

The Evil Demon Inside

Nicholas Silins

Cornell University

Penultimate Draft

Final Version to Appear in *Philosophy and Phenomenological Research*

Abstract: this paper examines how new evil demon problems could arise for our access to the internal world of our own minds. I start by arguing that the internalist/externalist debate in epistemology has been widely misconstrued---we need to reconfigure the debate in order to see how it can arise about our access to the internal world. I then argue for the coherence of scenarios of radical deception about our own minds, and I use them to defend a properly formulated internalist view about our access to our minds. The overarching lesson is that general epistemology and the specialized epistemology of introspection need to talk---each has much to learn from each other.

Suppose you're not radically deceived by an evil demon, and consider your counterpart who is. Things seem just the same way to your twin as they do to you, and your twin has the same beliefs about how things are as you, but your twin is in general not reliable about how things around them are. Are their beliefs about the external world still just as justified as yours? If your answer is "yes," you are an internalist in epistemology. If your answer is "no," you are an externalist in epistemology. (If your answer is "maybe," it is time to pick a side.)

The case I have just described poses the "new evil demon" problem in epistemology, introduced by Cohen and Lehrer (1983). The problem arises for

For their help with this paper, I'd like to thank Robert Beddor, Alex Byrne, Ophelia Deroy, Julien Dutant, Carl Ginet, Anna Giustina, Anil Gomes, Patrick Greenough, Uriah Kriegel, Neil Mehta, Lisa Miracchi, Ram Neta, Adam Pautz, Antonia Peacocke, Abelard Podgorski, Adriana Renero, Eric Schwitzgebel, Sydney Shoemaker, Susanna Siegel, Declan Smithies, Weng Hong Tang, Hannah Trees, Ru Ye, Jonna Vance, Jonathan Vogel, Timothy Williamson, and two anonymous referees. Special thanks to Lauren Ashwell, Karen Bennett, Pericles Lewis, Odette Lienau, Andrew McGonigal, Carlos Montemayor, David Orr, Padmaja Rajagopalan, Aziz Rana, Mira Seo, and Pramod Vijayasankar. I'm also grateful to audiences at workshops or other events at the University of Geneva, the Institut Jean-Nicod, the Universidad Panamericana, the Ohio State University, Oxford University, Bled, the University of Bergen, Nanyang Technological University and the National University of Singapore.

those externalist theories that seem to predict the victim lacks justified beliefs about the external world, given the failure of the victim's beliefs to be knowledgeable or reliable or otherwise properly connected to the truth. Strikingly, self-described externalists have tried to accommodate some internalist verdicts about the case (Goldman 1986, 2011; Sosa 2003), to explain (false) internalist verdicts about the case (Lasonen-Aarnio forthcoming; Williamson forthcoming), or to construct analogues of the new evil demon scenario about which internalist verdicts are implausible (Srinivasan forthcoming). Most or all epistemologists, internalist or externalist, seem to agree that the new evil demon scenario raises a challenge for externalist views.

My aim here is to examine how new evil demon problems might arise for our access to the internal world of our own minds. In part 1, I will argue that the internalist/externalist debate in epistemology has actually been widely misconstrued—we need to reconfigure the debate in order to see how it can arise about our access to the internal world. In part 2, I will argue for the coherence of scenarios of radical deception about our own minds, and I will use the scenarios to defend a properly formulated internalist view about our access to our minds. The overarching lesson will be that general epistemology and the specialized epistemology of introspection need to talk—each has much to learn from each other.

1. Understanding Internalism vs. Externalism about Introspection

First, some preliminaries about “introspective justification”, my main focus throughout the paper. To get the phenomenon in view, consider how you usually access what you're thinking about, and contrast how other people access what you're thinking about. While other people can have justification to believe you're thinking that it's time for lunch (L), you have justification to believe L in a way such that other people can't have justification to believe L in that way. The crucial point here is negative, leaving open the positive details about how introspective justification works. Whether or not you have anything like perceptions of your own

mental states, you do have justification for beliefs about your own mind in a way that does not extend to your beliefs about other minds. In the terms of Byrne (2005), introspective justification is “peculiar”. For all we have said, introspective justification need not be perceptual in any way, despite the misleading term of “introspection”, and introspective justification need not be superior to your justification for beliefs about other minds, or “privileged” in the sense of Byrne (2005). This neutrality is all for the good, allowing us to accommodate as many approaches to introspection as possible.

I'll now review a standard picture of what internalist/externalist debates in epistemology are about. I'll then revise that picture by showing how it neglects the possibility of internalist/externalist debates about introspective justification.

As we saw in the introduction, internalist views are standardly motivated by consideration of new evil demon scenarios. The lesson is supposed to be that you and your counterpart are equally justified in your beliefs, regardless of how you are hooked up to the world. Since Conee and Feldman's influential “Internalism Defended” (2001), it has become common to think of internalism in epistemology as then potentially coming in two flavors, each trying to specify what you and your deceived counterpart importantly have in common (see e.g. Srinivasan 2015: 276). According to the first version of internalism,

(Accessibilism): If two people are the same with respect to what they have introspective and a priori justification to believe, then they are the same with respect to what they have justification to believe.

This approach proceeds in epistemological terms to identify a base for justification, and privileges the sources of introspection and a priori reflection. As Conee and Feldman note, accessibilists vary in terms of what exactly they privilege, ranging from “direct access” to “suitable awareness” to “introspective awareness” (2001: 2). I will generally pass over these variations here (Alston 1986: 91-2 surveys further candidates).

The second flavor of internalism is quite different:

(Mentalism): If two people are the same with respect to their internal mental states, then they are the same with respect to what they have justification to believe.

Here the relevant mental states are meant to be specified somehow other than through introspective access. For example, Wedgwood (2002, forthcoming) or Schoenfield (2015) focus on those of your mental states that are non-factive, i.e. such that you can bear them to the proposition that p even if it is not the case that p. For the sake of simplicity, I will simply stick with the place-holder "internal mental state."¹

According to philosophers such as Timothy Williamson (2000, 2007) or Declan Smithies (2012a), internalism is ultimately best understood along the lines of accessibilism (where Williamson goes on to reject accessibilism because of its allegedly excessive accessibilist demands). According to philosophers such as Conee and Feldman, Wedgwood, and Schoenfield, internalism is ultimately best understood along the lines of mentalism. On a more ecumenical approach, internalism can be understood in either way, where neither has a better claim to capture the spirit of internalism.

I will now argue that all of these options are wrong.

I will first argue that Accessibilism is not the way to understand internalism, and not even a way to understand internalism. The core problem is simple. Given that the antecedent of Accessibilism is formulated using the notion of justification, Accessibilism cannot be used to formulate an internalist/externalist debate about introspective or a priori justification themselves. The approach takes introspective and a priori justification as given, without incorporating any stand on their own requirements.

To see that Accessibilism cannot be used to formulate an internalist/externalist debate about introspective and a priori justification, look at what happens when we try:

¹ Whether the restriction to non-factive mental states is adequate is not so clear---perhaps some internalists would want to assign a role to mental states that are not propositional attitudes at all. Consider the role traditionally assigned to mental states such as acquaintance.

(Accessibilism+): If two people are the same with respect to what they have introspective and a priori justification to believe, then they are the same with respect to what they have introspective and a priori justification to believe.

This view is trivial, and must be accepted by all regardless of their orientation. The core of internalist/externalist debates in epistemology cannot then be about Accessibilism. Accessibilism is not a statement of one version of internalism. Given its neutrality about the character of introspective and a priori justification, it fails to be a statement of any version of internalism.

For an example of the approach I am criticizing, consider Neta and Pritchard (2007), who distinguish the following array of claims:

- (1) S's epistemic justification for believing that p is constituted solely by S's mental states.
- (2) S's epistemic justification for believing that p is constituted solely by facts that S can know by reflection alone.
- (3) S's epistemic justification for believing that p is constituted solely by properties that S has in common with her recently envatted physical duplicate (2007: 382).

According to a view they defend on behalf of John McDowell, (1) and (2) are true although (3) is false.² Nevertheless, they hold that McDowell's acceptance of (2) "seems to make his view internalist (2007: 386)." The idea seems to be that accepting an accessibilist claim is sufficient to make you an internalist. However, there is room for internalist/externalist disputes over what it takes to know by reflection alone that p. In particular, there is room for internalist/externalist disputes over what it takes to have any associated justification to believe that p. Accepting (2) is then not a way of committing to internalism, given that (2) is itself open to internalist/externalist disputes.

As an objection, someone might say that I have misconstrued the accessibilist approach by formulating it as a supervenience thesis. According to proponents of internalism such as Smithies (2012a, forthcoming), and opponents such as Lasonen-

² Pritchard gives an extended further defense of a similar position in his 2012 book.

Aarnio (2014), internalism is to be understood as the following sort of level-connection thesis:

(Levels Accessibilism): If you have justification to believe that P, then you have justification to believe that you have justification to believe that P.

This approach gets us a non-trivial form of Accessibilism about introspective justification:

(Levels Accessibilism +): If you have introspective justification to believe that P, then you have justification to believe that you have justification to believe that P.

In response, I would say that the levels approach has the same sort of problem as the initial supervenience approach—the levels approach is silent about the nature of the higher-level justification the view demands. The view leaves open the possibility that, in order to have justification to believe that p, you must have some reliabilist or otherwise externalist form of justification to believe you have justification to believe that p. Figuratively speaking, the view only says that, if you have a scoop of justification at the first floor, then you also have a scoop at the second floor, and so on, without telling us more about what it is that you have a scoop of at any floor. The view makes only a structural claim about justification, without further information about what stands in the structure, leaving us without a distinctively internalist form of the view.

To further bring out that Levels Accessibilism is not itself a version of internalism, we may return to new evil demon scenarios. Suppose you are justified in believing that you are sitting down, and consider your radically deceived internal duplicate with the same belief—is that person justified in his or her belief? Levels Accessibilism is entirely silent about how to compare the cases. It predicts that you have higher-level justification, since we built in that you have justification at the ground floor. But it says nothing about whether your deceived counterpart has

justification at any level. Indeed, even the following strengthened form of Levels Accessibilism is also silent about the comparison across cases:

(Strong Levels Accessibilism): You have justification to believe that P if and only you have justification to believe that you have justification to believe that P.

On the approach I am taking, if a view fails to predict equal justification in new evil demon scenarios (while allowing for their possibility), it is not a form of internalism. To defend my approach, I would emphasize that we need some relatively pre-theoretical way of grounding our discussion, to avoid having a merely terminological dispute. Given the traditional role of new evil demon scenarios as focal points for dispute between internalists and externalists, those cases seem to be the best way of grounding our discussion.

Let's now return to the initial supervenience statement of Accessibilism. To defend using it to capture internalism, someone might insist that internalism is trivially true about introspective and a priori justification, although internalism is not trivially true about other forms of justification. On this approach, Accessibilism remains a good way to understand internalist/externalist debates about those areas where there is room for internalist/externalist dispute.

To see why there needs to be room for internalist/externalist debate about introspective justification, reconsider you and your deceived counterpart. Arguably you have introspective justification to believe you are in various factive mental states such as *seeing that* the sun is shining, or *remembering that* you ate oatmeal for breakfast. After all, you can come to a justified belief that you are in such states in a way such that no one else can in that way. Now, an externalist would say your deceived counterpart does not have any justification to believe that they are in such factive mental states, given how unreliable your counterpart is about what factive mental states they are in. Here there is room for internalists and externalists to disagree about whether the pair of subjects is the same with respect to introspective justification. Still, the externalist remains free to endorse Accessibilism. If you and your deceived twin differ with respect to introspective justification, then you and

your twin fail to be counterexamples to Accessibilism—the antecedent of the thesis simply fails to apply to the pair of you.

It is important that Accessibilism fails to be an articulation of internalism. According to some critics, internalism is false because it makes accessibilist demands that are too strong for us to satisfy.³ Whether or not the critics are right about our ability to satisfy accessibilist demands, there is room for internalism to be true even if Accessibilism is false. While some mentalists have made a similar observation, my point is that objections to Accessibilism do not target any version of internalism, let alone one version of internalism.⁴ A similar point can be made in response to problem cases for Accessibilism presented by Siegel (2012, 2017). In Siegel’s key cases of “bad cognitive penetration”, two people are the same with respect to introspective justification, although very different in terms of the hidden etiology of their perceptual experience. In particular, the role of prejudice in generating one person’s perception arguably makes the pair differ with respect to justification, in a way that goes under their introspective radars. We do have a challenge here to Accessibilism, but not yet to any version of internalism.⁵

Now setting Accessibilism aside, the standard alternative would be to understand internalism about introspective justification along mentalist lines:

(Mentalism +): If two people have the same internal mental states, then they are the same with respect to what they have introspective justification to believe.

We have made some progress: this formulation does leave room for internalist/externalist disputes about our access to our factive mental states.

The formulation does not leave clear enough room for internalist/externalist disputes about our access to our *internal* mental states. The classic approach would be to set up the debate via a new evil demon scenario. And the usual recipe for such

³ See Williamson (2000, 2007), Lasonen-Aarnio (2014) or Srinivasan (2015).

⁴ See Wedgwood (2002, forthcoming) or Schoenfield (2015). The possibility of internalism without Accessibilism is perhaps suggested as early as Sosa (1999: 146-148).

⁵ Puddifoot (2016) argues in a similar fashion to Siegel against Accessibilism, using pairs of subjects who differ with respect to their putatively inaccessible implicit biases, and arguably also with respect to their justification of their beliefs. See Toribio (2018) for a response.

a scenario would be to hold internal mental states fixed, and to vary the relation between internal mental states and the relevant domain of the internalist/externalist dispute. Internalists would then hold that your beliefs are justified even in the scenario where they are systematically mistaken. But we can't hold someone's internal mental states fixed, and vary whether they have *true* beliefs about what internal mental states they are in. If one person is systematically mistaken in believing that they are in internal mental states M1...Mn, and another person is by and large correct in believing that they are in internal mental states M1...Mn, they can't also be in the same internal mental states.

If you use the classic recipe, you can't cook up a coherent new evil demon scenario about our access to our internal mental states. Here the problem is not directly that Mentalism+ is trivially true, but rather that it is impossible to formulate the classic sort of scenario an internalist would use to motivate it, and that an externalist might use to reject it.⁶

Contrary to received wisdom, a fully general version of internalism can't be understood either along accessibilist lines, or along mentalist lines. On the alternative approach I will promote, arguably tracing all the way back to Goldman (1979) and Cohen and Lehrer (1983), the core issue at stake is the kind of truth-connection constitutively required for justification. The question is about how we get justification from a given source, and specifically about whether the ability of a source to justify beliefs comes from its ability to supply true beliefs. On this approach, contrary to that taken by Alston (1986) and many others, formulations of externalism strictly speaking come first, with versions of internalism then constructed as their denials.

Externalist views come on a spectrum here according to how strong a truth-connection they demand. For example, a view framed in terms of infallibility will be more externalist than a view framed in terms of reliability. To demarcate the

⁶ For discussion of whether impossible new evil demon cases could still be useful in adjudicating internalist/externalist disputes, see Miracchi (forthcoming). Even if Miracchi is right that some (subtly) metaphysically impossible scenarios are still useful, I would say that her point does not carry over to new evil demon scenarios for internal mental states: they wear their incoherence on their face, and are not subtly impossible scenarios.

externalist range of the spectrum of the various truth-connections one might demand, I say a view is externalist just in case it predicts an absence of justification in evil demon scenarios of systematic error. These are cases where a source is not "reliable" on my preferred use of the term (more at the end of 2.2. about further flavors of reliability). In using demon scenarios to demarcate externalism, I rely on my earlier point that we need to avoid having a merely terminological dispute about how to capture "externalism" and "internalism", and that evil demon scenarios give us a ladder to see the key substantive issue at stake.

Now that I have characterized externalism, I can describe a further important problem for using Mentalism to try to capture internalism. Mentalism (and Accessibilism) are stated as supervenience theses, telling us that if two people are the same with respect to such and such factors, then they will be the same with respect to justification. They do not explicitly address what justifies your beliefs, or what makes it the case that whatever privileged source is able to justify your beliefs. In the case of Mentalism, this is a serious problem, since Mentalism ends up being compatible with a paradigmatically externalist form of reliabilism, even though Mentalism is stated in terms of internal mental states. To see the problem, consider views on which you visually experience a color only if your visual experiences of that color reliably track it. Two people could agree about the account of color experience, and yet disagree about whether the reliability of your experience in any way accounts for the justification you get from your visual experiences for color beliefs. Even if your experience of redness must be reliable to be an experience of redness, the reliability of your experience could still be an idle wheel when it comes to justification. Either way, Mentalism turns out to be compatible with paradigmatically externalist positions in epistemology—duplicates with respect to internal mental states could still turn out to be duplicates with respect to reliability, and so could still be duplicates with respect to justification even if an externalist form of reliabilism is true.⁷ Mentalism does not capture internalism.

⁷ For views that combine externalist individuation of mental states with externalism in epistemology, see Burge (2003) or Sawyer and Majors (2005). For some recent discussion of the combined approach, see Gerken (forthcoming).

It is by no means novel to understand externalism as accounting for justification in terms of connections to truth, similar formulations can be found as early as Goldman (1979).⁸ However, theorists who use such formulations do not address the limitations of mentalist or accessibilist formulations of internalism, and sometimes still use mentalist formulations of internalism (e.g. Miracchi forthcoming: 2). Since the letter of Mentalism does not say anything about truth-connections, we do not have a transparently exhaustive set up of the debate if we use Mentalism to capture internalism and truth-connections to capture externalism. To set up the debate in an exhaustive way, we need to formulate internalism simply as the negation of externalism. And to repeat the point from the previous paragraph, Mentalism even turns out to be compatible with externalism.

A further lesson here is that verdicts about new evil demon scenarios are not at stake in all internalist/externalist disputes. New evil demon scenarios are instead a useful way into the debate, and a typically convenient proxy for the core issue at stake. However, we cannot always set up externalist/internalist debates by constructing a typical new evil demon pair of cases, and asking whether the subjects in it are equally justified. In the case of internal mental states, typical new evil demon scenarios are impossible to construct—we cannot hold the subject's perspective fixed, and at the same time vary the internal mental states that in part constitute their perspective. Nevertheless, we can still count a view as externalist according to the strength of the truth-connection it uses to account for justified beliefs about your internal mental states.

In sum, a standard approach about how to understand internalist/externalist disputes in epistemology is mistaken. First, Accessibilism fails to capture any form of internalism. Second, Mentalism fails to capture internalism about introspective justification in particular. To set up internalist/externalist debates across the board, we should instead understand internalist and externalist views as coming on a spectrum, where views are more externalist as they account for justification in terms of a stronger connection to truth

⁸ See also Graham (2012), Gerken (forthcoming), or Miracchi (forthcoming)

In concluding this section, I will formulate the two particular families of externalism about introspective justification I will oppose in the remainder of the paper.

First consider

(Constitutivism): For any mental state *M*, if you are introspectively justified in believing that you are in *M*, your belief that you are in *M* is justified at least part in virtue of your being in mental state *M*.

Here you cannot have an introspectively justified belief that is false, due to the nature of introspective justification. On the approach to understanding the externalist/internalist debate I have defended, this is then an extreme form of externalism about introspective justification.⁹

I will also oppose the following weaker claim:

(**Introspective Reliabilism**): For any mental state *M*, if you are introspectively justified in believing that you are in *M*, then your belief that you are in *M* is justified at least in part in virtue of being formed in a reliable way.

According to this view, on my intended reading of "reliable", introspection will not give justification in any demon scenario in which it systematically leads to false beliefs. Introspective Reliabilism is often assumed by critics of the reliability of introspection such as Dennett (1998) or Schwitzgebel (2008), who seem to think that a lack of introspective justification will follow from a lack of reliability.¹⁰ The view might also be assumed by some traditional theorists who think that

⁹ See Shoemaker (1996, 2009), Zimmermann (2006), Neta (2008), or Smithies (2012b, forthcoming). For a similar view in the case of perceptual justification, consider Johnston (2006). On his approach, when you are perceptually justified in believing that *a* is *F*, that is in part in virtue of your being perceptually aware of *a*'s *F*ness, where you could be so perceptually aware only if it is the case that *a* is *F*. Here you cannot have a perceptually justified belief that is false, due to the nature of perceptual justification. Here I count this view as an extreme form of externalism.

¹⁰ For example, Schwitzgebel writes that introspective judgments are not "sound" or "trustworthy" if they fail to be reliable (2008: 267-8), and states that they are "untrustworthy" after having argued for their unreliability (2008: 246, 2008: 259). (He uses the same wording in his 2011: ch. 7). That said, the connection between being trustworthy and a source of justification is open to debate. For direct use of the term "unjustified", with a strong implication that reliability is necessary for justification, see Schwitzgebel 2011: 137, 139.

introspection is reliable, and who think we have introspective justification (you could accept Introspective Reliabilism without accepting reliabilism across the board).

In principle, you could oppose Introspective Reliabilism by arguing only that the justified status of introspective beliefs fails to come from their reliability—perhaps reliability always coincides with introspective justification while remaining an idle wheel. In what follows, I will defend the following stronger form of internalism:

(Introspective Internalism): It is possible to be introspectively justified in believing that you are in M, even though your belief that you are in M is not formed in a way that is reliable.

I will argue that the reliability of introspective beliefs is not even a necessary condition for introspective justification, taking the cases of belief and consciousness in turn.

2. Adjudicating Internalism vs. Externalism about Introspective Justification

2.1. Judgment and Belief

I will now argue that our judgments can give us introspective justification about our beliefs even when our judgments are not a reliable guide to what we believe.

To get to this conclusion, I will start by using some examples to clarify what I have in mind by "judgment" and "belief". When someone asks you your name, and you sincerely answer "My name is N.N.", you judge that your name is N.N. You can also make the judgment when you simply answer the question in your mind but do not speak up. We get an initial grip on the idea of judgments as outer or inner assertions, but we do not have to build in that they go along with outer or inner speech—perhaps you can judge that p without entertaining any form of words. Now, before you were asked the question about your name, you had a standing

belief that your name is N.N. without judging that your name is N.N. While we'll discuss the relation between judgment and standing belief further in what follows, for now we may note one important difference: judgments are always conscious, standing beliefs are not always conscious.

Let's now look at how judgments can justify you in believing that you have this or that standing belief, in a way that supports Introspective Internalism.

I will start with the idea that, when you answer the question whether p, that puts you in a position to answer the question whether you believe that p.¹¹ In particular, I will put the idea as follows:

(Transparency): When you judge that p, you get introspective justification from your judgment to believe that you believe that p.

We can defend Transparency by using it to explain what's wrong with judgments of the following "Moore-Paradoxical" character:

(MoPa): P and I don't believe that P.

Other things being equal, it is irrational to make such judgments. As Goldstein (2000) puts it so well, such Moore-paradoxical judgments tend to be "Mooronic".¹²

When you judge the conjunction, I take it you judge each of its conjuncts. However, if Transparency is true, when you judge the first conjunct, that will put you in a position where you have justification actually to reject the second conjunct. You will now be irrational in judging the overall conjunction to be correct. In making a Mooronic judgment, you flout the justification given to you by your judgment that p.

¹¹ While commonly assigned to Evans (1982), the idea of Transparency arguably traces back to Edgley (1969). For further debate about how to refine and explain the Transparency idea, see Moran (2001), Byrne (2005), Gertler (2011), Silins (2012, 2013), Cassam (2014), James Barnett (2016), and A. Peacocke (2017).

¹² For further discussion of Moore's paradox and issues about self-knowledge, see Shoemaker (1996) and (2009), also Smithies (forthcoming).

Transparency also helps to capture the distinctive ways in which Mooronic judgments are defective. It is somehow rationally defective for you to judge that you do not exist, but there is nothing rationally worse about you judging that [snow is white and you do not exist]. There is no tension between judging that snow is white and judging that you do not exist. However, there is a tension between judging that snow is white, and judging you do not believe that snow is white. To capture what is distinctively defective about Mooronic judgments, Transparency points out that judging one conjunct gives you justification to reject the other. (While someone might try to deny Transparency's claim that your judgment is the specific source of your justification, it is doubtful that there just happens to be a correlation between judging that *p* and having justification to believe you believe that *p*).

We do need to say more to secure that judgment is a source of specifically introspective justification. When you judge that *p* after all, that might also seem to supply *me* with justification to believe that you believe that *p*. However, if your judgment ever does supply me with justification to believe you believe that *p*, it does so only via publicly available evidence I have such as your having said that *p*. In contrast, I need not use publicly available evidence when my judgments give me reason to believe that I believe that *p*. Also, if your judgment ever gives me reason to believe you believe that *p*, I must rely on the premise that you have judged that *p*. In my own case, I do not need to rely on the premise that I have judged that *p*. To see why, consider the odd case of a thorough-going skeptic about whether judgments exist, who makes a judgment of the form MoPa. The judgment is still Mooronic, and we can give a good explanation of why using Transparency, even though such a person does not believe or otherwise endorse the premise that she has judged that *p*. Given these asymmetries, it looks like your judgments give you distinctively introspective justification to believe you believe that *p*.

I have offered an explanation of what is wrong with Moore-paradoxical judgments, but what about Moore-paradoxical beliefs? Something seems rationally defective about having a standing belief of the form MoPa, and Transparency might seem to be able to explain why. Perhaps some other story is superior to

Transparency since that other story provides a uniform treatment of judgment and belief.

In reply, I first emphasize that Transparency might be able to explain what is rationally defective with Mooronic beliefs if beliefs are suitably linked to dispositions to judgment (see C. Peacocke 1998 for such an approach). Second, in any case I take it to be *more* rationally defective to make a Moore-paradoxical judgment than to have a Moore-paradoxical belief. You are flouting your epistemic resources in a more egregious way when you make a Moore-paradoxical judgment, where Transparency is optimally positioned to explain why (for agreement about the asymmetry between belief and judgment, see e.g. Kriegel 2004: 110-1). If a rival approach says that we do not get introspective justification from our judgments, but instead only from our beliefs themselves or some other source, that approach seems unable to explain why it is worse to make Moore-paradoxical judgments than to have Moore-paradoxical beliefs. So it looks like Transparency has the best explanation of what is wrong with Mooronic judgments.

Having argued that we get introspective justification from our judgments, I will now argue that we get introspective justification from our judgments even when they are not a reliable guide to our beliefs.

As a first step, consider that your judging that *p* is not sufficient for you to believe that *p*. To see how, consider cases where your judgment misfires, failing to reflect your beliefs. For example, consider how in a philosophical discussion, you might assert and judge that *p*, and then quickly retract. “Scratch that,” you might say, “that’s not what I believe”. (See Martin 1998 for more discussion of such philosophical cases.) For a more quotidian example, suppose you have been gazing at photos of the Sydney Opera House, and are asked to identify the capital of Australia in a game of trivia. You might answer that the capital of Australia is Sydney, judging that is so, but then soon correct your mistake, realizing that is not what you believe (Smithies forthcoming, ch. 5.2).¹³

¹³ For a sample view on which judgment is insufficient for belief, see C. Peacocke (1999), for an objection, see Zimmermann (2006). For further discussion of judgment and belief, and dispositional requirements for belief, see Schwitzgebel (2002, 2010, 2013), but also the critical response to

How can such mismatches between judgment and belief arise? A plausible diagnosis is that, in order to believe that *p*, you need to have enough dispositions associated with believing that *p*. I will not take a stand on exactly what the dispositions are, since I am trying to convince as many interlocuteurs as possible, across a wide range of views of what it takes to have a belief. As Schwitzgebel (2002: 252-4) points out in his survey of candidate behavioral, inferential, and phenomenal dispositions, there may not even be any particular disposition necessary or sufficient for belief that *p*, perhaps having enough of a cluster will do. The core problem is that the conscious surface of your mental life when you judge that *p* does not guarantee the presence of non-conscious dispositions to underwrite a standing belief that *p*. To use a toy example, suppose that believing that *p* requires being disposed to plan on the assumption that *p*. You can make a conscious judgment that *p* without having a disposition to plan on the assumption that *p*. The same holds for a disposition to act on the assumption that *p*, or a disposition to infer something that follows from the proposition that *p*, and so on. Your conscious perspective when you judge that *p* will not necessarily fix what you will do across the range of circumstances for manifesting dispositions associated with believing that *p*. Now, when you judge that *p*, you might manifest dispositions defined solely in the terms of the actual circumstances in which you are judging that *p*. But I take it that dispositions defined so narrowly trivialize the view that beliefs requires dispositions, and are not the dispositions proponents of the non-trivial approach had in mind.

An objector might insist that, whenever someone judges that *p*, they at least have a short-lived conscious belief that *p* insofar as they judge that *p*. This objection speaks past the topic of our discussion. We are concerned with how judgments can justify you in self-ascribing standing beliefs that need not be conscious. The objection changes the subject away from judgment's relation to standing belief, to conceive instead of judgment itself as a form of conscious belief.

Schwitzgebel by Borgoni (2016). For a recent book-length defense of pragmatic requirements for having a belief, see Zimmerman (2018). For some relevant empirical evidence about belief attributions, see Murray et al (2013) or Buckwalter et al (2015).

There is a gap between judgment and (standing) belief. Let's now consider how an evil demon might insert the skeptical knife.¹⁴ Suppose that whenever Rene judges at a time that p, an evil demon typically but indiscernibly ensures that Rene at that time fails to believe that p. In particular, the evil demon momentarily erases enough dispositions associated with believing that p, so that Rene ends up failing to believe that p while he judges that p and considers whether he believes that p. The demon subsequently restores the dispositions. Here Rene's use of the transparency method is not reliable, without Rene having any inkling of the disruption. (Now, you might have disagreed with my use of dispositions to explain the gap between judgment and belief. If so, you will also disagree with some of my description of Rene's case. However, if you still agree that it is possible to judge that p without believing that p, as when the quiz contestant missteps in judging that Sydney is the capital of Australia, please substitute in your own explanation of the gap to flesh out the case of Rene.)

Even though Rene's use of the Transparency Method is unreliable, it remains irrational for Rene to make a judgment along the lines of "p, but I don't believe that p". And Transparency remains as good an explanation of why as before. So it looks like Rene continues to get introspective justification to self-ascribe beliefs from his judgments. And further, Rene still seems to form a justified belief that he believes that p when he does so on the basis of his judgment that p. Given that he does not form a belief in a way that is reliable, we now have an example that confirms Introspective Internalism.

We can use our example to raise a further objection to the following particularly strong form of constitutivism:

¹⁴ Contrast Evans, who writes: "If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental state, even the most determined sceptic cannot find here a gap in which to insert his knife (1982, 225)."

You might think that, given his anti-skeptical point, Evans must be using "belief" in a way that is synonymous with "judgment". But Evan's discussion of "self-ascription of a belief" opens with the example of someone asking "Do you think there is going to be a third world war?" The person is presumably concerned with whether I have a standing belief---unless I am currently stockpiling gas masks or the like, it would be bizarre to ask whether I am occurrently thinking that there will be WWII.

(Strong Constitutivism): If you believe that p, then you have introspective justification to believe you believe that p; if you do not believe that p, then you have introspective justification to believe you do not believe that p.

On this version of the view, endorsed by Smithies (2012b, forthcoming), absences of belief are self-intimating. The view has extraordinary implications for the case of Rene, allowing a kind of introspective bootstrapping. We can use standard formulations from the literature to set up this instance of the problem.¹⁵ Consider that each time Rene judges that p, it looks like he will end up with justification to believe the following sort of conjunction:

Track Record: I am judging that p but do not believe that p.

Once enough justification for each of the Track Record conjunctions has stacked up, it looks like Rene will end up with justification to believe the following hypothesis as the best explanation of what's going on:

Explanation: there is a systematic mismatch between my judgments and my beliefs.

But it is absurd that Rene could get even a low degree of justification to believe Explanation, whether or not Rene can muster a justified belief in Explanation on the basis of the justification he has (Smithies is careful to point out that you can have justification for a belief without being able to use it to form a justified belief on its basis). The whole point of an evil demon is meant to be that you can have no hint of how your present appearances are misleading.

Let me now address some objections to my case for Introspective Internalism.

First, one might try to explain away the internalist verdict about the case as tracking some property other than justification. For example, on the approach taken by Lasonen-Aarnio (2014, forthcoming), such judgments track the property of

¹⁵ For further discussion of the nature of bootstrapping problems and of who they might be problems for, see Cohen (2000), White (2006), Weisberg (2010), or Barnett (2014).

reasonableness rather than justification. The rough idea here is that a belief is reasonable when formed on the basis of a disposition that yields knowledge in suitably normal conditions, where that disposition might fail to lead to knowledge in the thinker's actual conditions. And perhaps when we judge that Rene is justified, we simply confuse the presence of reasonableness with the presence of justification.

To contest the proposed explanation, we can use variations of the original case. In one variant, Rene has a fluke-y one-off belief that he believes that *p* in response to his judging that *p*. In another variant, Rene mimics the disposition of our original Rene, but only because of further manipulation by an evil demon. In these variants, Rene does not manifest any disposition to believe he believes that *p* on the basis of judging that *p*, but he still presumably has justification to believe he believes that *p*—our earlier Moorean argument applies just as well as before. Here his belief is not “reasonable” in the objector's sense, and does not seem to be reasonable in the objector's sense, but still does seem to be justified.

There is also a more general problem with the objectors' strategy, one that reveals the irrelevance of trying to explain away internalist intuitions about Rene. I did not appeal to internalist intuitions in setting up the case—I instead used consideration of Mooronic judgments to build an argument.

As a further objection, one might appeal to variant formulations of reliabilism in epistemology. Some versions of the approach use a lower bar for the reliability they demand, with the aim of predicting that victims of evil demons can retain justification. These versions of reliabilism are then compatible with the verdict that Rene retains introspective justification. For example, a reliabilist might merely demand a source that is reliable in normal worlds (Goldman 1986), or a source that is reliable in the actual world (Goldman 2001).¹⁶ Even though the transparency method leads Rene astray in his non-actual, abnormal world, he still uses a method that is reliable in the actual world (ours), and that is reliable in normal worlds such as ours. Some reliabilists therefore have room to maintain that Rene gets

¹⁶ For further discussion of how to formulate reliabilism, see Goldman and Beddor SEP, Sosa (2001), Comesaña (2002), Graham (2017), or Miracchi (forthcoming).

introspectively justified beliefs through the transparency method. So perhaps my case for Introspective Internalism fails.

I take the objector to neglect my reconfiguration in section 1 of the externalist/internalist debate. The test for externalism I proposed, based on the historical role of new evil demon scenarios, was whether the view rules out the presence of justification in cases of systematic error. According to the resulting understanding of Introspective Internalism, it is possible to be introspectively justified in believing that you are in M, even if your belief that you are in M is formed in a way that systematically leads you to error. The case of Rene establishes Introspective Internalism, since the transparency method does systematically lead him to error, and yet still gives him introspectively justified beliefs. Here it is beside the point whether the transparency method is reliable in the actual world, or in normal worlds—what matters is that the method systematically leads him to error in his scenario. In sum, there might well be versions of reliabilism that are consistent with Introspective Internalism. But these will simply be internalist forms of reliabilism.

2.3. Consciousness and Attention

In this section I'll use the case of consciousness outside attention to argue for Introspective Internalism. Let me start by partly clarifying the debate about whether consciousness requires attention.

Rather than trying to define terms such as “experience” or “consciousness”, I will simply draw your attention to some of your own experiences, such as your visual experience of this text. That's the sort of thing I put under the heading of “experience” and “consciousness”. I also won't try to define the term “attention”, and instead would like to draw your attention to your attention itself. For instance, fix your eyes on the period at the end of this sentence, and at the same time shift your attention to your left. That is the sort of thing I put under the heading of “attention”.

Now, are you conscious only of that to which you attend? According to what we may call the *highlight view*, the answer is “no”—you sometimes experience more entities than those to which you attend (be they objects, properties, or other entities).¹⁷ According what I will call the *searchlight view*, the answer is “yes”—just as you might see only those things illuminated by a searchlight, you experience only those entities to which you attend.¹⁸

Much more can be said to clarify the debate about consciousness and attention, in particular by distinguishing different kinds of attention, but I will proceed on the assumption that we now have a good enough grip on the key notions and the key question.¹⁹

I will now rehearse a classic common sense motivation for the highlight view. The common sense motivation is that, when you turn your attention to something you had not previously been attending to, say a sound in the background, it can seem to you that you had been experiencing the sound before turning your attention to it. The temporal point here is important. The motivation is not just that, when I turn my attention to the sound, I experience the sound, but also that, when I turn my attention to the sound, I sometimes seem to have been experiencing the sound before I turned my attention to it.²⁰

The common sense motivation relies on the claim that our memories of a recent experience of *x* can justify us in believing that we had the experience of *x*, even when we also remember that we were not attending to *x* at the time. Since these memories are distinctively first-personal memories of our own experience, and not just of the scene experienced, they are also sources of introspective justification here if they are sources of justification at all. Perhaps I can rely on my memory of a scene to form a justified belief about the scene in a way you can too, but you can't rely on your memory of my experience to form a justified belief about my experience in the way I apparently can.

¹⁷ See Searle (1992), Koch and Tsuchiya (2007), or Jennings (2015).

¹⁸ See Mack and Rock (1998), Prinz (2011), or M. A. Cohen et al (2012).

¹⁹ For useful further clarificatory discussion, see chapter 5 of Wu (2014) or chapter 12 of Watzl (2017).

²⁰ See Block (1995), and for related discussion of the visual case, see Martin (1992, 2001) and Smithies (forthcoming).

Proponents of the searchlight view have a response readily available, according to which this common sense motivation for the highlight view is simply an instance of the “refrigerator light illusion” (cf. Thomas 1999). Although a naïve subject might think that the refrigerator light is on when the fridge door is closed, really the subject has overgeneralized from the cases in which the door is open and the light is on. The idea is that we are doing no better when we think about consciousness and attention. The “refrigerator light illusion” hypothesis—henceforth just the “illusion hypothesis”—is in particular that, when we turn our attention to something we had not previously been attending to, we “turn the light on” and generate our conscious experience of the thing, rather than ever discovering an antecedent experience of the thing. Even if it does seem to us that we antecedently experienced the thing, we are misled if we believe that we did antecedently experience the thing.

The illusion hypothesis is in effect an evil demon hypothesis about our access to consciousness, one on which the evil demon is inside us. The illusion hypothesis entails that our memories of our recent conscious experiences in the absence of attention are unreliable.

To see that we can set up a case for Introspective Internalism here, let us start by supposing that the illusion hypothesis is true. Presumably there’s a possible situation in which the hypothesis is true of us whether or not it is not actually true of us. Next, contrast the situation of an ordinary subject with that of the highlight theorist who has been challenged with the illusion hypothesis in a debate. It could be that the highlight theorist ceases to be justified in relying on memory once challenged with the illusion hypothesis (or not, it’s not clear how much of a challenge the illusion hypothesis poses unless further evidence is offered in its favor). Be that as it may, the ordinary subject unaware of the hypothesis is presumably still justified in relying on memory of her recent experiences even if the illusion hypothesis is true. To see that this verdict is plausible, imagine another subject in the illusion scenario who seems to remember consciously hearing the bell ring even though she didn’t attend to it. Without any reason to suspect that the illusion hypothesis is true, this skeptical subject denies that or suspends judgment

about whether she consciously heard the bell ring without attending to it. Such a stance would seem justified if the subject has gained reason to suspect that her memory is unreliable, but in our description of the case she does not. A good explanation of what has gone wrong here is that she is flouting justification she receives from her memory. Going back now to the deceived subject who is not suspicious of her memory, she is introspectively justified in believing that she consciously heard the bell ring without attending to it, even though she in fact has not formed her belief in a way that is reliable. Here we have another case that confirms Introspective Internalism.

Given how one can defend the highlight view, and given how searchlight theorists respond, we can construct another counterexample to externalist requirements on introspective justification. Our peculiar justification for beliefs about the internal world again stays even when our reliability goes.

Conclusion

Discussions of introspection are often conducted in isolation from general epistemology, without attention to the full spectrum of debates and distinctions out there. In particular, while philosophers such as Schwitzgebel and his critics have debated whether introspection is reliable, we have lacked a much-needed internalist/externalist debate about the role of reliability in introspective justification.²¹ That said, discussions of general epistemology themselves suffer from neglecting the case of introspection. In particular, when philosophers discuss new evil demon problems, they standardly fail to address whether new evil demon problems could arise for our access to our own minds. This lapse causes serious trouble for the traditional set up of internalist/externalist debates in epistemology. In order to understand the disagreement properly, we need to set aside the role of our internal mental states, as well as the role of access requirements, and instead focus directly on what sort of truth-connection allows us to understand justification.

²¹ See Schwitzgebel (2008, 2011), and Bayne and Spener 2010, Smithies 2013, or Bayne 2014.

Having now done the needed work, I hope we can see we could remain introspectively justified even if we did have an evil demon inside.

REFERENCES

- Alston, William. P. (1986). Internalism and Externalism in Epistemology. *Philosophical Topics* 14 (1):179-221.
- Barnett, David. J. (2014). What's the matter with epistemic circularity? *Philosophical Studies*, 171(2): 177-205.
- (2016). Inferential justification and the transparency of belief. *Noûs*, 50(1), 184-212.
- Block, Ned. (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences* 18 (2):227--247.
- Borgoni, C. (2016). Dissonance and Irrationality: A Criticism of The In-Between Account of Dissonance Cases. *Pacific Philosophical Quarterly*, 97(1), 48-57.
- Buckwalter, W., Rose, D., & Turri, J. (2015). Belief through thick and thin. *Noûs*, 49(4), 748-775.
- Burge, Tyler. (2003). Perceptual entitlement. *Philosophy and phenomenological research*, 67(3): 503-548.
- Byrne, Alex. (2005): "Introspection", *Philosophical Topics*, 33: 79-104.
- Cassam, Quassim. (2014). *Self-knowledge for humans*. Oxford University Press, USA.
- Cohen, Stewart and Lehrer, Keith (1983). Justification, truth, and coherence. *Synthese* 55 (2):191-207.
- (2002). "Basic Knowledge and the Problem of Easy Knowledge", *Philosophy and Phenomenological Research*, 65: 309-29.
- Cohen, M. A., Cavanagh, P., Chun, M. M., & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in cognitive sciences*, 16(8), 411-417.
- Comesaña, Juan. (2002). The diagonal and the demon. *Philosophical Studies*, 110(3), 249-266
- Conee, Earl and Feldman, Richard (2001). Internalism Defended. *American Philosophical Quarterly* 38 (1):1 - 18.
- Dennett, Daniel. C. (1988). Quining qualia. In *Consciousness in modern science*. Oxford University Press.
- Edgley, Roy. (1969). *Reason in Theory and Practice*. London: Hutchinson.
- Evans, Gareth. (1982). *The Varieties of Reference*. Oxford: OUP.
- Gerken, Mikkel (forthcoming). The New Evil Demon and the Devil in the Details. In *The Factive Turn in Epistemology*. Cambridge University Press. Available at https://philarchive.org/rec/GERTNE-2?all_versions=1
- Gertler, Brie (2011). Self-Knowledge and the Transparency of Belief. In Anthony Hatzimoysis (ed.), *Self-Knowledge*. Oxford University Press.
- Goldman, Alvin. I. (1979). What is justified belief?. In *Justification and knowledge* (pp. 1-23). Springer Netherlands.

- (1986). *Epistemology and Cognition*. Harvard University Press.
- (2011). Toward a Synthesis of Reliabilism and Evidentialism. In T. Dougherty (ed.), *Evidentialism and its Discontents*. Oxford University Press. pp. 254.
- Goldstein, Laurence. (2000), 'Moore's Paradox', in P. Engel (ed.), *Believing and Accepting* (Dordrecht: Kluwer Academic Publishers), 65 — 92.
- Graham, Peter J. (2012). Epistemic Entitlement. *Noûs* 46 (3):449-482.
- Graham, P. J. (2017). Normal Circumstances Reliabilism: Goldman on Reliability and Justified Belief. *Philosophical Topics*, 45(1), 33-61.
- Jennings, Carolyn Dicey. (2015). Consciousness without attention. *Journal of the American Philosophical Association*, 1(2), 276-295.
- Johnston, Mark (2006). Better Than Mere Knowledge? The Function of Sensory Awareness. In John Hawthorne & Tamar Gendler (eds.), *Perceptual Experience*. Oxford University Press. pp. 260--290.
- Koch, Christoph., & Tsuchiya, N. (2007). Attention and consciousness: two distinct brain processes. *Trends in cognitive sciences*, 11(1), 16-22.
- Kriegel, Uriah. (2004). Moore's paradox and the structure of conscious belief. *Erkenntnis*, 61(1), 99-121.
- Lasonen-Aarnio, Maria. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research*, 88(2), 314-345
- (forthcoming). Virtuous Failure and Victims of Deceit. In Julien Dutant and Fabian Dorsch, eds., *The New Evil Demon*, Oxford: Oxford University Press.
- Lyons, J. C. (2013). Should reliabilists be worried about demon worlds?. *Philosophy and Phenomenological Research*, 86(1), 1-40.
- Mack, Arien, and Irvin Rock. 1998. *Inattentional blindness*. Vol. 33. Cambridge, MA: MIT press.
- Majors, Brad., & Sawyer, Sarah. (2005). The epistemological argument for content externalism. *Philosophical Perspectives*, 19(1), 257-280.
- Martin, Michael G. F. (1992). Perception, concepts, and memory. *Philosophical Review* 101 (4):745-63.
- (2001). Out of the past: Episodic recall as retained acquaintance. *Time and memory: Issues in philosophy and psychology*, 257-284.
- Miracchi, Lisa (forthcoming). Competent Perspectives and the New Evil Demon Problem. In Julien Dutant (ed.), *The New Evil Demon: New Essays on Knowledge, Justification and Rationality*. Oxford University Press. Available at <https://philarchive.org/rec/MIRCPA>
- Moran, R. 2001: *Authority and Estrangement*. Cambridge: Harvard University Press.
- Murray, D., Sytsma, J., & Livengood, J. (2013). God knows (but does God believe?). *Philosophical studies*, 166(1), 83-107.
- Neta, Ram., & Pritchard, Duncan. (2007). McDowell and the new evil genius. *Philosophy and Phenomenological Research*, 74(2), 381-396.
- Neta, Ram (2008). The nature and reach of privileged access. In Anthony E. Hatzimoysis (ed.), *Self-Knowledge*. Oxford University Press.
- Peacocke, A. (2017). Embedded mental action in self-attribution of belief. *Philosophical Studies*, 174(2): 353-377
- Peacocke, Christopher. 1999: *Being Known*. Oxford: OUP.
- Prinz, Jesse. (2012). *The conscious brain*. Oxford University Press

Pritchard, Duncan. (2012). *Epistemological disjunctivism*. Oxford University Press.

Puddifoot, K. (2016). Accessibilism and the challenge from implicit bias. *Pacific Philosophical Quarterly*, 97(3), 421-434.

Schoenfield, Miriam (2015). Internalism without Luminosity. *Philosophical Issues* 25 (1):252-272.

Schwitzgebel, Eric. (2008). The unreliability of naive introspection. *Philosophical Review*, 117(2), 245-273.

----- (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91(4), 531-553.

----- (2011). *Perplexities of consciousness*. MIT press.

Searle, John. R. (1992). *The rediscovery of the mind*. MIT press.

Shoemaker, Sydney. (1995), 'Moore's Paradox and Self-Knowledge', *Philosophical Studies*, 77: 211 – 28

----- 1996: *The First-Person Perspective and Other Essays*. Cambridge: CUP.

----- (2009). Self-intimation and second order belief. *Erkenntnis*, 71(1), 35-51.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs*, 46(2), 201-222.

Silins, N. (2012). Judgment as a Guide to Belief. In Smithies and Stoljar (eds.) *Introspection and consciousness*, 295-328.

Silins, N. (2013). Introspection and inference. *Philosophical Studies*, 163(2), 291-315.

Smithies, Declan (2011). What Is the Role of Consciousness in Demonstrative Thought? *Journal of Philosophy* 108 (1):5-34.

----- (2012a). Moore's Paradox and the Accessibility of Justification. *Philosophy and Phenomenological Research*, 85(2), 273-300.

----- (2012b). A Simple Theory of Introspection. In Smithies and Stoljar (eds.), *Introspection and Consciousness*. Oxford University Press.

----- forthcoming. *The Epistemic Role of Consciousness*. OUP. Available at <https://sites.google.com/site/declansmithies/Home/the-epistemic-role-of-consciousness>

Sosa, Ernest (1999). Skepticism and the internal/external divide. In John Greco & Ernest Sosa (eds.), *The Blackwell Guide to Epistemology*. Blackwell. pp. 145--57.

----- and Bonjour, Laurence (2003). *Epistemic Justification: Internalism Vs. Externalism, Foundations Vs. Virtues*. Wiley-Blackwell.

Sosa, E. (2001). Goldman's reliabilism and virtue epistemology. *Philosophical Topics*, 29(1/2), 383-400.

Srinivasan, Amia (2015). Normativity without Cartesian privilege. *Philosophical Issues* 25 (1):273-299.

----- (forthcoming). Radical Externalism. Available at http://users.ox.ac.uk/~corp1468/Research_&_Writing_files/Radical%20Externalism_current.pdf

Toribio, J. (2018). Accessibility, implicit bias, and epistemic justification. *Synthese*, 1-19.

Watzl, Sebastian (2017). *Structuring Mind. The Nature of Attention and How it Shapes Consciousness*. Oxford, UK: Oxford University Press.

Wedgwood, Ralph (2002). Internalism Explained. *Philosophy and Phenomenological Research* 65 (2):349-369.

- forthcoming. Internalism Re-explained. In Julien Dutant (ed.), *The New Evil Demon: New Essays on Knowledge, Justification and Rationality*. Oxford University Press. Available at http://www-bcf.usc.edu/~wedgwood/Internalism_reexplained.pdf
- Weisberg, Jonathan. (2010). Bootstrapping in general. *Philosophy and Phenomenological Research*, 81(3), 525-548.
- White, Roger. (2006). "Problems for Dogmatism", *Philosophical Studies*, 131: 525-557.
- Williamson, Timothy. (2000). *Knowledge and its Limits*. Oxford: OUP.
- (2007). On being justified in one's head. In *Rationality and the good: Critical essays on the ethics and epistemology of Robert Audi*, 106-22.
- (forthcoming). Justification, Excuses, and Skeptical Scenarios. In Julien Dutant (ed.), *The New Evil Demon: New Essays on Knowledge, Justification and Rationality*. Oxford University Press. Available at http://media.philosophy.ox.ac.uk/assets/pdf_file/0018/35343/Luxembourg.pdf
- Wu, Wayne (2014). *Attention*. Routledge.
- Zimmerman, Aaron. Z. (2006). Basic self-knowledge: Answering Peacocke's criticisms of constitutivism. *Philosophical Studies*, 128(2), 337-379.
- Zimmerman, A. Z. (2018). *Belief: A pragmatic picture*. Oxford University Press.