Synthese (2021) 198:9763–9785
https://doi.org/10.1007/s11229-020-02680-6

The evil demon in the lab: skepticism, introspection, and introspection of introspection

Nicholas Silins[1]

Received: 23 August 2019 / Accepted: 29 April 2020 / Published online: 28 May 2020

Abstract
In part one, I clarify the crucial notion of "introspection", and give novel cases for the coherence of scenarios of local and global deception about how we access our own minds, drawing on empirical work. In part two, I evaluate a series of skeptical arguments based on such scenarios of error, and in each case explain why the skeptical argument fails. The first main upshot is that we should not over-estimate what it takes to introspect: introspection need not be accurate, or non-inferential, or exclusive of perception, or even exclusive of confabulation. The second main upshot is that, while skeptical challenges by figures such as Carruthers, Doris, and Schwitzgebel are rich and empirically informed, these skeptical challenges founder on how they are epistemologically under-informed.

Keywords Introspection · Self-knowledge · Skepticism · Reliability

1 Introduction

When Descartes tried to whittle down his beliefs to those that are certain, he used the knife of an interfering evil demon. It is now standard to understand skeptical problems as follows: given the possibility of being radically deceived by an evil demon, what if anything can we know or justifiedly believe? While such evil demon problems for our access to the external world are widely discussed, they are much less discussed in the case of our access to our own minds. That extension will be my focus here.

In particular, I will focus on our access to how we access our own minds. Let me briefly explain, with more to come soon. When philosophers do question the scope of our introspective knowledge, they often focus on whether introspection is reliable (e.g. Dennett 1988; Schwitzgebel 2008, 2011), or on whether our being in a mental state is sufficient for us to have introspective knowledge that we are in that mental state (e.g.

✉ Nicholas Silins
ns338@cornell.edu

[1] Cornell University, Ithaca, USA

Williamson 2000 or Srinivasan 2015). In these contexts it is generally assumed that we do introspect, the question is just about the epistemic quality of our introspection. The challenges I will survey are more startling, they question whether we know we are introspecting at all. These challenges are important because of how radical they are, and because of how they require us to evaluate necessary conditions for introspection, thereby giving us greater understanding of what introspection is and what it is not.

In part one, I will begin to clarify the crucial notion of "introspection", and give novel arguments for the coherence of scenarios of local and global deception about how we access our own minds. Here my conclusion is that we can be unreliable about when we introspect. In part two, I will tease apart a series of skeptical arguments based on these scenarios of error. Here my conclusion is that, even though we can be unreliable about when we introspect, no skeptical argument departing from such scenarios is successful.

At some stages of the paper, I will respond to empirically driven skeptical work by figures such as Carruthers, Doris, and Schwitzgebel. The evil demon need not be a figment of the philosophers' imagination—the evil demon has entered the lab (through the developmental work of Gopnik (1993), the evil demon even has access to our children). While their focus is not squarely on introspection of introspection, in ways I will detail later, we can apply the templates of the arguments they use to our topic, and learn how their arguments fail when applied to our topic and when applied to theirs. While the work of these skeptical figures raises fascinating challenges from psychology, and should be addressed by anyone interested in self-knowledge, I will try to show that their work neglects crucial points in epistemology. In the course of seeing how their skeptical arguments fail, we will gain a sharper understanding of what it does and doesn't take for us to introspect.

2 Set up and some cases of error

The first step is to clarify "introspection." Our best way in is through examples rather than definitions. Consider how you seem to ordinarily determine what someone else is thinking about, and contrast how you seem to ordinarily determine what you are thinking about. Apparently, when you ordinarily determine what you are thinking about, you do so in a way that is unavailable to others to determine what you're thinking about. In turn, it seems that other people can determine what they are thinking about in a way that is unavailable to you. In such cases, I will say that people make "introspective judgments" about their minds, singling out those judgments we form about our own minds in a way available only to us.

Introspective judgments are characterized by how they are formed rather than just by their mental subject matter. The characterization of introspective judgments thereby opens up some possibilities of error—even when you are right in thinking that you made a judgment about your mind, you might be wrong about how you made that judgment about your mind. That said, our gloss of "introspective judgment" does not commit to any positive account of how they are made, and certainly does not commit to any perceptual model of how they are made. The central negative point is that introspective judgments are not made in a way that can be used for our judgments

about other minds. In the useful terms of Byrne's (2005) work on the topic, they are "peculiar". It's a good thing that our gloss is circumspect in this way, since it allows for us to have substantive, non-verbal disputes over the positive details of how introspection works.

Given the negative characterization of introspection by Byrne, myself, and many others, a judgment might well be introspective even though it is also formed in an inferential way.[1] Importantly, we have also left open the positive details about how we form judgments about other minds. While one might assume that our access to other minds is straightforwardly observational or inferential, philosophers such as Stump have argued that at least our second-person knowledge of other minds is much more intricately grounded in social interaction.[2] So it is best that we avoid commitment to a simplistic model of knowledge of other minds. Finally, for all we have said, two judgments may be introspective even if they are formed in two different ways. There is no need for such a thing as *the* way or even *the* primary way such that no one else can make a judgment about our minds in that way. With our judgments as with so many areas, there may be many ways to be peculiar.[3]

Finally, my focus is on ordinary introspective judgments, not characterized by having an especially strong epistemic status, or "privilege" in the terms of Byrne (2005). One might instead try to identify a subset of introspective judgments that are (also) characterized by the strength of their epistemic status. Perhaps some introspective judgments are epistemically optimal thanks to training and careful reflection.[4] Or perhaps some introspective judgments are epistemically optimal thanks to having a special kind of "self-verifying" or otherwise reliable content.[5] These are introspective judgments on steroids. I address introspective judgments we make using ordinary mental categories and without special training. These introspective judgments are open to evil demon problems even if more sophisticated introspective judgments potentially are not.

We can now start to get more clear about the distinction between our access to our minds, and our access to our access about our minds. In what follows, I will separate two ways for us to be reliable with respect to introspection, contrasting the following two questions:

(**Reliability OF Introspection**): When we make introspective judgments, how reliable are they?
(**Reliability ABOUT Introspection**): How reliable are we about whether we are making an introspective judgment?

The first question concerns our introspective judgments at the first floor, and roughly looks at how likely they are to be true. For example, when we think we are in pain, how

[1] For some examples, see Smithies and Stoljar (2012: p. 4), Dretske (2012: p. 49), Siewert (2012: p. 129), or Spener (2012: 384).

[2] See Stump (2010: chs. 3, 4), Talbert (2015, 2017), or Benton (2017).

[3] Contrast the set up in Schwitzgebel 2008 that does build in a commitment to uniqueness: "Think of introspection as you will—as long as it is *the* primary method by which we normally reach judgments about our experience in cases of the sort I'll describe. *That* method, whatever it is… (2008: p. 248, emphasis mine)." (The same wording committed to uniqueness is used in Schwitzgebel 2011, ch: 7).

[4] See Spener (2015) for further discussion.

[5] See Chalmers (2003), Horgan and Kriegel (2007), or Gertler (2012).

reliable are we about whether we are in pain? This is the sort of question discussed by figures such as Dennett and Schwitzgebel. The second question ascends a level. This question roughly looks at how likely we are to be introspecting when we think we are introspecting. For example, when we think we've made an introspective judgment that we're in pain, how reliable are we about whether we've made an introspective judgment that we're in pain?

To illustrate the distinction between the reliability OF introspection and reliability ABOUT introspection, I'll now explore a surprising way in which they're related, one that shows we must make some introspective mistakes about our own minds. The link will emerge when we assess the following answer to the OF question:

> (**Introspective Infallibility**): If you introspectively judge that you are in a mental state M, then you are in M.

On this extreme view, formulated without restriction to particular kinds of mental states, we are the popes of our own minds. While Introspective Infallibility is now widely rejected, it is important not to reject it for the wrong reasons. By closely assessing a classic case against Introspective Infallibility, we'll gain understanding of how questions OF and ABOUT introspection interact. We'll also get a new argument for the old conclusion we must make some introspective mistakes about our minds.

Here is a classic case attributed to Rogers Albritton, here recounted by Christopher Hill:

> The case involves a college student who is being initiated into a fraternity. He is shown a razor, and is then blindfolded and told that the razor will be drawn across his throat. When he feels a sensation he cries out: he believes for a split second that he is in pain. However, after contemplating the sensation for a moment, he comes to feel that it is actually an experience of some other kind. It is, he decides, a sensation of cold. And this belief is confirmed when, a bit later, the blindfold is removed and he is shown that his throat is in contact with an icicle rather than a razor (1991: pp. 128–9).

Let's focus on the student's judgment that he is in pain. According to the critics of Introspective Infallibility, the student makes an introspective judgment that is false.

In response, the defender of Introspective Infallibility might argue he makes an introspective judgment that is true. Perhaps his thinking that he feels pain causes him to feel pain. But this response has a time lag problem. Even if he can think his way into being in pain, there is presumably some however brief amount of time when he thinks he is in pain, but is not yet in pain.[6]

[6] Perhaps other mental states of the student result in him being in pain. Perhaps they even jointly cause him to be in pain, and to believe he is in pain, so that there is no time lag problem. Contrast the following pathways:

Belief that in pain → pain	
Suitable cluster of mental states	] → pain
	] → belief in pain

A more promising line is that the student does not make an introspective judgment at all. Introspective Infallibility actually allows that we can make mistakes about what mental states we are in, provided that we do not make those judgments in an introspective way. In particular, given the way that the student's judgment relies on his testimonial belief that there is a razor on his throat, perhaps it fails to be introspectively made.

While the suggestion is initially promising, it ultimately backfires. To see why, we need to consider a variant of the case in which the student is quite reflective. In this variant, the student has taken "Intro to the Philosophy of Mind," and he has opinions about the status of his beliefs. In particular, he judges on the basis of introspection that [he believes he is in pain on the basis of introspection]. On the current approach to saving Introspective Infallibility, now we have a new introspective judgment that is false. In the effort to avoid an introspective error at one floor, the defender of Introspective Infallibility now has simply introduced an introspective error at the next floor. Their move is self-defeating.

More generally, in order for Albritton's challenge to succeed, the student need not make an introspective mistake at the ground floor. In our reflective variant of the case, the student must make an introspective mistake at either the ground floor or the next floor. We need not identify the floor where the error occurs to see that there is an error.

In response, you might insist that whether a judgment is introspective or not is a matter of how it is caused, and that no one should expect us to have introspective access to how our mental states are caused (even critics of traditional approaches to introspection such as Carruthers (2011) seem to endorse this move). To see the move in action, consider it as a response to the classic work by Nisbett and Wilson (1977). According to philosophical lore, Nisbett and Wilson (1977) asked subjects to choose between pairs of stockings and to give reasons for their choice, finding that subjects gave reasons for their choice such as the greater silkiness or smoothness of the chosen stockings. As the lore continues, the stockings were identical, and the subjects' choice was caused by the right-most location of the chosen stockings rather than any reason involving non-spatial differences between the stockings.

Footnote 6 continued
One challenge for this response is metaphysical—to spell out the more detailed causal story in a plausible way. For many conditions C, it might seem that we'll be in pain and believe we're in pain when C obtains, even though we don't end up even believing we're in pain when C rolls around. Just consider the range of cases in which we might say "oh, that actually wasn't bad". Another challenge for the response is epistemological—if the belief that you are in pain fails to be caused by your pain, and also fails to cause your pain, it's less clear whether they're suitably related for you to know that you're in. Presumably a defender of infallibility wants to defend the knowledgeable status of our introspective judgments as well as their truth.

Ironically, crucial details of this flagship case of confabulation have been confabulated—Nisbett and Wilson's classic paper is widely misunderstood. Nisbett and Wilson (1977) give no examples of what reasons subjects did state for their choices, and never say that the subjects had false beliefs about differences between the stockings such as their feel. The paper thereby leaves entirely open the possibility that subjects said they picked at random, with no illusions about the identical character of the stockings. Also, Nisbett and Wilson themselves speak against a causal explanation in terms of spatial position, mentioning a possible role of temporal bias towards the most recently seen garment.[7]

Assuming that Nisbett and Wilson' subjects did have false beliefs of some form or the other about their reasons for their choices, a natural defensive move is to restrict the scope of introspection. The idea is that our reasons for our choices are causes of our choices, and that we shouldn't be expected to have introspective access to considerations about how our choices were caused.[8] Since the introspective status of a judgment is a matter of its causal history, perhaps introspection should not be expected to extend to whether we introspect.

I suspect that both skeptics and non-skeptics about introspection have been too quick to exclude introspective access from considerations about the causal history of our mental states. On reflection, the exclusion is quite demanding. Consider a case in which you form a belief on the basis of conscious reasoning, and indeed self-consciously do so. Here I would say it is plausible that you can have introspective access to the fact that you have formed your belief on the basis of conscious reasoning. You presumably know that you formed your belief on the basis of conscious reasoning, and you don't seem to have relied on some way of knowing available to someone else for that conclusion, so it seems that whatever way of knowing you have used is introspective.

There is also empirical evidence against the orthodox exclusion of causal processes from introspection. For experimental work suggesting that we sometimes introspectively access to the way that we have guided our attention when performing a visual search for a target, see Reyes and Sackur 2017. And for experimental work suggesting that we sometimes introspectively access whether parallel or serial memory processing has played a (causal) role in our answering a question, see Reyes and Sackur 2018.

[7] For critical discussion of Nisbett and Wilson (1977), and references to further critical discussions, see section 1.3 of Newell and Shanks (2014).

Bortolotti points out that, even if the Nisbett and Wilson explanation in terms of spatial position fails, their stocking study could still supply a case of confabulation (2018: p. 4).

[8] For references to examples of philosophers making this move, see Engelbert and Carruthers (2010: 249). They also describe the stocking study in accord with the lore.

At a minimum, there is room for reasonable debate about the scope of introspection, and we should not assume in advance that introspection is excluded from the causal history of our present mental states.

In particular, I would say it is reasonable to expect that introspection can extend to whether a judgment is introspective as well (later we will see that our introspective access to our introspecting seems to be assumed even by a broadly skeptical figure such as Schwitzgebel). Notice that, if introspection never extends to our own introspection, it is not clear how it would be so easy to use first-person reflection on ordinary examples to get across intended cases of the phenomenon. When we draw a person's attention to cases in which they form a judgment introspectively, we do not ask them to use public evidence that they have introspected (it is not even clear what such public evidence would be). Rather, it seems sufficient to rely on the person's own sense that they have made a judgment in a special, introspective way.

Stepping back, we now see how debates about the reliability OF introspection turn out to be connected to debates about our reliability ABOUT introspection. You might try to protect the perfect reliability OF introspection by giving up perfect reliability ABOUT introspection. However, since judgments ABOUT introspection are themselves sometimes introspectively made, the move turns out to give up on the perfect reliability OF introspection after all.

We must make some introspective mistakes, where those mistakes might be at the ground floor, or at the next level up. To continue building up to evil demon problems, let's now consider some scientific threats to our perfect reliability ABOUT introspection, and even our reliability period ABOUT introspection.

To try to make some scientifically grounded progress here, I will leave behind Albritton's student, and turn to the *Journal of Obesity Research* for Brian Wansink et al.'s (2005) study of a bottomless bowl of soup.

Wansink et al. asked participants to eat a soup lunch and 20 min later to answer questions about how much they had eaten and their level of satiety. While all participants thought they had an ordinary bowl of soup, only some participants did, and the others actually had a bottomless bowl of soup that would slowly refill as they ate. Wansink's finding was broadly that, while bottomless eaters consumed 73% more soup than ordinary eaters, bottomless eaters gave a similar estimation of how much they had eaten, and still rated their hunger and satiety broadly the same as ordinary eaters.[9]

[9] Here is the full table from Wansink et al. (2005):

Footnote 9 continued

Biased visual cues unknowingly influence overconsumption*

	Visual cues consumption		
	Accurate visual cue (normal soup bowls)	Biased visual cue (self re-filling cue bowls)	F test (1, 5)
Actual consumption volume			
Actual ounces of soup consumed	8.5 ± 6.1	14.7 ± 8.4	8.99[c]
Actual calories of soup consumed	15.4.9 ± 110.3	267.9 ± 153.5	8.99[c]
Estimated consumption volume			
Estimated ounces of soup consumed	8.2 ± 6.9	9.8 ± 9.2	0.46
Estimated calories of soup consumed	122.6 ± 101.0	127.4 ± 95.6	0.03
*Consumption monitoring**			
"I carefully paid attention to how much I ate"	4.9 ± 2.3	5.3 ± 2.4	0.69
"I carefully monitored how much soup I ate"	4.7 ± 2.5	4.7 ± 2.8	0.00
"I usually eat until I reach the bottom of the bowl"	6.2 ± 2.1	6.6 ± 2.5	0.31
"I always try to clean my plate (or bowl) at home"	6.4 ± 2.2	6.1 ± 2.7	0.20
*Presence of others**			
"If other people keep eating, I am more likely also to"	5.5 ± 2.4	5.4 ± 5.7	0.03
"Eating with other people distracted me from how much I was eating"	4.7 ± 2.8	4.6 ± 2.5	0.00
Self-perceptions of satiety[a]			
"How hungry are you right now?"	3.4 ± 2.1	3.0 ± 1.9	0.63
"How full are you right now?"	5.7 ± 1.9	5.1 ± 2.7	1.03
"How nauseated are you right now?"	3.3 ± 2.3	2.6 ± 2.0	1.47
"How much food do you think you could eat right now?"	7.1 ± 1.7	7.0 ± 1.8	0.04

Values are mean ± SD
*Measured with agreement scales (1 = strongly discharge; 9 = strongly agree)
[a]Measured with semantic differential scales (e.g. 1 = a little; 9 = a lot)
[b]$p < 0.05$
[c]$p < 0.01$

I work with the example of bottomless soup for its vividness, but don't forget to add salt. Much of Wansink's work has recently suffered after a notorious blog post by Wansink himself about dubious research practices that he has deleted, and subsequent close critical scrutiny by van der Zee et al. (2017).[10] Still, van der Zee et al., as meticulous as they are, do not raise any direct challenges to the experiments I'll discuss here. Also, note that we could also use similar studies by Barbara Rolls involving smoothies puffed out with air (Rolls et al. 2000), manipulations of sandwich portions that did not affect hunger ratings (Rolls et al. 2004), or cheese puffs with lesser and greater aeration but the same caloric value (Osterholt et al. 2007). It is finally also important to note the availability of the original "bottomless bowl" study by Eva Daemmich reported in Pudel and Oetting (1977: pp. 383–4).[11]

Wansink et al. do not directly engage with philosophical debates about introspection, but they do make some suggestive remarks:

> These findings build on prior work by showing that individuals can base their satiation on visual cues related to portion size. In effect, people use their eyes to count calories and not their stomachs. Those shown biased visual cues had satiety ratings that were uncorrelated with actual consumption (2005: p. 98).

Here I defend an interpretation of their results that connects with our question of reliability ABOUT introspection.

First, I understand the participant judgments to concern how hungry or full they *feel*, so that we are properly focusing on judgments about mental states rather than non-mental bodily states.

Second, given the vastly greater amount of soup bottomless eaters had (73% more), it's plausible that bottomless eaters felt significantly less hungry than the ordinary eaters (I respond below to alternative descriptions of the case).

Third, given that the bottomless eaters had significantly different feelings of hunger from ordinary eaters, while having the broadly the same visual cues as ordinary eaters, it seems that bottomless eaters did not base their assessment of their own hunger by introspecting on how they feel. If they had, the bottomless eaters should have ended up with self-assessments that differed more dramatically from those by ordinary eaters given their different levels of feeling of hunger. Instead of making their judgment introspectively, the bottomless eaters seem to have made their judgment on the basis of how much soup they apparently ate—"biased visual cues"—where others could make a judgment on that same basis. On this interpretation, bottomless eaters thus did not make introspective judgments about how hungry they feel. (You might propose that the bottomless eaters judged their level of hunger on the basis both of visual cues, and introspection of their felt level of hunger or other "internal cues". But then the bottomless eaters should have ended up with levels of self-assessed hunger that were more divergent from ordinary eaters, again given their divergent levels of feeling of hunger.)

[10] A cached version is here: https://web.archive.org/web/20170312041524/http://www.brianwansink.com/phd-advice/the-grad-student-who-never-said-no.

[11] For surveys of a range of portion size effects, see Wadhera and Capaldi-Phillips (2014) and Benton (2015). For a meta-analysis of the studies, see Zlatevska et al. (2014). And for discussion of potential mechanisms of portion size effects, see the (2011) paper of Burger et al.

Fourth, while Wansink et al. don't ask the question, I take it that reflective participants in their experiment would think that they are making introspective judgments about how hungry they feel (whether or not the reflective subjects would use our philosophical jargon). We thus have cases where people can make mistakes about whether they are making an introspective judgment.

Finally, given the picture Wansink et al. defend, there is the threat of widespread error about when we introspect. In a wide range of cases where we seem to be introspecting about how hungry we feel, we would in effect be forming our opinions on the basis of the following sort of inference equally available to others: "My portion visibly isn't finished, so I must still feel hungry". Here we would be forming our opinions about how hungry we feel in ways such that other people could form them about us in that way. Now we have reached the possibility that we are not even reliable about when we introspect about how hungry we feel. In other words, we may be in an actual scenario of widespread deception for our access to when we introspect about how hungry we feel. In particular, if you are in the "clean plate club", you might often judge that you feel full on the basis of seeing your empty plate rather than on how you feel after finishing your portion of food.

Let me now address some objections before we proceed.

First, one might insist that the use of perception is compatible with the use of introspection. Here I'll address the most developed version of the challenge I am aware of in print, from Schwitzgebel 2012. While I think the point is ultimately correct, I do think that the details of Schwitzgebel's discussion need to be challenged, in a way that should further clarify what it is for a judgment to be introspective.

Schwitzgebel (2012) floats a two-fold form of pluralism about introspection.[12] First, and most importantly for our purposes, his pluralism arises "within cases" since many processes—including perceptual ones—can be involved within an episode of introspection. Second, his pluralism arises "between cases" since different batches of processes can be involved across episodes of introspection. On the sort of positive view Schwitzgebel suggests, there is nothing distinctive about your introspective judgments beyond their being made in a way intended to be sensitive to your mental states, using at least some processes that can only be applied to your mental states (so understood, his view looks to be compatible with tradition, but I'll set that worry aside). Applying the overall view to our current case, even if bottomless eaters do rely on visual cues, that is compatible with them making introspective judgments about how hungry they feel.

What reason do we have to believe Schwitzgebel's suggested form of pluralism? Some of his discussion draws our attention to the broad range of (perceptually-based) attitudes about the external world that can influence our judgments about our mental states. For example, in part using a case like that of Albritton's student, Schwitzgebel writes:

> If I see you move behind me with a red-hot poker and then suddenly I feel a startling touch on my neck, I might swiftly and readily judge that I'm feeling heat and pain, not coolness, even if you have actually touched me with an ice-cube (2012: 32).

[12] For discussion of further forms of pluralism, see Prinz (2004), Boyle (2009), or Samoilova (2016).

Schwitzgebel's purist opponent can agree that such cases occur. The problem would be that Schwitzgebel hasn't yet given us a reason to think the judgments are introspectively made.

A different challenge proceeds as follows:

> Here is another phenomenon that strains against the idea that introspection is a cognitively distinct process sharply separable from the process of outward perception. Judgments about sensory experience can easily collapse into judgments about the outside world with no crisp border between; and the two sorts of judgments, in such cases, are often seemingly driven by virtually identical processes… We gradually, insensibly traverse the distinction between introspective and non-introspective judgment. In such cases, introspection might be best regarded as perception with a twist or with a slightly different aim that can be half forgotten. The processes of perception, then, would be part of the process of introspection (2012: pp. 34–5).

It's not clear to me whether there are cases where our judgments are somehow indeterminate between being about our own mental states and being about the external world. Be that as it may, so long as there are some clear cases where we are making judgments about our own mind rather than the world, we do seem to be able to sharply separate introspection and perception. For a vivid illustration, consider someone who is sensorily deprived, and makes a true introspective judgment they are not outwardly perceiving. Since that person can accurately introspect they are not perceiving, there must be some sharp separation between processes of introspection and processes of perception.

While I think Schwitzgebel (2012) fails to establish his conclusion, he is importantly right that a judgment can be made in a way that is both introspective and perceptual. To make an introspective judgment, all you need to do is to make a judgment about your mind in *a* way such that no one else can make it in that way. Now, you might make a judgment about your mental state in a way that is peculiar, and that yet still recruits perception, testimony or some other source, so long as your judgment is not made solely using sources available to others. Partial peculiarity guarantees the presence of introspection.

It is a coherent possibility that the bottomless eater makes a judgment about how hungry they feel that is genuinely introspective, and yet also partially based on visual cues. But we have no argument that this possibility is actualized in the soup scenario. We also have evidence that the possibility is not actualized, given the convergence of assessments of hunger between bottomless and ordinary eaters, who do have broadly the same visual cues, and who we have assumed to significantly diverge in their levels of feeling of hunger.

A second objection promotes an alternative interpretation of the data unaddressed by Wansink et al., one that denies my crucial assumption of divergent levels of felt hunger. When Wansink et al. suggest that "the amount of food on a plate… lessens one's reliance on self-monitoring" (2005: p. 93), or Wansink writes that "we believe our eyes not our stomach" (2006: ch. 2), the suggestion is that we assess how hungry we feel without introspecting. An important alternative is that our stomach believes our eyes. On this line of thought, the misleading external cues received by bottomless

eaters cause them to feel much less full than they otherwise would, and indeed equally full as the ordinary eaters. Here the bottomless eaters' judgments about how full they feel could still based on introspection of "internal cues" rather than external cues. The internal cues would simply have varied in response to external cues.

The problem can pop up in many cases where one tries to experimentally produce an error in judgment about our own minds. Perhaps in manipulating the conditions that lead to judgments about our mental states, you have manipulated the mental states those judgments are about. For example, Dutton and Aron (1974) notoriously claim to get people to mistakenly think they are attracted to someone, as a result of misinterpreting their excitement after crossing a scary bridge as attraction to an interviewer. However, perhaps the rush of crossing a scary bridge makes one more likely to be attracted to someone on the other side.[13]

While I cannot definitively rule out the alternative interpretation just sketched of Wansink's experiment, it is less plausible than the more orthodox interpretation of the Wansink experiment. The crucial point is the dramatic scale of the effect in the Wansink experiment, where it takes 73% more food to reach roughly the same assessed feeling of fullness. On one natural development of the alternative interpretation, how full the bottomless eater feels could be mediated by their beliefs about how much they ate, in a case of cognitive penetration. The problem here is that, even in what are arguably some of the best candidates for being cases of cognitive penetration (of vision), the scales of the effect are fairly small. To pick just one commonly discussed example, Hansen et al. (2006) asked subjects to adjust an image of a banana until it is achromatic, and they at most moderately overadjusted into the range of blue.[14] Putative cases of cognitive penetration thus do not provide a strong enough precedent for the objector. On another way to go, how full the bottomless eater feels could be mediated by their other senses, as in cases of cross-modal effects on perception. Here again I think the challenge is to find suitably close precedents with a suitably scaled effect. While I cannot rule out that such precedents can be found, the objector owes us some, and we do not have them yet.[15]

Overall, given the extraordinary excess consumed by the bottomless eaters, we end up with empirical evidence that we are imperfectly reliable about when we introspect about how hungry we feel.

3 The epistemic implications of cases of error

We have now seen potential cases of error and even unreliability (in a certain domain) about when we introspect. Let's now see whether they can be used to build successful skeptical arguments. Here I will look most closely at templates for skeptical arguments we can already find used by figures such as Sinnott-Armstrong, Doris, or Carruthers.

[13] For discussion of a similar problematic over the interpretation of cognitive dissonance experiments, see Fiala and Nichols (2009).

[14] For a (skeptical) survey of potential cases of cognitive penetration, see Firestone and Scholl (2016).

[15] For some recent reflections on how to understand cross-modal interactions in perception, potentially as perception that is somehow multi-modal, see the essays in Part II of Stokes et al. (2014). For a sample recent discussion focusing on (images of) food, see Spence et al. (2016).

The simplest route is perhaps the following:

Argument from Error
(Case of Error): You sometimes falsely believe that you introspectively believe that p.
(Generalization 1): If you sometimes falsely believe that you introspectively believe that p, then you never know that you introspectively believe that p.
So,
You never know that you introspectively believe that p.

The standard problem here is that this argument moves too quickly from fallibility to ignorance. Falllibility might well entail the absence of certainty, but we need a much more extensive argument to get to the absence of knowledge. As things stand, we have no reason to deny you can gain knowledge from a newspaper that an earthquake has occurred, even if that newspaper has sometimes led you to false beliefs about the spellings of names.[16]
It is much more promising to hold that knowledge is incompatible with the widespread actualization of error. So a better skeptical strategy might focus on the potential prevalence of errors about introspection (see e.g. Scaife 2014, also the similar skeptical arguments against moral intuitionism of Sinnott-Armstrong 2006):

Argument from Widespread Error
(Widespread Error): You are not reliable when you believe that you introspectively believe that p.
(Generalization 2): If you are not reliable when you believe that you introspectively believe that p, then you never know that you introspectively believe that p.
So,
You never know that you introspectively believe that p.

This argument relies on a substantially stronger starting claim about how often we are mistaken about when we introspect. Indeed, too strong a starting claim—in order to reach skepticism about our access to when we introspect, the argument has assumed a hefty dose of non-skepticism about how much we have learned from the challenging studies. While we have seen surprising evidence from Wansink and Rolls in favor of a very restricted version of Widespread Error, the jury remains out on whether the bold, across-the-board claim made by Widespread Error is true. Most importantly, the two most promising ways of trying to pick up the burden of proof fail.
The best bet of the skeptic here is probably to adapt work by Doris (2015), himself focused on issues about reflective agency, and lean on the wide range of studies of confabulation about our own minds.[17] But there are two major reasons why leaning on studies of confabulation will not help.

First, even if confabulating subjects come to a false conclusion about their own minds, it's perfectly possible that they still came to that conclusion through intro-

[16] For a survey of how to articulate and how to evaluate fallibilism, see Fantl and McGrath (2009).

[17] For overviews of the studies, see Scaife (2014) section 2, Doris (2015: ch. 4), or Bortolotti (2018). Note that, while Schwitzgebel argues that introspection is not reliable, he still seems to assume that its function and standard for success is to be understood in terms of accuracy (2008: pp. 265–6). This assumption is importantly challenged in Doris (2015: ch. 4).

spection. Making a mistake does not entail that you failed to introspect. To think otherwise is to be misled by the term "introspection", where that term unfortunately suggests a perceptual relation to a mental state that is really there. But we shouldn't let a misleading word push us towards a perceptual model of introspection.

Second, even if confabulating subjects come to a conclusion about their own mind through a broadly inferential or interpretative route, it's still entirely possible that they came to that conclusion through introspection. To assume otherwise is to use a non-inferential, non-interpretative requirement for introspection that is too theoretically loaded. There are paradigm cases of introspection in which we interpret or make inferences about our own minds. For example, you might come to an introspective judgment that you hope for a birthday call by imagining not getting one, realizing that you would feel disappointed in that situation, and coming to the conclusion that you hope for a call. There is a step involving interpretation or inference here, since the data point that you would feel disappointed in the scenario of not getting a call is not identical to the conclusion that you hope for a call. While this process is broadly inferential or interpretative, it remains a paradigmatic case of introspection. (On the negative conception of introspection used here, if you infer from an introspective judgment that you would feel disappointed to a judgment that you hoped for a call, your judgment that you hoped for a call will inevitably be introspective, since no one could reach that conclusion in the way you did, since no one could reach your key premise in the way you did).[18]

The skeptic could alternatively lean on the work of Schwitzgebel (2008, 2011), frequently cited as having established the unreliability of introspection (e.g. by Srinivasan 2015). But there are many difficulties here.[19]

First, if Schwitzgebel (2012)'s own between-cases form of pluralism about introspection turns out to be correct, with very different processes coming under the heading of "introspection" in different cases, it becomes much harder to establish that various species of introspection are not reliable.[20] Schwitzgebel's earlier critique is targeted at "the primary method" we use to form beliefs about our own minds, but if his later work is correct there is no such method, and the critique will need to pick through multiple methods in their plurality. That method-by-method critique has yet to be made.

Second, our skeptic's focus is on a second-order question, our reliability about when we introspectively believe we feel angry, or introspectively believe we have richly detailed peripheral vision. Schwitzgebel's focus in his (2008, 2011) is on a first-order question, our reliability more directly about when we feel angry, or when we have richly detailed peripheral vision. Even if we weren't reliable about being in such first-order mental states, that leaves open whether we are still reliable about when we introspectively judge we are in them. Indeed, in order to find examples where we are introspectively mistaken about what first-order mental states we are in, Schwitzgebel must work with cases in which we do make introspective judgments in the first place about our first-order mental states (as we reviewed in Sect. 1). But then we won't be at

[18] For further discussion of such examples of introspection, see Lawlor (2009) or Cassam (2014: ch. 11).

[19] Here I try to present some novel problems for appealing to Schwitzgebel. For further challenges (that I largely endorse) to Schwitzgebel's critique, see Bayne and Spener (2010), Smithies (2013), or Bayne (2014).

[20] Thanks to Carolina Flores for this point.

risk in his cases of mistakenly thinking that we are introspecting. Also, Schwitzgebel describes his central cases without hesitation as ones of introspecting, and given that he does not mention third-person evidence in his descriptions of them, presumably uses introspection to reach the conclusion that he is introspecting.[21]

Third, the definition of "unreliability" used in Schwitzgebel's critical work is quite different from the one needed by the skeptic. In the sense used in Schwitzgebel's critique,

> There are two kinds of unreliability. Something might be unreliable because it often goes wrong or yields the wrong result, or it might be unreliable because it fails to do anything or yield any result at all (2008: p. 265).

Call the second kind the broad sense of "unreliability". We have been using a narrow sense in which only the mistakes put out by a process or method are relevant to its reliability, and where the silences of a process or method are irrelevant to its reliability. Given the focus on necessary conditions for knowledge, the skeptic needs to use the narrow sense of reliability. Consider: if my taciturn neighbor answers virtually none of my questions, my taciturn neighbour is "unreliable" in the broad sense. However, if my neighbour is very likely to say something true on those rare occasions when he does say something, he remains reliable in the narrow sense and a perfectly good candidate to supply knowledge. Since our skeptic needs to establish unreliability in the narrow sense, Schwitzgebel's work on the broad sense of unreliability won't directly be a useful crutch.

In sum, the skeptic's most two most promising strategies fail to establish that we are unreliable (in the narrow sense) about when we introspect.

As a further problem for the skeptic, Generalization 2 is shaky as well. Even if it turns out that we are not reliable on some general level of description about when we introspect, the possibility still remain of knowing that we introspect when we are in good or optimal conditions for introspection. Compare: even if counterfeiting of $100 bills is rife, so that we are never reliable enough to know that we are holding a $100 bill, we might still be perfectly able to know that we are holding a penny. Even if we are not in general reliable about when we introspect about how hungry we feel, we could still be in a perfectly good position to know that we have introspected when we feel absolutely starving or when we feel absolutely stuffed.

The arguments so far rely on claims about actual error that are either too weak to support a skeptical conclusion, or so strong that they merit skepticism themselves. A more promising strategy is to somehow leverage the possibility of being in error about introspecting. For an example of such a strategy in another area, consider John Doris' initial statement of his own skeptical challenge to reflective theories of agency, arguing that we do not have good epistemic access to whether we engage in reflective agency:

[21] Consider the following passage:

My wife mentions that I seem to be angry about being stuck with the dishes again (despite the fact that doing the dishes makes me happy?). I deny it. I reflect; I sincerely attempt to discover whether I'm angry—I don't just reflexively defend myself but try to be the good self-psychologist my wife would like me to be—and still I don't see it. I don't think I'm angry. But I'm wrong, of course, as I usually am in such situations: My wife reads my face better than I introspect (2008: p. 252, see also p. 255).

> A skeptical hypothesis is one that cannot be ruled out, and would falsify some belief, or category of beliefs, if true…. The present skeptical hypothesis maintains that for any putative exercise of agency, one cannot rule out a defeater (or defeaters) in the explanation of that behavior. Where one cannot rule out this alternative, one cannot justifiably posit an instance of morally responsible agency. Therefore, one is never justified in positing an instance of morally responsible agency (2015: 65, see also 64).

In other words, since we cannot rule out certain skeptical hypotheses in which people fail to exercise reflective agency, we can never determine that anyone exercises reflective agency. I will export the strategy to the case of introspection, and I will focus on a subject's own on knowledge of whether they introspect. The idea is to identify specific scientific/skeptical hypotheses incompatible with a subject introspecting, and then to exploit the subject's lack of knowledge that they are false. To ultimately reach a skeptical conclusion, the following template implicit in the quote from Doris is a start:

Argument from Open Possibility of Error
(Must Rule Out): If SK is incompatible with S introspectively believing that p, then S knows S introspectively believes that p only if S knows that SK is false.
(Ignorance): SK is incompatible with S's introspectively believing that p, and S does not know that SK is false.
So,
S doesn't know that S introspectively believes that p.

An initial hurdle is that the Must Rule Out requirement is far too strong. Consider how any necessary falsehood is incompatible with every proposition. It seems absurd to require that, for you to know any proposition that P, for every necessary falsehood that F, you must know it's not the case that F. Your inability to rule out some skeptical hypotheses can be irrelevant to ruling in that you are introspecting.

While Doris in effect starts with Must Rule Out in the quote, he goes on to restrict the claim to skeptical hypotheses that have a non-trivial probability of being true, and that are also of practical relevance (2015: pp. 65–67). To see why the adjusted requirement remains too strong, consider the vast range of cases where complicated scientific hypotheses are incompatible with ordinary claims. For example, consider the ordinary claim that the chewing gum does not taste like pineapple, made by someone chemically uninformed who is allergic to pineapple. Now consider a complicated scientific hypothesis incompatible with the ordinary claim, that the chewing gum contains a lot of allyl hexanoate (a compound found in pineapples used to create pineapple flavoring). The scientific/skeptical hypothesis is of practical relevance given the person's allergies, and we can suppose that the hypothesis also has a non-trivial probability of being true. Since our character has never heard of the scientific hypothesis, and doesn't even have the concepts required to entertain it, our character by no means knows that the hypothesis is false. But our character still seems perfectly able to know that the chewing gum does not taste like pineapple. Doris' restriction of Must Rule Out is still too demanding.

A standard skeptical fix here (not discussed by Doris) is to appeal to known incompatibility rather than merely to incompatibility:

Argument from Known Open Possibility of Error:
(Must Rule Out2): If you know that SK is incompatible with your introspectively believing that p, then you know you introspectively believe that p only if you know that SK is false.
(Mixed Ignorance): You know that SK is incompatible with your introspectively believing that p, and you don't know that SK is false.
So,
You don't know that you introspectively believe that p.

Now we have a skeptical argument that again relies on a crucial dose of non-skepticism in Mixed Ignorance. The reliance leads the argument into trouble. That's because at most experts are aware of the subtle scientific/skeptical hypotheses that threaten to be incompatible with our introspecting, and consequently the argument challenges only the knowledge of experts. As far as the vast range of non-experts is concerned, they end up being saved from the skeptical argument by their own ignorance (I am unclear on whether Doris' own argument about reflective agency is meant to go beyond the case of experts' knowledge).[22]

One challenge (for each variant of the current strategy) is to find specific hypotheses that permit the template to be run. As I argued above, even if you have a false belief about your mind, you could still be introspecting, and even if you are confabulating about your own mind, you could still be introspecting. It is not so easy to find a skeptical hypothesis that is genuinely incompatible with our introspecting.

In any case, even if the argument does manage to apply to experts in some cases, it is unclear whether the now restricted conclusion will be surprising. It is not so unlikely that, once you bite from the apple of the *Journal of Obesity Research*, and become apprised of evidence about your inaccuracy about when you introspect, your knowledge of when you introspect ends up being undermined. We would at most have a case of defeat of the defeasible knowledge that non-experts still get to possess.[23]

So far the skeptic has used the templates of classic skeptical arguments without seeing any surprising success, and while seeing new hurdles generated by the use of empirically based scenarios of error.

A more promising strategy is to shift the target of the skeptical argument, aiming for a conclusion that might be more easily reached from cases of error. For example, in Carruthers' own discussion of closely related issues, he acknowledges that "as philosophers will know, there are numerous strategies for replying to such arguments [e.g. skeptical arguments concerning knowledge] (2011: p. 43)." He pivots as follows:

> … the split-brain data seem to show decisively that we don't have any *subjectively accessible warrant* for believing that we ever have transparent access to our own attitudes. This is because patients can report plainly confabulated explanations

[22] For further discussion of challenges for formulating skeptical arguments in terms of known incompatibility, see Blome-Tillmann (2006) and David and Warfield (2008).

[23] Although see Lasonen-Aarnio (2010, 2014) or Baker-Hytch and Benton (2015) for challenges to standard assumptions about how knowledge can get undermined.

> with all of the same sense of obviousness and immediacy as ordinary people (2011: p. 43, emphasis mine).

Carruthers is focusing on what it is like on the inside for split-brain patients when they seem to themselves to access their attitudes, and in particular on what he assumes to be the felt "obviousness and immediacy" of their self-ascriptions. The apparent fact that ordinary subjects share the same felt "obviousness and immediacy" of their own self-ascriptions is meant to entail that ordinary subjects don't have any "subjectively accessible warrant" for their self-ascriptions.

To further elucidate the strategy he pursues in this passage, we should clarify Carruthers' own terms and goals—they are not quite the same as ours.

First, when Carruthers speaks of "transparent" access to our attitudes, he has in mind a form of access that is not only "introspective" in our sense from page 3 of peculiarity. Carruthers has in mind a form of access to our attitudes that is also non-interpretive. Our own focus is on introspective access that might well also be interpretive in some way. Here "transparent access" entail introspective access but not vice versa. Now, Carruthers' main claim is about a hurdle for our access to whether we have transparent access to our attitudes, and he is silent about our access to whether we "introspect" in the sense used here. (While Carruthers' target is narrower than our own, the problems I go on to discuss will still arise for his own project.)

Second, while Carruthers does not define "subjectively accessible warrant", I take him to mean any reason to believe that p such that you are able to know that you have that reason to believe that p (more soon on an alternate reading of his term).

Finally, Carruthers' primary aim in his work is to argue directly for the conclusion that we do not transparently access our attitudes—never mind what sort of (misleading) reason we might have to believe we sometimes do transparently access our attitudes. He develops his primary line of argument by delineating the predictions made by his own theory and its rivals, and accounting for how his own theory allegedly has the best fit with the overall empirical evidence (this discussion is largely spread across chapters 5–12 of his 2011).[24] My own focus will remain on whether we have ever have reason to believe we transparently/introspectively access our attitudes (this argument in his 2011 book is only in his chapter 5).

Adapting the template of Carruthers' argument away from the specific case of transparency to the more general case of "introspection" in our sense, the new strategy is this:

> **New Argument from Error**
> (Case of Error): Confabulating split-brain patients seem to themselves to have introspective access to their mental states when they don't.
> (Generalization 3): If Case of Error is true, then none of us have subjectively accessible warrant to believe that we have introspective access to our mental states.
> So,

[24] For critical discussion of those arguments, see Goldman (2006), Fiala and Nichols (2009), Rey (2013), or Andreotta (2019).

(Qualified Skepticism): None of us have subjectively accessible warrant to believe that we have introspective access to our mental states.

Some problems arise right away from the example chosen. Confabulating split-brain patients need not be mistaken if they think they introspect, even if they would be mistaken in thinking they have transparent access to their mental states. As we saw above, confabulation is compatible with introspection.

Now, it might well be that split-brain patients do fail to introspect. We actually do not need to take a stand here on this question. The New Argument from Error in any case fails because of its reliance on the dubious Generalization 3.

One challenge arises because it is unclear whether split-brain patients are mentally similar enough to ordinary subjects to bear on the epistemic standing of ordinary subjects—consider for instance the notorious inability of split-brain patients to verbalize what is shown in the left of their visual field. These cognitive differences remain whether or not split-brain subjects indeed are the same as ordinary subjects in the apparent "obviousness and immediacy" of their self-ascriptions of their attitudes. These cognitive differences block an immediate inference from the errors of split-brain patients to an epistemic threat to ordinary subjects. Compare: if we learn that there are poorly executed fake barns around, that manage to fool only subjects who are cognitively impaired, we are not yet able to conclude that there is an epistemic hurdle for subjects who are not cognitively impaired, regardless of any shared sense of obviousness in everybody's judgments about when a barn is in the field. This challenge to Generalization 3 arises whether the skeptic targets introspection or only transparency.

Second, even if split-brain patients did have mental lives that are overall sufficiently similar to those of ordinary subjects, there is too long a road from their being in error to an epistemic threat to our having subjectively accessible warrant, and indeed even to their having subjectively accessible warrant. When someone has reason to believe that p, and knows that they have reason to believe that p, it could still fail to be the case that p. For example, when the wall looks red to me in apparently good conditions, I presumably have reason to believe that the wall is red, and know that I do, regardless of whether the wall really is red. In the error scenario where the wall looks red to me but is not red, but all else seems to be going well, I presumably still have (misleading) reason to believe that the wall is red while knowing that I have reason to believe that the wall is red. In particular, notice that such a fallibilist view should be compelling to you if you allow that suggestive but inconclusive experiments can give rational support for hypotheses, as the empirically-minded Carruthers presumably does. As a result, both the split-brain subject potentially in error and the ordinary introspecting subject could easily still have subjectively accessible reasons to believe they are introspecting, simply reasons that fail to guarantee truth. Again, these would be reasons potentially of the same strength as reasons we obtain from strongly suggestive but inconclusive scientific experiments.

In pursuing the new skeptical strategy, we seem to have fallen back into the trap of assuming that fallibility leads directly to a skeptical conclusion. This problem undermines the skeptical strategy regardless of whether it targets our access to introspection in general or our access to transparent access in particular.

One way to respond would be to emphasize the quantity of errors made by split-brain subjects and others—I take myself already to have addressed this strategy in my discussion of the Argument from Widespread Error.

In what I take to be the most promising available line of response, the opponent could say that when they speak of "subjectively accessible warrants", they have a more demanding standard in mind than I have considered so far. On a more stringent use of the expression, when you have a subjectively accessible *warrant* to believe that p, your warrant is such that you have it only if it is the case that p. Such warrants are conclusive in the sense that they guarantee truth. (I will assume they are subjectively accessible in the sense that you are able to know both that you have them and that they are conclusive).

It is in one way easy to get from cases of error to the absence of a conclusive warrant—it is by definition impossible to have a conclusive warrant to believe that p in a case where you falsely believe that p. So the more demanding reading of "subjectively accessible warrant" would make it easy to get to a skeptical conclusion at least about a subject who is in error. But the skeptical point so far holds only for the subject who is in the case of error, and the status remains open of the subject who is not in the case of error. The introspecting subject could easily still have a subjectively accessible conclusive warrant even though her potentially deceived counterpart does not. Compare: the subject who genuinely sees that the wall is red might have a subjectively accessible conclusive warrant consisting of her seeing that the wall is red, even though her deceived counterpart does not. The subject in the "good case" would have a warrant consisting of a genuine perceptual relation to something like the fact that the wall is red, whereas the subject in the "bad case" would at best have a warrant consisting of some sort of appearance that does not guarantee that the wall is red (views with this structure are defended by McDowell 2008, Pritchard 2012, or Schellenberg 2016). This sort of point is especially plausible when we reconsider the cognitive differences between split-brain patients and ordinary subjects. As compared to the perceptual case, I take it to be less clear whether the split-brain patients and ordinary subjects even are overall the same "from the inside" in their perspectives on the world.[25]

The skeptic might insist that, even if the subject in the good case had a conclusive warrant, she would lack the ability to know that they have a conclusive warrant—the skeptic insists that there is no *subjectively accessible* conclusive warrant in the bad case or the good case. But the denial of subjective accessibility is its own skeptical thesis, in need of its own skeptical argument. Even if the subject in the bad case is unable to know something on the basis of the appearances, it does not automatically follow that the subject in the good case is unable to know the same sort of thing on the basis of appearances—that question is just the sort of thing we need a good skeptical argument to adjudicate, one that has not yet been produced.

[25] For views with a similar structure in the case of introspection, inspired by the perceptual case, see Hellie (2006) or Macpherson (2010). The perceptual flavor is optional however. "Constitutivists" such as Shoemaker (2009) or Smithies (2012) thoroughly reject perceptual models of introspection, but still hold that you in general have introspective reason to believe that you are in a mental state only if you are in that mental state.

Also, if we interpret the skeptical challenge in terms of the more demanding epistemic notion, the conclusion becomes dramatically weaker. The new conclusion would be that we have no subjectively accessible *conclusive* reason to believe that we introspect or have transparent access to our minds. The new conclusion would leave open the important possibility that we have a subjectively accessible non-conclusive reason to believe that we introspect or have transparent access to our minds. In other words, a reason of the same quality as the sorts of reasons we can get from reflection on scientific studies.

Each of the evil demon problems we have discussed tries to move from a case of error to a skeptical conclusion. None of them lines up an error hypothesis that gets us to a skeptical conclusion with bite.

4 Conclusion

When Descartes raised his hypothesis of a deceiving evil demon, he worried about radical error about such elementary questions as whether squares have four sides, and yet strikingly did not seem to worry about radical error about his own mental states. I have opposed Descartes to some extent by arguing for novel scenarios of error and radical error about how we access our own minds. My work still has remained in the spirit of Descartes: I have also argued that we ultimately remain safe from evil demon problems for our access to how we access our own minds.

One major lesson is that we should not over-estimate what it takes to introspect. Contrary to what you might assume, introspection need not be accurate, or non-inferential, or exclusive of perception, or exclusive of confabulation. Another major lesson is that, while skeptical challenges by figures such as Carruthers, Doris, and Schwitzgebel are rich and empirically informed, these skeptical challenges founder on how they are epistemologically under-informed.

Acknowledgements For their help with this paper, I'd like to thank Jack Barnett, Alex Byrne, Ophelia Deroy, Carolina Flores, Carl Ginet, Anna Giustina, Anil Gomes, Patrick Greenough, Lisa Miracchi, Ram Neta, Shaun Nichols, Adam Pautz, Adriana Renero, Barbara Rolls, Eric Schwitzgebel, Sydney Shoemaker, Susanna Siegel, Declan Smithies, Hannah Trees, Ru Ye, Jonna Vance, Timothy Williamson, and several anonymous referees. I'm also grateful to audiences at workshops or other events at the University of Geneva, the Institut Jean-Nicod, the Ohio State University, Oxford University, Bled, CSU Chico, Nanyang Technological University, and the University of Bergen.

References

Andreotta, A. (2019). Confabulation does not undermine introspection for propositional attitudes. *Synthese*. https://doi.org/10.1007/s11229-019-02373-9.

Baker-Hytch, M., & Benton, M. A. (2015). Defeatism defeated. *Philosophical Perspectives, 29*(1), 40–66.

Bayne, T. (2014). *Introspective insecurity. Open MIND*. Frankfurt am Main: MIND Group.

Bayne, T., & Spener, M. (2010). Introspective humility. *Philosophical Issues, 20,* 1–22.

Benton, D. (2015). Portion size: what we know and what we need to know. *Critical Reviews in Food Science and Nutrition, 55*(7), 988–1004.

Benton, M. A. (2017). Epistemology personalized. *The Philosophical Quarterly, 67,* 813–834.

Blome-Tillmann, M. (2006). A closer look at closure scepticism. In *Proceedings of the aristotelian society* (vol. 106, no. 1, pp. 383–392). Oxford: Oxford University Press.
Bortolotti, L. (2018). Stranger than fiction: Costs and benefits of everyday confabulation. *Review of Philosophy and Psychology, 9*(2), 227–249.
Boyle, M. (2009). Two kinds of self-knowledge. *Philosophy and Phenomenological Research, 78*(1), 133–164.
Burger, K. S., Fisher, J. O., & Johnson, S. L. (2011). Mechanisms behind the portion size effect: Visibility and bite size. *Obesity, 19*(3), 546–551.
Byrne, A. (2005). Introspection. *Philosophical Topics, 33,* 79–104.
Carruthers, P. (2011). *The opacity of mind: an integrative theory of self-knowledge*. Oxford: OUP.
Cassam, Q. (2014). *Self-knowledge for humans*. Oxford: Oxford University Press.
Chalmers, D. (2003). The content and epistemology of phenomenal belief. *Consciousness: New Philosophical Perspectives, 220,* 271.
David, M., & Warfield, T. A. (2008). Knowledge-closure and skepticism. In Q. Smith (Ed.), *Epistemology: New essays*. Oxford: Oxford University Press.
Dennett, D. C. (1988). Quining qualia. In A. J. Marcel & E. Bisiach (Eds.), *Consciousness in modern science*. Oxford: Oxford University Press.
Doris, J. M. (2015). *Talking to our selves: Reflection, ignorance, and agency*. OUP Oxford.
Dretske, F. (2012). Awareness and authority: Skeptical doubts about self-knowledge. *Introspection and Consciousness,* 49–64.
Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology, 30*(4), 510.
Engelbert, M., & Carruthers, P. (2010). Introspection. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(2), 245–253.
Fantl, J., & McGrath, M. (2009). *Knowledge in an uncertain world*. Oxford: Oxford University Press.
Fiala, B., & Nichols, S. (2009). Confabulation, confidence, and introspection (Commentary on Peter Carruthers). *Behavioral and Brain Sciences, 32,* 144–145.
Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences,* 39, E229. https://doi.org/10.1017/S0140525X15000965.
Gertler, B. (2012). Renewed acquaintance. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness* (pp. 89–123). Oxford University Press.
Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*(1), 1.
Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience, 9*(11), 1367–1368.
Hellie, B. (2006). Beyond phenomenal naivete. *Philosophers' Imprint, 6,* 1–24.
Hill, C. S. (1991). *Sensations: A defense of type materialism*. Cambridge: Cambridge University Press.
Horgan, T., & Kriegel, U. (2007). Phenomenal epistemology: What is consciousness that we may know it so well? *Philosophical Issues, 17*(1), 123–144.
Lasonen-Aarnio, M. (2010). Unreasonable Knowledge. *Philosophical Perspectives, 24*(1), 1–21.
Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research, 88*(2), 314–345.
Lawlor, K. (2009). Knowing what one wants. *Philosophy and Phenomenological Research, 79*(1), 47–75.
Macpherson, F. (2010). A disjunctive theory of introspection: a reflection on zombies and Anton's syndrome. *Philosophical Issues, 20*(1), 226–265.
McDowell, J. (2008). The disjunctive conception of experience as material for a transcendental argument. In F. Macpherson & A. Haddock (Eds.), *Disjunctivism: Perception, action, knowledge*. Oxford University Press.
Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences, 37*(1), 1–19.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231–259.
Osterholt, K. M., Roe, L. S., & Rolls, B. J. (2007). Incorporation of air into a snack food reduces energy intake. *Appetite, 48*(3), 351–358.
Prinz, J. (2004). The fractionation of introspection. *Journal of Consciousness Studies, 11*(7–8), 40–57.
Pritchard, D. (2012). *Epistemological disjunctivism*. Oxford University Press.

Pudel, V. E., & Oetting, M. (1977). Eating in the laboratory: Behavioural aspects of the positive energy balance. *International Journal of Obesity, 1*(4), 369–386.
Rey, G. (2013). We are not all 'self-blind': A defense of a modest introspectionism. *Mind & Language, 28*(3), 259–285.
Reyes, G., & Sackur, J. (2017). Introspective access to implicit shifts of attention. *Consciousness and Cognition, 48,* 11–20.
Reyes, G., & Sackur, J. (2018). Introspection during short-term memory scanning. *Quarterly Journal of Experimental Psychology, 71,* 2088–2100.
Rolls, B. J., Bell, E. A., & Waugh, B. A. (2000). Increasing the volume of a food by incorporating air affects satiety in men. *The American journal of clinical nutrition, 72*(2), 361–368.
Rolls, B. J., Roe, L. S., Meengs, J. S., & Wall, D. E. (2004). Increasing the portion size of a sandwich increases energy intake. *Journal of the American Dietetic Association, 104*(3), 367–372.
Samoilova, K. (2016). Transparency and introspective unification. *Synthese, 193*(10), 3363–3381.
Scaife, R. (2014). A problem for self-knowledge: The implications of taking confabulation seriously. *Acta Analytica, 29*(4), 469–485.
Schellenberg, S. (2016). Phenomenal evidence and factive evidence. *Philosophical Studies, 173*(4), 875–896.
Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review, 117*(2), 245–273.
Schwitzgebel, E. (2011). *Perplexities of consciousness*. Cambridge: MIT Press.
Schwitzgebel, E. (2012). Introspection, what. In D. Smithies & D. Stoljar (Eds.), *Introspection and Consciousness* (pp. 29–48). Oxford: Oxford University Press.
Shoemaker, S. (2009). Self-intimation and second order belief. *Erkenntnis, 71*(1), 35–51.
Siewert, C. (2012). On the phenomenology of introspection. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness* (pp. 129–168). Oxford: Oxford University Press.
Sinnott-Armstrong, W. (2006) Moral intuitionism meets empirical psychology in *Metaethics after Moore* (ed.) Horgan, T. Oxford University Press.
Smithies, D. (2012). A simple theory of introspection. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness*. Oxford: Oxford University Press.
Smithies, D. (2013). On the unreliability of introspection. *Philosophical Studies, 165*(3), 1177–1186.
Smithies, D., & Stoljar, D. (Eds.). (2012). Introduction. In *Introspection and consciousness*. Oxford: Oxford University Press.
Spence, C., Okajima, K., Cheok, A. D., Petit, O., & Michel, C. (2016). Eating with our eyes: From visual hunger to digital satiation. *Brain and Cognition, 110,* 53–63.
Spener, M. (2012). Mind-independence and visual phenomenology. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness*. Oxford: Oxford University Press.
Spener, M. (2015). Calibrating introspection. *Philosophical Issues, 25*(1), 300–321.
Srinivasan, A. (2015). Are we luminous? *Philosophy and Phenomenological Research, 90*(2), 294–319.
Stokes, D., Matthen, M., & Biggs, S. (2014). *Perception and its modalities*. Oxford: Oxford University Press.
Stump, E. (2010). *Wandering in darkness: Narrative and the problem of suffering*. Oxford: Oxford University Press.
Talbert, B. M. (2015). Knowing other people: A second-person framework. *Ratio, 28*(2), 190–206.
Talbert, B. (2017). Overthinking and other minds: The analysis paralysis. *Social Epistemology, 31*(6), 545–556.
van der Zee, T., Anaya, J., & Brown, N. J. (2017). Statistical heartburn: An attempt to digest four pizza publications from the Cornell Food and Brand Lab. *BMC Nutrition, 3*(1), 54.
Wadhera, D., & Capaldi-Phillips, E. D. (2014). A review of visual cues associated with food on food acceptance and consumption. *Eating Behaviors, 15*(1), 132–143.
Wansink, B. (2006). *Mindless eating: Why we eat more than we think*. New York: Bantam.
Wansink, B., Painter, J. E., & North, J. (2005). Bottomless bowls: Why visual cues of portion size may influence intake. *Obesity, 13*(1), 93–100.
Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
Zlatevska, N., Dubelaar, C., & Holden, S. S. (2014). Sizing up the effect of portion size on consumption: A meta-analytic review. *Journal of Marketing, 78*(3), 140–154.