



External world scepticism and self scepticism

Joshua Rowan Thorpe^{1,2}

Accepted: 11 December 2022

© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

A general trend in recent philosophical and empirical work aims to undermine various traditional claims regarding the distinctive nature of self-knowledge. So far, however, this work has not seriously threatened the Cartesian claim that (at least some) self-knowledge is immune to the sort of sceptical problem that seems to afflict our knowledge of the external world. In this paper I carry this trend further by arguing that the Cartesian claim is false. This is done by showing that a familiar sceptical argument that targets my knowledge of the external world can be adapted to target my belief that I exist, along with any of my self-knowledge that I know entails my own existence. Thus, my self-knowledge and my knowledge of the external world are subject to the same sort of sceptical problem.

Cartesian sceptical arguments involving dreams and evil demons generate the problem of external world scepticism. However, Descartes doesn't think that any sceptical argument threatens my self-knowledge, and so he doesn't think that there is a corresponding problem of what we might call self scepticism. While we may not agree with Descartes that *all* self-knowledge is immune to sceptical worry, his view may seem plausible when we restrict our attention to paradigmatic instances of self-knowledge, such as my knowledge that:

I am thinking.
I am thinking that water is wet.
I am in pain.
I am having a yellow visual experience.

✉ Joshua Rowan Thorpe
joshthop@gmail.com

¹ University of Glasgow, Glasgow, Scotland

² School of Humanities, 1 University Gardens G12 8QQ Glasgow, UK

This weakened version of Descartes' view is captured by what I will call the:

Cartesian claim: Some significant class of self-knowledge is immune to the sort of sceptical problem that seems to afflict our knowledge of the external world.¹

If true, the Cartesian claim marks an important difference between self-knowledge, and knowledge of the external world. And it seems that the Cartesian claim is widely endorsed (at least tacitly), even by philosophers who reject other aspects of Descartes epistemology and metaphysics of mind. Historically, there is much discussion of the problem of external world scepticism, but very little discussion of any problem of self scepticism. Presumably this is because it is thought that there is a real and pressing problem of external world scepticism, but no equivalent problem of self scepticism.

One might think that although the Cartesian claim has enjoyed widespread endorsement in the past, we now have good reason to reject it. Much recent philosophical and empirical work aims to undermine various traditional claims regarding the distinctive nature of self-knowledge, and it may seem that this work undermines the Cartesian claim. So far, however, it has not.

One strand in this work aims to reject the idea that we have particularly good epistemic access to the facts about our own mental states. For example, it aims to show that we are not infallible regarding our mental states,² or that our mental states are not luminous in the sense that when we are in a particular mental state we are always in a position to know that we are in it.³ However, the claim that we do not have particularly *good* (infallible or luminous) epistemic access to our mental states does not entail that we don't know anything about them, and so it does not raise a sceptical problem regarding self-knowledge. Another strand suggests that our beliefs about our own mental states are prone to errors of a sort to which we might have thought them to be immune.⁴ But the claim that a certain class of belief is prone to a certain sort of error does not entail that beliefs of this sort are never knowledge, and so this strand does not suggest that there is a sceptical problem regarding self knowledge either.

It has also been suggested that externalism about mental content threatens the Cartesian claim. Roughly, the thought is that since externalist mental states supervene on environmental facts, my justification for believing that I am in a particular externalist mental state will include empirical beliefs about my environment. Thus, my knowledge that I am in a particular externalist mental state is empirical, and so it is threatened by any scepticism that threatens my empirical knowledge of the external world.⁵

¹ For the purposes of this paper, we need not worry about how the advocate of the Cartesian claim should define this class.

² Shoemaker (1994).

³ Williamson (2002).

⁴ Nisbett & Wilson (1977) present empirical work suggesting that subjects routinely misidentify the factors that influence their reasoning process. Schwitzgebel (2008) argues that we are very unreliable when it comes to introspecting our own mental states, and it might be thought that when we are very unreliable on a topic we don't know (much) about it. However, most of his argument does not focus on the paradigmatic cases of self-knowledge with which the Cartesian claim is concerned.

⁵ See Boghossian (1989).

It is not clear to what extent this line of thought would threaten the Cartesian claim if it were plausible. The claim remains interesting even if it is understood as saying that there is an interesting class of self-knowledge of *non-externalist* mental states that does not give rise to a sceptical problem. More importantly, there is widespread consensus that it is a mistake to think that the fact that externalist mental states supervene on my environment entails that my knowledge that I am in a particular externalist mental state is empirical. Without this implication, there is no reason to think that scepticism that threatens my empirical knowledge will threaten my knowledge that I am in a particular externalist mental state.⁶

Thus, notwithstanding work that aims to undermine the idea that self-knowledge is distinctive in various ways, it may seem that the Cartesian claim does indeed capture an important distinction between self-knowledge and knowledge of the external world. As we shall see, however, the Cartesian claim is false after all. A familiar closure-based sceptical argument that targets my contingent beliefs about the external world can be adapted to target my belief that I exist, and thus to target any of my self-knowledge that I know entails my existence. The question of how to respond to this argument is the problem of self scepticism, which is analogous to the problem of external world scepticism.

1 External world scepticism and self scepticism

In this section I rehearse a familiar closure argument for external world scepticism, and explain how this can be adapted into an argument for self scepticism. The aim of this section is to get the argument for self scepticism on the table, and to show that it can be given the same initial motivations as those that drive the argument for external world scepticism. The rest of the paper deals with potential objections to this argument.

The external world sceptic starts by presenting me with a scenario in which I and my thoughts – my intentional and non-intentional mental states – remain the same, while the external world is drastically altered. Here is a version of the scenario described by Descartes in the first Meditation:⁷

The External World Sceptical scenario (EWS scenario): An evil demon destroys the external world, including my body. However, the demon ensures that I and my thoughts carry on exactly as they would have done had the demon not interfered with the external world.

To illustrate the EWS scenario, suppose that if the demon had not interfered with the world it would have rained today, and that as a result I would have had certain visual/olfactory/tactile/auditory experiences, and the belief that it is raining. In the EWS scenario there is no rain, but the demon ensures that I nevertheless have the same

⁶ See Gertler (2015), Falvey & Owens (1994) and Bar-On (2004).

⁷ Descartes *Meditations on First Philosophy*, Cottingham's (1986) translation.

rain-associated visual/olfactory/tactile/auditory experiences, and the belief that it is raining.

The sceptic employs the EWS scenario in the following closure-based sceptical argument:

- (E1) If I know that I have hands, I know that the EWS scenario is false.
- (E2) I don't know that the EWS scenario is false.
- (E3) I don't know that I have hands.

Call this the External World Sceptical argument (EWS argument).

The first premise of the EWS argument is supported by appeal to the idea that knowledge is closed under known implication. This idea is captured by principles such as the following:

Closure: for any subject *S*, and propositions *p*, *q*: if *S* knows that *p*, and *S* knows that *p* is incompatible with *q*, then *S* knows that not *q*.⁸

Closure, together with the fact that I know that my belief that I have hands is incompatible with the EWS scenario, entails (E1).

The sceptic can motivate (E2) by challenging me to provide a reason to prefer the hypothesis that I am not in the EWS scenario over the hypothesis that I am not in the EWS scenario. It will be useful later to have gone over some of the reasons that this challenge is so difficult to meet.

I can be seen as facing a dilemma when I attempt to respond to this challenge. On what we might call the empirical horn of the dilemma, I might try to offer empirical reasons to think that the EWS scenario is false. However, on the assumption that my experiences are my empirical evidence, my empirical evidence is, by stipulation, compatible with the EWS scenario. There are two ways of construing this idea, depending on how we think of sensory evidence. First, we might think of my sensory evidence as propositional. If we think of my sensory evidence in this way, the idea is that the propositions that are my sensory evidence are logically compatible with the EWS scenario. Second, we might think of my sensory evidence as a collection of non-propositional things. If we think of my sensory evidence in this way, the idea is that all these things could exist in the EWS scenario.⁹

On what we may call the a priori horn of the dilemma, I might try to give a purely a priori proof that the EWS scenario is false.¹⁰ On the face of it, however, the prospects

⁸This principle may require refinement; see for example Hawthorne (2014). However, this refinement is unlikely to affect the sceptical argument in any significant way.

⁹If I accept that my empirical evidence is compatible with the EWS scenario, I might still attempt to employ inference to the best explanation to show that my sensory evidence gives me reason to think that the EWS scenario is false. For example, I might try to reject the EWS scenario on the grounds that it is an ad hoc explanation of my experiences. However, it is difficult to formulate plausible principles for inference to the best explanation that support this response. See Beebe (2009) for discussion of this strategy.

¹⁰It is hard to find examples of people who have tried to reject the sceptical argument by arguing that the EWS scenario is logically or conceptually incoherent. Perhaps Descartes' argument that a necessarily existing God could not allow him to be radically deceived is an example.

for such a proof seem dim. The EWS scenario looks like a logical possibility, so how could I have an a priori reason to think that it is false?

Suppose that the EWS argument works. What are the consequences? The EWS argument concludes that I do not know that I have hands, but there is nothing special about my belief that I have hands. If the EWS argument works, versions of it can be employed to show that any belief of mine that I know to be incompatible with the EWS scenario is not knowledge. These include most of my contingent beliefs about the external world. However, there are certain contingent beliefs that I do not know to be incompatible with the EWS scenario, for example, my belief that I exist. (In fact, I know that the EWS scenario entails that I do exist.) So the EWS argument doesn't threaten this belief.

However, I will now suggest that a variation on the EWS scenario can be employed in a closure-based sceptical argument that does threaten my belief that I exist. Assume that 'I' is a referring term. For convenience, we will call the referent of 'I' as employed by me *myself*. When other people use the word 'I', they also refer, and for convenience we will call the things that they refer to *selves*. Assume nothing more about what selves are for the moment. Now consider the following sceptical scenario:

The Self Scepticism scenario (SS scenario): A demon destroys the external world, as in the EWS scenario. However, the demon doesn't stop there. It also destroys all selves, including myself and itself. Before doing so, the demon ensures that all thought carries on exactly as it would have done had it not destroyed the external world along with all the selves.

To get a grip on the SS scenario, suppose again that if the demon had not interfered with the world it would have rained today, and that as a result I would have had certain visual/olfactory/tactile/auditory experiences, and the belief that it is raining. In the SS scenario there is no rain, and I (i.e. myself) do not exist. However, the demon has ensured that the rain-associated visual/olfactory/tactile/auditory experiences occur, and that the belief that it is raining occurs.

Is the SS scenario possible, or even intelligible? One immediate reaction is that to get the SS scenario off the ground we must assume that selves are something more than the thinking activity that is usually attributed to them. For suppose that selves were identical with the thinking activity that is usually attributed to them. In that case, it would be impossible to 'destroy all selves' while ensuring that 'all thought carries on', since whatever selves exist are *identical* with certain stretches of the thinking activity that is being preserved. Let us make this not-improbable assumption for now. (In § 2 I will argue that the sceptic need only make a much weaker assumption.) Even on this assumption, some readers will immediately balk at the idea that the SS scenario is possible, or intelligible. Later in this section I will address the idea that the SS scenario is just *obviously* or *self-evidently* impossible, and later in the paper I will consider some attempts to argue that the SS scenario is impossible.

Let us return to the argument for self scepticism. The self sceptic employs the SS scenario in the following argument:

- (S1) If I know that I exist, then I know that the SS scenario is false.
 (S2) I don't know that the SS scenario is false.
 (S3) I don't know that I exist.

Call this the Self Scepticism argument (SS argument). Any self-knowledge that I know to be incompatible with the SS scenario is threatened by the SS argument, just as any empirical knowledge that is incompatible with the EWS scenario is threatened by the EWS argument. In particular, any self-knowledge that is existentially committed to myself is threatened by the SS argument, since any such knowledge is incompatible with the SS scenario. This includes the paradigmatic cases of self-knowledge noted at the beginning of this paper that are thought to be immune to scepticism, if any are: e.g. my knowledge that I am thinking, that I am thinking that water is wet, that I am in pain, that I am having a yellow visual experience, etc.¹¹

The sceptic's defence of the premises of the SS argument is analogous to her defence of the premises of the EWS argument. I know that the SS scenario is incompatible with my belief that I exist, so Closure entails (S1). No issues arise here that do not arise with the motivation for (E1).

The sceptic may motivate (S2) by challenging me to give a reason for thinking that (S2) is false. Again, I can be seen as facing a dilemma when I attempt to respond to this challenge. Because we are less familiar with it, this dilemma requires more explanation than the dilemma that arises when I attempt to rule out the EWS scenario. Both dilemmas are, however, fundamentally the same.

On the empirical horn of the dilemma, I try to give an empirical reason for thinking that the SS scenario is false. However, on the assumption that my sensory experiences are my sensory evidence, my sensory evidence is, by stipulation, compatible with the SS scenario. So my sensory evidence alone does not rule out the SS scenario. Again, there are two ways of interpreting this idea, depending on how we think about sensory evidence.

First, we might think of my sensory evidence as propositional. If we think of my sensory evidence in this way, the idea is that the propositions that are my sensory evidence are logically compatible with the SS scenario. If the propositions in question are impersonal propositions about my sensory experiences (e.g. *there is a yellow sensory experience*) they will be logically compatible with the SS scenario. I think that this is the right way to think about sensory evidence. Indeed, I think that this is one of the lessons that anyone who thinks that sensory evidence is propositional should take from Hume's insight that when I attend to my sensory experiences I do not attend to

¹¹ On the classic expressivist view sometimes associated with Wittgenstein, the utterance 'I am in pain' is akin to the utterance 'ouch', in that both express the feeling of pain, and do not express a proposition. On this view, the utterance 'I am in pain' is not incompatible with the SS scenario, and it might be thought that one could avoid my argument against the Cartesian claim by being this sort of expressivist. This is not a good strategy for two reasons. First, classical expressivism is widely seen as implausible and has few adherents. Neo-expressivist views of the sort endorsed by Bar-On (2019) are becoming more popular, but on these views 'I am in pain' does express a proposition, and thus is threatened by the SS argument. Second, since the classic expressivist view says that 'I am in pain' does not express a proposition they are committed to the claim that this expression does not express knowledge, and therefore does not express self-knowledge. Classic expressivism therefore offers no hope of vindicating the Cartesian claim by describing a class of self-knowledge that is immune to scepticism.

myself.¹² Another possibility is that the propositions that make up my sensory experience are (or that some are) first personal propositions about my sensory experiences (e.g. *I am having a yellow experience*). In that case, my sensory evidence is not logically compatible with the SS scenario. Despite this, thinking of my sensory evidence as first personal propositions about my sensory experiences will not allow me to rule out the SS scenario. However, a defence of this claim will have to wait until §3, where we will see that when attempting to respond to self scepticism I am not entitled to assume both that the concept ‘I’ refers and that any of the propositions that I am entertaining contains the concept ‘I’.

Second, we might think of my sensory evidence as a set of non-propositional things. If we think of my sensory evidence in this way, the idea is that these things could exist in the SS scenario. If we think of these things as objects that don’t include me as a part (e.g. a yellow sensory experience) then they will be able to exist in the SS scenario. I don’t think of sensory evidence as a set of non-propositional things, but for those who do Hume’s insight that when I attend to my experiences I do not attend to myself is a good reason to think of sensory evidence in this way.¹³

On the a priori horn of the dilemma, I try to dismiss the SS scenario on a priori grounds. But what would these grounds be? One suggestion is that the very definition of ‘thought’ ensures that it is a conceptual and therefore a priori knowable truth that thinking cannot take place without a self. When the SS scenario was introduced, thoughts were said to be *mental* states, and isn’t a mental state a state of the mind of a thinker, and therefore a self?¹⁴ So isn’t the idea of thoughts – i.e. states of the mind of a thinker/self – occurring without a self incoherent after all?

If this reasoning is correct, the only result is that we must refine our description of the SS scenario. Call expressions of content and experiences that are not states of the mind of a self thoughts*. Then, rather than saying that in the SS scenario there are thoughts but no selves, we will say that in the SS scenario there are thoughts* but no selves.

Is there any a priori reason to think that the SS scenario so redescribed is impossible? One possible reaction is that it is just *obvious* or *self-evident* that thoughts* are impossible and/or unintelligible, such that no further argument for this claim is needed. Analogously, one might think that it is just obvious that a square triangle is impossible. No argument or justification for this claim is required (even if it is possible to give one), and only a deeply confused or irrational person could fail to recognise this. If the SS scenario is obviously impossible or unintelligible in this sense, (S2) can never be motivated.

However, the claim that the SS scenario is obviously impossible in this sense is deeply implausible. Philosophers have held views on which there are thoughts*

¹² More precisely, I do not attend *directly* to myself; see Peacock (2012).

¹³ We noted in fn. 9 that one might try to rule out the EWS scenario by appealing to one’s sensory evidence and inference to the best explanation. There is no reason to think that this strategy is any more likely to succeed in ruling out the SS scenario.

¹⁴ I am happy to assume here that thinkers are selves, although I will suggest that the sceptic need say very little about the nature of the self in §2.

(e.g. Lichtenberg).¹⁵ So have certain religions (e.g. Buddhism).¹⁶ It may be that on close inspection these views will turn out to be necessarily false. However, the claim that these views are *obviously* impossible seems completely incredible. Minimally, a defence of this claim would have to include an explanation of how so many apparently intelligent people are actually as confused and irrational as people who believe that square triangles are possible. (The same response can be made to the even less plausible claim that the SS scenario is obviously unintelligible.)

We may conclude that there is no “quick and easy” way of ruling out the SS scenario on empirical or a priori grounds, just as there is no “quick and easy” way of ruling out the EWS scenario on empirical or a priori grounds. This is sufficient for *prima facie* motivation for (S2). Many readers will no doubt suspect that there it is possible to rule out SS via some more involved a priori reasoning. Versions of such reasoning will be considered in the remaining sections of this paper.

For now, we have seen that the SS argument and the EWS argument have the same structure, and the same initial motivations. Both arguments generate unacceptable sceptical conclusions. This suggests that, insofar as there is a problem of external world scepticism, there is also a problem of self scepticism. The rest of this paper aims to vindicate this suggestion by considering and rejecting various potential responses to the SS argument.

Note that, if they work, certain potential solutions to the problem of external world scepticism would apply equally to the problem of self scepticism. For example, someone who rejects (E2) on the grounds that my belief that the EWS scenario is false is safe and therefore knowledge is likely to reject (S2) on the grounds that my belief that the SS scenario is false is safe and therefore knowledge.¹⁷ Such solutions will not be considered, because they do not threaten the main claim of this paper, which is that the Cartesian claim is false. The Cartesian claim says that some self-knowledge is immune to the sort of sceptical problem that seems to afflict knowledge of the external world. The falsity of this claim is compatible with the view that the problem of external world scepticism and the problem of self scepticism have the same solution. In fact, the view that they have the same solution only strengthens the impression of parity between the two sceptical problems.

2 Assumptions about the self¹⁸

This section examines the question of what assumptions the sceptic must make about the self. It may seem that the sceptic is committed to the denial of the bundle theory of the self, or to the denial of a certain physicalist account of the self. I argue that the sceptic need only reject a particular version of the bundle theory that is anyway implausible, and that the sceptic is not committed to the denial of the physicalist

¹⁵ See Tester (2013) for this interpretation of Lichtenberg.

¹⁶ See Collins (1982) and Kalupahana (1987) for the prevailing ‘no-self’ interpretation of Buddhist doctrine.

¹⁷ See Sosa (1999) for this response to external world scepticism.

¹⁸ My thanks to an anonymous referee for encouraging me to address the issues raised in this section

account of the self. I conclude that the sceptic is committed to no particular positive account of the self.

When the SS scenario was first introduced we noted that the possibility of this scenario depends on the assumption that there is more to a self than the thinking activity that is usually attributed to it. If selves are identical with stretches of thinking activity then it is impossible that all selves are destroyed while ‘all thought carries on’, as described in the SS scenario. But doesn’t this assumption at least weaken the dialectical appeal of the SS argument? The sceptic has said nothing to defend it, and, moreover, there are philosophers who reject it, claiming that selves are identical with thinking activity.

Let us start with the point that there are philosophers who say that the self is identical with thinking activity. These philosophers endorse the bundle theory of the self, according to which selves are identical with bundles of thoughts.¹⁹ It is true that on the bundle theory of the self the SS scenario is impossible: selves cannot be destroyed while the bundles of thoughts that they are identical to are preserved. So isn’t the sceptic committed to the substantial assumption that the bundle theory of the self is false?

I do not think so. Although the SS argument as it stands may be committed to this assumption, it can be revised to assume much less. One of the principle questions for bundle theorists is, what relationship must a certain set of thoughts have to each other to count as a self? Different versions of bundle theory give different answers to this question, but answers divide broadly into those that say that the relevant thoughts must have appropriate causal relations, those that say that the relevant thoughts must have appropriate psychological relations (e.g. coherence), and those that say that the relevant thoughts must have both appropriate causal and psychological relations. For our purposes, the precise details of this relationship do not matter. The important point is that according to bundle theory thoughts that stand in the relevant relationship to each other count as a self. This allows distinct thoughts occurring at a particular time to be part of and belong to the same self. It also allows thoughts occurring at different times to be part of and belong to the same self. This second point allows the bundle theorist to identify selves with temporally extended bundles of thoughts, allowing the bundle theorist to say that selves persist through time.

The following revised version of the SS scenario is possible according to most versions of bundle theory that include this requirement of appropriate relations between thoughts:

Churning Self Sceptical scenario (CSS scenario): the demon destroys all selves, including itself, and the external world. Before doing so the demon ensures that after this destruction the only thing that will exist is a churning mess of thinking activity where thoughts occur at random. Purely by chance, it happens that for a moment in this churning mess exactly the thoughts that I would have had at this moment had the demon not interfered occur. After this moment, the churning of random thoughts continues.

¹⁹This view is usually associated with Hume (1975). See Campbell (1967) and Pike (2006) for more recent defences.

Does a self exist in the CSS scenario according to the bundle theorist? If, as on most versions of bundle theory, there is a requirement of a particular kind of causal connection, the answer is no, because the relevant causal connections do not obtain. Even if the only requirement is that certain psychological connections obtain it is still likely that the answer is no. The thoughts that occur at the relevant moment may have some of the relevant psychological connections to each other. But they cannot bear many of the psychological connections to each other that my current thoughts actually have, because most of these connections are *historical*: for example my thoughts are coherent with each other *now*, largely due to the fact that they come at the end of a narrative consisting of other thoughts that serves to explain why I have the thoughts that I currently do. True, it would be possible to abandon any mention of these historical psychological connections and endorse a version of bundle theory on which the thoughts that momentarily occur in the CSS scenario are sufficiently psychologically connected to each other to form a self. But this would be a niche view, and I'm not sure that anyone endorses it. If the sceptic must assume that this niche view is false, this is not an implausible assumption.

Can the CSS scenario be used in the SS argument? I know that my belief that I exist is incompatible with this scenario, so by closure (S1) is true. In addition, there is no reason to think that I can rule out the CSS scenario by appeal to my sensory evidence. The sensory experiences that I have in the relevant moment of the CSS scenario are the same as those that I actually have, including my thoughts and impressions about what happened in the past. If this is all the evidence that I have when attempting to rule out the CSS scenario, I will not get very far.²⁰ And there seems to be no more a priori reason to think that the CSS scenario is false than there is to think that the SS scenario is false.

A last-ditch attempt to save this response to the SS argument would be to simply *stipulate* a meaning for 'I' such that I may be identical with the thinking activity that occurs at the relevant moment in the CSS scenario. So long as we define the self in this way, the CSS scenario is impossible, insofar as it also says that no self would exist.

There is little to be gained by this desperate strategy. Let us write 'I*' to express the newly stipulated meaning. Perhaps it is true that 'I* exist' is immune to scepticism. But so long as we have no reason to believe that the meaning of 'I*' is the same as the meaning of 'I' as it occurs in belief that I am in pain, we have no reason to think that this belief is immune to scepticism. More generally, so long as we have no reason to think that the meaning of 'I*' is the same as the meaning of 'I' as it occurs in ordinary instances of self-knowledge we will have no reason to think that self-knowledge is immune to scepticism. But to claim that this is the right interpretation of 'I' in normal instances of self-knowledge is effectively to claim that implausible niche version of bundle theory is true.

Does the sceptic have any further commitments regarding the nature of the self, beyond the claim that the niche version of bundle theory is false? One thought is that it might be incompatible with certain physicalist views of the self. Plausibly,

²⁰The challenge here is similar to Russel's (1995) challenge to explain how I know that the world did not pop into existence in its entirety five minutes ago.

my thoughts have a physical supervenience base. The physicalist view of the self in question says that selves are identical with (parts of) these supervenience bases. For example, such a view might say that my thoughts supervene on my brain states, and that I am my brain. Doesn't this view entail that the SS/CSS scenario is impossible, since it describes a situation where my thoughts exist, and yet I, and therefore the supervenience base of my thoughts, do not exist? So, isn't the sceptic committed to the assumption that it is false?

No, for two reasons. First, both physicalism and claims about the supervenience base of thought are usually taken to be empirical, the main support of such theories coming from our best science. If these views are empirical, and if they entail that the SS/CSS scenario is impossible, this will be an empirical impossibility, like, e.g. the empirical impossibility that water has the chemical structure XYZ rather than H²O.²¹ Since the impossibility of the SS/CSS scenario is empirical, it is question begging to assume it in the face of a sceptical argument that threatens my empirical knowledge along with my self-knowledge. Analogously, it would be question begging to assume that water cannot be XYZ in the face of external world scepticism, since knowledge of this impossibility rests on the empirical knowledge that the sceptic threatens. The second reason is that physicalists usually present physicalism as contingently true, rather than necessarily true.²² So it is not clear that physicalism *does* entail that the SS/CSS scenario is impossible.

Moreover, whatever one thinks of these reasons, there is no disparity between the CSS/SS scenario and the EWS scenario here. Physicalism is incompatible with the EWS scenario, insofar as this scenario describes the demon destroying the external world, including my body (and thus my brain), while preserving me and my thoughts. This is not usually thought to pose a problem for EWS scepticism, but if it is the sceptic can reply as above by pointing to the empirical and contingent nature of physicalism. These moves are as plausible for the external world sceptic as they are for the self sceptic, and so there is no disparity between the EWS argument and the SS argument here.

Need the self sceptic give any positive account of the self, beyond the negative claim that the niche version of bundle theory described above is false? I do not think so. Various popular accounts of the self are compatible with the argument for self-scepticism. The sceptic need make no assumption about which of these accounts is true.

3 The cogito and the self-reference rule for 'I'

In this section, I examine a Cartesian argument that my belief that I exist is immune to sceptical doubt, and show that it is of no use as a response to the SS argument. One might think that the Cartesian argument can be saved by appeal to the popular idea that 'I' is governed by the self-reference rule. However, this will not work either.

²¹ Loar (1997) holds that physicalism is necessarily true, and knowable only by empirical means.

²² See for example Lewis (1983).

In the *Discourse* Descartes recounts arriving at the conclusion that his belief that ‘*I am thinking, therefore I exist*’ was so firm and secure that all the most extravagant suppositions of the sceptics were not capable of overthrowing it.²³ Why think that this belief – the cogito – might be immune to closure-based sceptical arguments? As Williams (2005, p. 75) points out, both parts of the cogito – ‘I am thinking’ and ‘I exist’ – are *incorrigible*, in the sense that ‘if anyone believes that he is thinking, or again, that he exists, then necessarily he has a true belief’.²⁴ The fact that my belief that I exist is incorrigible in this sense may suggest a response to the SS argument along the following lines: I can recognise that my belief that I exist is incorrigible, and this recognition is sufficient, without begging any questions against the sceptic, for me to know that this belief is true. Since I know that I exist, I know that the SS scenario is false, so (S2) is false.

Without getting into the details of this argument, it is clear that if we are already taking the SS scenario seriously, we cannot use it to rule it out. Williams is surely correct in saying that my belief that I exist is incorrigible in the sense that if *anyone* believes that he exists, then what he believes is true. However, to assume that there is someone who is doing the believing while attempting to rule out the SS scenario is simply to beg the question, in the same way that I would beg the question if I were to assume that I have hands in the course of an argument that I am not in the EWS scenario.

The fact that my belief that I exist is incorrigible in Williams’ sense does not on its own suffice to explain why this belief is immune to closure scepticism. However, it might be thought that the following widely endorsed rule can be combined with the incorrigibility of this belief to make the explanation work:

Self-Reference Rule: the concept ‘I’ always refers to the thinker of the thought in which it occurs.²⁵

If the Self-Reference Rule is true, then all beliefs involving ‘I’, and in particular my belief that I exist, belong to a thinker.

Someone who endorses the Self-Reference Rule might try to revive the response to the SS argument considered above by arguing along the following lines: I recognise that my belief that I exist is incorrigible, so if this belief is thought by someone it is true. That this belief includes ‘I’, together with the Self-Reference Rule entails that this belief is thought by someone. So, this belief is true. Running through this argument establishes that I exist without begging the question against the self-sceptic, and thus allows me to rule out the SS scenario, falsifying (S2).

It is not clear why I would be entitled to assume that the Self-Reference Rule is true when attempting to rule out the SS scenario. It may have seemed plausible

²³ Descartes, in P. J. Olscamp’s (1965) translation, Part iv: VI 32, HRI 101.

²⁴ Stroud (2011) makes the same point as Williams. Unfortunately, neither Williams nor Stroud explain why the incorrigibility of a claim would render it immune to closure scepticism, so the following explanation of why this might be thought to be so is extrapolation on my part.

²⁵ See for example O’Brien (1994). There is some discussion of how best to formulate the Self-Reference Rule, but the precise formulation doesn’t matter here.

that 'I' *always* refers to the thinker of the thought in which it occurs when we were assuming that thoughts always have thinkers. However, consideration of the SS scenario brings this assumption into question. Without it, it is not clear what entitles me to claim that 'I' always refers to the thinker of the thought in which it occurs, as opposed to the more cautious claim that 'I' is a device for referring to the thinker of the thought in which it occurs, *if indeed there is such a thinker*.

Some might resist these reasons for thinking that if I am worried about the SS scenario I cannot assume that the Self-Reference Rule is true. One way to do this would be to say that it is an integral part of our understanding of the 'I' concept that it is governed by this rule, so once we are entertaining the possibility that the concept is not governed by this rule we are no longer talking about the 'I' concept.²⁶ Another would be to simply stipulate that when we are talking about the concept 'I', we are talking about a concept that is governed by the Self-Reference Rule.

However, these moves will not help. In the course of attempting to rule out the SS scenario, I cannot assume both that the Self-Reference Rule is true, and that a thought that genuinely involves 'I' is occurring. If the Self-Reference Rule is true, then thoughts that involve 'I' cannot occur in the SS scenario. However, thoughts that do not involve 'I' can occur in the SS scenario (or at least, we've been given no reason to think that they can't). If I am taking the SS scenario seriously, why should I be entitled to assume that when I utter the words 'I exist' is a belief that involves the concept 'I'?

Here is an analogy. Suppose that I have the singular thought that this bottle is blue. There are views of singular thought on which it is a necessary condition of my having this singular thought that there actually be a bottle in front of me.²⁷ On these views, it follows from the fact that I am having the singular thought in question that there is a bottle in front of me. Opinions diverge on what happens if there turns out to be no bottle: perhaps I will have a thought with a different logical form, or perhaps there will be no genuine thought at all, but only the appearance of a thought.

Those who hold this view of singular thought do not typically think that I can start with only my introspective knowledge of what I am thinking and deduce that there is a bottle in front of me, because they do not think that I can know by introspection that I am having the singular thought that this bottle is blue. Those who think that if there is no bottle then I will have a thought of a different logical form may say that I know by introspection that I am thinking a thought about a blue bottle, but that I do not know whether this is a *singular* thought. Those who think that if there is no bottle there will be no thought at all may have to say that I cannot know by introspection that there is any genuine thought at all.

Suppose that I know that the Self-Reference Rule is true. Even so, it does not follow that I will be able to deduce from my introspective knowledge of what I am thinking that I am not in the SS scenario, for the same reasons that the view of singular thought that we have been considering does not entail that I can deduce from my introspective knowledge of what I am thinking that there is a bottle in front of me. Perhaps the SS scenario is true, and so the thought that I express by saying 'I exist' is

²⁶ See for example Kaplan (1989).

²⁷ See for example Evans (1989) and McDowell (1994).

not a thought that genuinely involves ‘I’. Or, perhaps the SS scenario is true, and so although the appearance of a thought might prompt me to say ‘I exist’, in fact there is no thought at all. I can rule these possibilities out only if I have some prior entitlement to assume that the SS scenario is false.

In §1 I said that even if my sensory evidence includes first-person propositions about my sensory experiences, this will not allow me to rule out the SS scenario by appealing to my sensory experiences. My reasons for saying this should now be clear. Even if my sensory evidence does in fact consist of first-person propositions about my experiences, it is not obvious that I am entitled to the assumption that the concept ‘I’ that occurs in these propositions refers when attempting to rule out the SS scenario. However, I am certainly not entitled both to the assumption that it refers, and to the assumption that the propositions that make up my sensory evidence include the genuine ‘I’ concept. And yet, without both assumptions I cannot appeal to the propositions that make up my sensory evidence to rule out the SS scenario.

4 Is the SS argument self-defeating?

So far, we have considered attempts to show that the SS argument is unsound. However, one might think that there is no need to consider the *soundness* of the argument, because the SS argument is somehow self-defeating, and so can never generate a sceptical problem. In this section we will see that the SS argument is not self-defeating in any way that would prevent it from generating a sceptical problem.

One might think that the SS argument is self-defeating in the sense that *the conclusion is not something that I can rationally believe*. Here are two reasons that one might think this. First, I surely cannot reasonably endorse a claim that I know entails the existence of some entity, while recognizing that I do not know that the entity in question exists. But wouldn’t I have to do exactly that to endorse (S3), which says that I don’t know that I exist? For this claim (i) entails my existence, and (ii) says that I do not know that I exist. Another reason is that, plausibly, to claim to know anything is to claim knowledge *for oneself*.²⁸ But I cannot claim knowledge of (S3) for myself. If I accept (S3) I accept that I don’t exist, and if I don’t know that I exist I cannot meaningfully claim anything for myself.

One might also think that the SS argument is self-defeating in the sense that *it has no stable upshot*.²⁹ One reason for thinking this is, again, that I cannot reasonably endorse a claim while denying knowledge of an existential commitment of that claim. Premise (S2) says that I don’t know that the SS scenario is false, and thus appears to entail my existence. Perhaps I can endorse this claim while working through the SS argument, but upon completing the argument I must conclude that I don’t know that I exist. But if I do, won’t I lose the ability to believe (S2), which was part of my grounds for this very conclusion? Another reason is that, again, it is plausible that to claim to know something is to claim that knowledge for oneself. Having accepted (S3), I surely cannot claim knowledge of (S2) for myself. If either of these reasons

²⁸ I am grateful to an anonymous referee for this point.

²⁹ I am grateful to the same anonymous referee for pushing me to address this point.

is correct, the result is that the SS argument cannot have a stable upshot, in the sense that although I may find the SS argument rationally compelling so long as I believe that I exist, as soon as I take the final step and conclude that I don't know that I exist, I lose my grounds for this conclusion. Of course, there is then nothing to prevent me from endorsing my belief that I exist again, and again working through the SS argument until I reach the conclusion that I don't know that I exist...

I have grouped these two ways in which the SS argument might be self-defeating because my response to both reasons is the same. If I cannot rationally endorse (S3), or if I cannot work through the SS argument to reach a stable rational belief in (S3), this will not prevent the SS argument from generating a sceptical problem. Sceptical problems or paradoxes arise because a sceptical argument appears to generate pressure to endorse an intuitively unacceptable sceptical conclusion. If it turns out that I cannot rationally endorse that conclusion, or that the argument cannot provide a stable motivation for this conclusion, this only underlines the fact that the conclusion is unacceptable. It doesn't do anything to relieve the pressure generated by the sceptical argument to endorse that conclusion. Sceptical problems are solved when this pressure is relieved by an explanation of what is wrong with the sceptical argument, not by additional reason to think that the conclusion of the argument is unacceptable.

The same is true of other philosophical problems, or paradoxes. For example, the conclusion of the liar paradox cannot reasonably be believed, since it is a logical contradiction. To note this is not to solve the problem presented by the liar. Such a solution must take the form of an explanation of what is wrong with the apparently compelling argument that seems to lead to this conclusion. Indeed, it is sometimes argued that we cannot rationally endorse the conclusion of the EWS argument. As Wright (1985) points out, this would not solve the problem of external world scepticism for the same reason.³⁰

Another possibility is that one of the *premises* of the SS argument is self-defeating, in the sense that I cannot rationally endorse that premise, even when considered in isolation from the other premise. If we knew that one of the premises of the argument was self-defeating in this sense, we would have an explanation of what is wrong with the SS argument, and so we would have a solution to the sceptical problem that it seems to generate. I see no prospect that (S1) is self-defeating in this sense. However, one might wonder whether I could ever coherently endorse (S2), which says that I don't know that the SS scenario is false. I know the SS scenario to be incompatible with my own existence. So, won't I be endorsing a claim that commits me to my own existence, whilst also saying that I don't know that I exist? Surely this is something that I cannot do.

It may be that I cannot. But this is not what I will be doing insofar as I am endorsing (S2) *independently* of (S1). When I endorse (S2), I endorse the claim that I don't know that the SS scenario is false, which, we may suppose, commits me to the belief that I exist. We may also suppose that I believe that the SS scenario is incompatible with my existence. But I won't see anything problematic about this unless I *also* believe (S1), which says that if I know that I exist, then I know that the SS scenario is

³⁰ See also Beebe (2011) for a similar point regarding sceptical arguments intended to threaten my a priori knowledge.

false.³¹ If I endorse both (S2) and (S1) I effectively run the SS argument, which puts pressure on me to endorse the unacceptable sceptical conclusion that I do not know that I exist, thus generating the problem of self-scepticism.

5 Conclusion

Just as the EWS argument threatens my contingent knowledge of the external world, the SS argument threatens my self-knowledge. The first argument generates the problem of external world scepticism, while the second argument generates the problem of self scepticism. One might think that there are solutions to the problem of self scepticism that do not apply to the problem of external world scepticism. I have considered several of what I take to be the most promising such solutions, and argued that they do not work. By doing so, I hope to have made a convincing case that the Cartesian claim is false: insofar as there is a problem of external world scepticism, there is also a problem of self scepticism. We must give up on the idea that self-knowledge is distinctive in its immunity to sceptical attack.³²

References

- Bar-On, D. (2004). Externalism and Self-Knowledge: Content, Use, and expression. *Nouse*, 38(3), 430–455.
- Bar-On, D. (2019). Neo-Expressivism: (Self-)Knowledge, meaning, and Truth. *Royal Institute of Philosophy Supplement*, 86, 11–36.
- Beebe, J. R. (2009). The Abductivist reply to Scepticism. *Philosophy and Phenomenological Research*, LXXIX(3), 605–636.
- Beebe, J. R. (2011). A priori scepticism. *Philosophy and Phenomenological Research*, LXXXIII(3), 583–602.
- Boghossian, P. A. (1989). Content and self knowledge. *Philosophical Topics*, 17, 5–26.
- Campbell, S. (2006). The conception of a person as a Series of Mental events. *Philosophy and Phenomenological Research*, LXXIII(2), 339–358.
- Collins, S. (1982). *Selfless persons*. Cambridge: Cambridge University Press.
- Descartes, R. (1965). In P. J. Olscamp (Ed.), *Discourse on Method, Optics, geometry, and Meteorology*. Indianapolis: Bobbs-Merril.
- Descartes, R. (1986). In J. Cottingham (Ed.), *Meditations on First Philosophy*. Cambridge: Cambridge University Press.
- Evans, G. (1989). *The varieties of reference*. Oxford: Oxford University Press.
- Falvey, K., & Owens, J. (1994). Externalism, Self Knowledge, and Scepticism. *The Philosophical Review*, 103(1), 107–137.
- Gertler, B. (2015). Internalism, externalism and accessibilism. In S. C. Goldberg (Ed.), *Externalism, Self-Knowledge, and scepticism: New Essays*. Cambridge: Cambridge University Press.
- Hawthorne, J. (2014). The Case for Closure. In M. Steup, J. Turri, & E. Sosa (Eds.), *Contemporary Debates in Epistemology*. Oxford: Wiley-Blackwell.

³¹ Without this extra belief I am like Nozick (1981), who rejects Closure, and thus sees no incompatibility between my knowing that I have hands, knowing that if I have hands I am not a handless brain-in-a-vat, and my not knowing that I am not a handless brain-in-a-vat.

³² Thanks to Bernhard Salow, David Horst, Crispin Wright, Adrian Haddock, Peter Sullivan, and Giada Fratantonio for comments and helpful discussion. I'm grateful to the Arts and Humanities Research Council (Grant No. AH/W008424/1) for funding this research.

- Hume, D. (1975). *A Treatise of Human Nature*. In L. Selby-Bigge, *A Treatise of Human Nature* Oxford: Clarendon Press.
- Kalupahana, D. J. (1987). *The principles of buddhist psychology*. Albany: State University of New York Press.
- Kaplan, D. (1989). Demonstratives: an essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan*. Oxford: Oxford University Press.
- Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61, 343–377.
- Loar, B. (1997). Phenomenal States. In N. Block (Ed.), *The nature of consciousness: philosophical Debates*. Cambridge, MA: MIT Press.
- McDowell, J. (1994). *Mind and world*. Cambridge, Mas: Harvard University Press.
- Nisbett, R., & Wilson, T. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge: Cambridge University Press.
- O'Brien, L. F. (1994). Anscombe and the self-reference rule. *Analysis*, 54(4), 277–281.
- Peacocke, C. (2012). Subjects and consciousness. In A. Coliva (Ed.), *Self and self-knowledge*. Oxford: Oxford University Press.
- Pike, N. (1967). Hume's Bundle Theory of the self: a Limited Defence. *American Philosophical Quarterly*, 4(2), 159–165.
- Russel, B. (1995). *The analysis of mind*. Routledge.
- Schwitzegel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, 117(2), 245–273.
- Shoemaker, S. (1994). Self-knoweldge and 'inner sense'. *Philosophy and Phenomenological Research*, 54, 249–314.
- Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives*, 13, 141–153.
- Stroud, B. (2011). Our debt to Descartes. In B. Stroud (Ed.), *Philosophers past and Present: selected Essays*. Oxford: Oxford University Press.
- Tester, S. (2013). G.C. Lichtenberg on self-consciousness and personal identity. *Archiv Fur Geschichte der Philosophie*, 95(3), 336–395.
- Williams, B. (2005). *Descartes: the project of pure enquiry*. Oxford: Routledge.
- Williamson, T. (2002). *Knowledge and its limits*. Oxford: Oxford University Press.
- Wright, C. (1985). Facts and certainty. *Proceedings of the British Academy*, 71, 429–472.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.