspended judgment to be well-founded.

Could it be that suspended judgment does not need to be well-founded to be rational? It might if suspension is a kind of default attitude, not requiring a basis in evidence like belief does. If so, Watson might still rationally suspend in his disagreement with Lestrade, despite knowing his evidence supports another attitude. Now I doubt belief and suspension really are so different. But if they were, it would only raise further problems for cases like:

**Watson's Conflicting Evidence:** In the morning, Watson believes the butler did it based on clues that inductively support his guilt. In the afternoon, Watson learns new evidence that abductively supports that really the maid did it, and so suspends judgment on whether the butler did it. In the evening, Watson is reminded by Sherlock that he has an abysmal track record at abductive reasoning, though he has never had any trouble with inductive reasoning.

Unless we are steadfasters in general, we should say that by evening Watson ought to believe the butler did it, since his doubts have been defeated. But to explain this, reflectivists seem bound to appeal at least to a negative belief that suspension is prohibited, if not a positive belief that belief is required. If guidance reflectivism cannot allow the former to make suspension irrational, or the latter to make belief rational, it cannot handle the full range of higher-order evidence cases.

## 8. Conclusion

Where does all this leave the central motivations for reflectivism? Probably some have just got to go, like common intuitions about epistemic akrasia and Moore-paradoxicality. But I still have hope for certain intuitions about higher-order evidence cases.

The problem raised by the cases is this. There are many people out there whose track record or biases I might learn about. But only one is me, the same person whose reasoning led to these very beliefs now before my mind. It seems there should be some way for me to single out for special consideration what I know about this person in particular. But it won't come just by my following ordinary evidential standards, like those governing deductive, inductive, and abductive inferences. Since these standards are invariant under third-personalization, I can conform to them without making any special use of information about myself. 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.

Even if conforming to ordinary evidential standards is all it takes to be rational, it might seem somehow suboptimal. Wouldn't an ideal agent have some further capacity that allows it to coordinate third-personal information about its reasoning with how it conducts that reasoning? Maybe this is the kernel of truth behind the reflectivist approach to higher-order defeat, and even to the ambitious claim that reflective reasoning as the additional capacity that distinguishes ideal agents from the brutes. Maybe an ideal agent would have such a capacity, and maybe we do, too. But even if all that is so, reflectivism offers a false theory of what it is.

## References

- 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.
- 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.
- Burge, Tyler (2013) *Cognition Through Understanding*. OUP.
- Byrne, Alex (2018) *Transparency and Self-Knowledge*. OUP.
- Chalmers, David and Alan Hájek (2007) 'Ramsey + Moore = God' *Analysis* 67(2): 170-172.
- Christensen, David (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.
- 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.
- Dutant, Julien and Littlejohn, Clayton (MS) 'Defeaters as Indicators of Ignorance'
- Egan, Andy and Elga, Adam (2005) 'I Can't Believe I'm Stupid' *Philosophical Perspectives* 19(1): 77-93.
- Elga, Adam (2000) 'Self-Locating Belief and the Sleeping Beauty Problem' *Analysis* 60(2).
- (2005) 'On Overrating Oneself... And Knowing It' *Philosophical Studies* 123(1-2): 115-124.
- Fernandez, Jordi (2013) *Transparent Minds: A Study of Self-Knowledge*. OUP.

- Gibbons, John (2006) 'Access Externalism' *Mind* 115(457): 19-39.
- (2013) *The Norm of Belief*. OUP.
- 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.
- 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.
- 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.
- 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 (MS) 'The Rationality of Practical Hope'
- 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.

- 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 Velleman, J. David (2005) 'Doxastic Deliberation' *Philosophical Review* 114(4): 497-534.
- Shoemaker, Sydney (1996) *The First-Person Perspective and Other Essays*. Cambridge University Press.
- 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.
- Turri, John (2012) 'A Puzzle about Withholding' *Philosophical Quarterly* 62(247): 355-364.
- 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.
- (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.
- 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.