

How both you and the brain-in-a-vat can know whether or not you are envatted

Author: Ofra Magidor, University of Oxford.

Magdalen College, Oxford, OX1 4AU, UK

ofra.magidor@philosophy.ox.ac.uk

Abstract

Epistemic Externalism offers one of the most prominent responses to the sceptical challenge. Externalism has commonly been interpreted (not least by externalists themselves) as postulating a crucial asymmetry between the actual world agent and their brain-in-a-vat (BIV) counterpart: while the actual agent is in a position to know she is not envatted, her BIV counterpart is not in a position to know that she is envatted, or in other words only the former is in a position to know whether or not she is envatted.

In this paper, I argue that there is in fact no such asymmetry: assuming epistemic externalism, both the actual world agent and their BIV counterpart are in a position to know whether or not they are envatted. After an introduction (I), I present the main argument (II). I examine to what extent the argument survives when one accepts additional externalist-friendly commitments: semantic externalism, a sensitivity condition on knowledge, and epistemic contextualism (III). Finally, I discuss the implications of my conclusion to a variety of debates in epistemology (IV).

I. Introduction

i. *The Sceptical Challenge and the Externalist Response.* Contrast Actual-Jill an ordinary flesh and blood person, with her counterpart, BIV-Jill, who is a handless brain-in-a-vat ('BIV') which is hooked to a computer feeding her brain with electrical stimuli that make things appear to her just as (or at least indistinguishably from how) they actually do. Suppose that Actual-Jill truly believes that she has hands, that she is not a BIV, and that there are trees nearby. The sceptical challenge is to explain how these beliefs can constitute knowledge. After all, claims the sceptic, mere true beliefs are insufficient for knowledge: one also needs to have good justification, warrant, or evidence for one's belief. But Actual-Jill's belief cannot have sufficiently strong justification, because if BIV-Jill has all the corresponding beliefs, these would have exactly the same justification as that of Actual-Jill, yet fail to constitute knowledge. And nor will it do to account for the epistemic difference between the two agents by appeal to the factivity of knowledge: after all, if there happen to be trees outside BIV-Jill's vat laboratory then her belief that there are trees nearby will also be true, but would nevertheless fail to constitute knowledge.

One of the most prominent responses (or family of responses) to the sceptical challenge is the *epistemic externalist* response. According to epistemic externalists there is a necessary condition on knowledge other than truth which Actual-Jill satisfies but BIV-Jill does not.

There are various variants of this strategy. According to *process reliabilism*, a necessary condition on knowledge is that one's belief be formed using a reliable process: a process that has a tendency to produce true beliefs in similar environments to those of the original belief.¹ Actual-Jill is able to form the belief that she has hands using such a process: perception. Moreover, once she uses this process to obtain knowledge that she has hands, she can come to know that she is not a BIV by applying a further reliable process: valid inference from known premises.² By contrast, BIV-Jill does not have access to these same processes. As she is not perceiving hands, she cannot form the belief that she has hands using perception, and as she does not know that she has hands and cannot infer that she is not a BIV using either deductive or inductive inference from this kind of premise.

Another variant of this strategy is the *safety-theoretic* view of knowledge, which maintains that a necessary condition on knowledge is that one's belief be safe: one could not have easily had a (relevantly similar) false belief.³ Actual-Jill's belief that she has hands is safe, because there are no nearby worlds where she falsely believes she has hands. Nearby worlds are either ones where she is perceiving her hands and (truly) believes that she has hands, or ones where she no longer has hands (e.g. because she has lost them in an accident), and fully realises that she lacks hands hence avoiding the (false) belief that she has hands. Moreover, her belief that she is not a BIV is also safe, simply because all nearby worlds are ones where (we are assuming) she still believes that she is not a BIV, and indeed she is not a BIV. But BIV-Jill's beliefs about the external world are not safe, even when they happen to be true. For example, even if she truly believes that there are trees nearby, she could have easily had a false belief, as the nature of the stimuli she is receiving are not sensitive to her environment: there are nearby worlds where

¹ The locus classicus for the reliabilist view is Goldman (1979), but this formulation follows that of Heller (1995): 504. As Heller notes, if one individuates processes in a sufficiently fine-grained way, one can drop reference to the environment.

² Admittedly, this latter process will not help Actual-Jill rule out slightly different sceptical hypotheses: for example, that she is a full-bodied person with hands but her brain is being fed electrical stimuli just as in the BIV scenario. However, process reliabilists can still account for how Actual-Jill is able to know that she's not in those scenarios: using the reliable process of perception she could obtain many pieces of knowledge of the form 'It appears sunny outside, and it is sunny outside', 'It appears that I have hands, and I do have hands', and so forth. And she can then use the reliable method of an inductive inference from known premises to conclude that her external world beliefs are formed via perception, and hence that she is not in a vat-like scenario. (For critical discussion of whether these kind of belief forming processes are legitimate see for example Bergman (2006), chapter 7 and Weisberg (2012).

³ See Sosa (1999), Williamson (2000), and Pritchard (2005).

there are no trees nearby, but she is nevertheless receiving stimuli of trees and thus forming the false belief that there are trees nearby.⁴

A third variant of the externalist strategy is the *externalist-evidentialist* view, which maintains that a necessary condition on knowledge is that one's belief is supported by one's evidence, but where the notion of evidence is construed so that Actual-Jill and BIV-Jill do not share the same evidence.⁵ There are a wide range of views concerning what kind of entities count as evidence, and what sort of relation counts as evidential support. I will not discuss them in detail but to get a feel: Actual-Jill's evidence might include the very state of perceiving she has hands or alternatively the proposition that she is perceiving hands. By contrast, BIV-Jill does not have the same evidence: at best she has a *sensation* as of having hands or the proposition that it *appears* to her as if she has hands. This leaves open the possibility that Actual-Jill's evidence provides sufficient support for knowledge, while BIV-Jill's evidence does not.

ii. *The Alleged Asymmetry.* The externalist response to the sceptical challenge is standardly understood as postulating an asymmetry between an agent in the actual world and their BIV counterpart. One kind of asymmetry is that Actual-Jill knows that she is not a BIV, while BIV-Jill does not know that she is not a BIV. This asymmetry, however, is not very significant: BIV-Jill's belief that she is not a BIV obviously does not constitute knowledge, since it is false. Moreover, if Actual-Jill happens to be a sceptically-inclined individual who falsely believes that she *is* a BIV, then Actual-Jill would lack knowledge as to whether or not she is a BIV.⁶ This does not seem to show anything very epistemically significant. Externalists, however, typically suppose that there is a much deeper asymmetry at stake.

Before pointing out this deeper asymmetry, we need some terminology. Following Chalmers (2005), let us say that an agent is *envatted* if they have a cognitive system which receives its inputs from, and sends outputs to, a computational simulation of a world. A BIV is one example of an agent that is envatted, but there are many other examples (for example, an agent who is

⁴ Of course we can imagine variants of the BIV-scenario where the stimuli BIV-Jill is receiving are very sensitive to the environment around her. But this is plausibly not a sceptical scenario at all – in that case, BIV-Jill arguably does know that there are trees nearby.

⁵ Williamson (2000) is in one sense an externalist evidentialist, though since he takes one's evidence to consist of the propositions one knows, his evidentialism is trivial (if you know that p, then p is part of one's evidence and hence trivially fully supported by your evidence). More interestingly, there are various externalist views of evidence which do not accept that all evidence is knowledge (see e.g. Schellenberg (2013), Mitova (2014), and Neta (*forthcoming*)). Combining any such view with the claim that knowledge requires evidence would yield a substantive externalist evidentialist view (see e.g. Schellenberg (*forthcoming*) for an explicit endorsement of this kind of view).

⁶ Similar points apply to having accidentally true beliefs.

a brain in a bottle, or a fully-bodied agent whose brain is hooked to electrical stimuli just as in the BIV scenario). The hypothesis that one is *not* a BIV follows from the more general claim that one is not envatted. The claim that one *is* a BIV, on the other hand, is a very specific hypothesis which does not follow from the general claim that one is envatted. It should not be surprising that BIVs are not be able to know, at least not by quick reflection on their sense experiences, that they are BIVs (just as, in the actual world it is not easy for agents to discover, e.g. that their bodies are made of cells). The more interesting question, however, is whether BIVs can know that they are envatted.

The deeper asymmetry concerns the notion of being *envatted* (rather than being a BIV) and the more general question what the agent is *in a position to know* (rather than what she happens to know). Externalists maintain that an agent such as Actual-Jill is in a position to know whether or not she is envatted, namely she in a position to know that she is not. By contrast, they typically assume that an agent such as BIV-Jill is not in a position to know whether or not she is envatted, i.e. she is not in a position to know that she is. To put things otherwise: even if Actual-Jill happens to have a false beliefs about this matter, she is nevertheless in a position to form a true and knowledgeable belief. By contrast, even if BIV-Jill happens to form a true belief beliefs on this issue, her belief would still fail to constitute knowledge.

Here, for example, is a typical expression of this claim from Williamson: ‘If one is in the bad case then one does not know that one is not in the good case. Even if one pessimistically believes that one is not in the good case, one’s true belief does not constitute knowledge;...Part of the badness of the bad case is that one cannot know just how bad one’s case is.’ (Williamson (2001): 165).⁷

Moreover, it is often thought that this kind of asymmetry is what lies at the heart of the externalist solution to scepticism: where the sceptic goes wrong is in assuming that an agent and their BIV-counterpart are epistemic equals. Here is Williamson again: ‘For the sceptics, the two cases are symmetrical: just as it is consistent with everything one knows in the bad case that one is in the good case, so it is consistent with everyone one knows in the good case

⁷ This perspective is by no means unique to Williamson. Here are some other relevant quotes:

- ‘No subject in the bad case can figure out she’s not in the good case’ (Silins (2005): 380)

- ‘Yet the brain in a vat is presumably in no position to tell that it lacks this evidence [that it has hands - OM]: if it were, it could conclude that it is in a very unusual situation, and the tragedy of the brain’s predicament is exactly that it is in no position to figure this out.’ (Salow (*forthcoming*))

- ‘But subjects in the bad case aren’t in a position to know that they are’ (Fratantonio and McGlynn (*forthcoming*))

that one is in the bad case. For the sceptic's opponent, the two cases are not symmetrical'. (Williamson (2000): 165))⁸

My aim in this paper is to argue that epistemic externalists need not be committed to this kind of asymmetry. On the contrary, I argue that the very same considerations that lead externalists to claim that Actual-Jill is in a position to know whether or not she is envatted, naturally lead to the claim that BIV-Jill is equally in a position to know whether or not she is envatted.⁹ My claim is thus that the sceptic's mistake is not in assuming that there is an epistemic symmetry between Actual-Jill and BIV-Jill. Rather, the sceptic's mistake is in assuming that the only way to resolve this symmetry is by maintaining that both agents are in an epistemically unfortunate situation. My proposal, by contrast, is that the symmetry can be resolved in the other direction, by affording knowledge of whether they are envatted to both agents.

II. How a BIV can know that she is envatted

i The General Argument. How can BIV-Jill come to know that she is envatted? The literature on scepticism tends to focus on brains-in-vats who, like their typical actual-counterparts, believe that they are not envatted and do not believe that they are envatted. Of course, BIV-Jill cannot not know that she is envatted while having that belief pattern. But the importance of this point should not be overstated: after all, if Actual-Jill happens to (falsely) believe that she *is* envatted, few epistemologists take that to prove that she is not in a position to know that is not envatted. Rather, they would claim that while Actual-Jill does not in fact know, she is nevertheless *in a position to know* that she is not envatted, because it's possible for her to form the right belief (in the right manner), and were she to do so, she would know that she is not envatted. My claim is that the parallel holds for BIV-Jill.

To simplify the discussion, I follow Chalmers (2005) in assuming that the computational process that feeds stimuli to BIV-Jill's brain gives rise to an ontology of virtual objects. These are non-material objects which are constituted by computational processes. On this way of construing things, while BIV-Jill does not have hands she does have *virtual-hands* – because

⁸ Technically, in this passage Williamson talks about what is consistent with what one knows, rather than what one is in a position to know – but the sceptic maintains that both agents are not even in a position to know.

⁹ I am being intentionally hedgy here. Few, if any, externalists offer a fully reductive analysis of knowledge, so there is always room for a logically consistent view where, were BIV-Jill to believe that she is envatted, her belief would not constitute knowledge. However, my claim is that this pessimistic conclusion is in no way forced on the externalist, and moreover, on the most natural way of fleshing out the details of the externalist views, BIV-Jill *is* in a position to know that she is envatted. It is also worth noting that some forms of externalism, ones based on normality notions, essentially build in the relevant kind of asymmetry by fiat into the view – my arguments are not intended to apply to such views.

there is a computational processes, connected to her in the right sort of ways, which simulates hands. This assumption, however, is not crucial. If one wishes to eschew the ontology of virtual objects, one can replace talk of BIV-Jill's belief that she has virtual-hands with more general beliefs of the form 'I am receiving inputs from a computational process which presents a hand-shaped appearance'.

To keep the discussion tractable, I focus on the case of process reliabilism (with some brief remarks on other externalist views in §iii). Assume then that on the basis of her vat experiences, BIV-Jill forms the belief that she has virtual-hands, and infers from this belief that she is envatted. Her beliefs are true, but do they constitute knowledge?

BIV-Jill's belief that she has virtual-hands is formed by a reliable process: roughly, inferring what her virtual reality is like from her experiences in the vat. In environments similar to that of BIV-Jill, the experiences one has in the vat are an extremely reliable guide to one's virtual reality¹⁰, and thus the reliabilist has no reason to deny that BIV-Jill's belief constitutes knowledge. And since the claim that she is envatted is entailed by the claim that she has virtual-hands, then her belief that she is envatted is also formed by a reliable process: valid inference from known premises.¹¹

Some might worry that the belief forming process that BIV-Jill uses isn't ultimately reliable: after all, BIV-Jill's vat experiences are indistinguishable from Actual-Jill's concrete-world experiences. But had BIV-Jill been in the concrete world she might have applied precisely the same processes (e.g. 'concluding on the basis of experiences that seem like thus that she has virtual-hands'), which would have resulted in many false beliefs.

The problem with this objection, is that it ignores that fact that reliabilism only requires truth-tendency in similar environments (or alternatively, individuates processes sufficiently finely that the agent has no access to the relevant process in very different environments). Of course,

¹⁰ That's not to say that BIV-Jill is infallible about the nature of her virtual reality. For example, if BIV-Jill is presented with a virtual Müller-Lyer illusion, she might falsely conclude that she is presented with two virtual lines of the same length, or if she is having an experience of a virtual-chair painted in red-48 she might falsely conclude that it is a virtual-chair painted in red-49. And of course, she might have lots of false higher-level beliefs, e.g. about the causal underpinnings of her virtual reality.

¹¹ Some might question whether the claim that BIV-Jill is envatted really is entailed by the claim that she has virtual-hands. (Perhaps, for example, if she is temporarily playing a computer game in which her character seems to have hands then she counts as having virtual-hands without counting as being envatted, not even temporarily). In that case, though, we can imagine a slightly different belief forming process: in the same manner that BIV-Jill gains knowledge that she has virtual-hands, she is able to gain knowledge that she has virtual-legs, is virtually-walking around virtual-streets, virtually-lives in a virtual-house, and so forth. She can then infer via the reliable process of inductive inference from known premises that she is envatted. In this manner, she can thus also gain knowledge that she is envatted.

the question of what is the right way to individuate processes and environments constitutes the (in)famous generality problem for reliabilism. However, it is worth remembering that precisely parallel worries can be raised with respect to the reliabilist's claim that Actual-Jill forms the belief that she is *not* envatted using a reliable process. If reliabilists succeed in individuating processes/environments in a way that renders the BIV-scenario irrelevant for undermining Actual-Jill's reliability, then presumably, they can maintain that Actual-Jill's scenario is equally irrelevant to undermining BIV-Jill's reliability.

ii. The Argument Refined. Some, however, might raise an entirely different worry about my line of argument. The worry is that the case is under-described: it's hard to figure out what sort of belief forming process would lead BIV-Jill to form the belief that she has virtual-hands or that she is envatted, and without details about her belief formation process, it is hard to assess whether this process really does yield knowledge.

The worry can be broken down to three closely related challenges. The first challenge is to show that there is a psychologically plausible story on which BIV-Jill succeeds in forming these beliefs. The second challenge is to show that this psychological story is still consistent with the claim that BIV-Jill forms her belief via a reliable process. (After all, if the psychological story is, for example, that BIV-Jill happens to read in the tea-leaves that she has virtual-hands then her belief forming process is obviously not reliable.) The third challenge is a bit more delicate. The challenge concerns the notion of defeaters, and which form it takes (if it arises at all), depends on specific details of one's externalist view.

In addition to the standard externalists requirements on knowledge (e.g. that one's belief be safe or formed by a reliable process), many externalists also postulate an additional necessary condition on knowledge, one which introduces a minimal internalist component to the view: a 'no defeater' condition on knowledge.¹² Thus, for example, according to some externalists, if Norman the Clairvoyant forms a belief that the president is in town via a perfectly reliable mechanism of clairvoyance, his belief might still fall short of constituting knowledge if he also believes (or according to some, if he justifiably believes) that clairvoyance is utterly unreliable and that he ought not to believe that the president is in town. The third challenge, then, is to show that BIV-Jill forms her beliefs in a manner that is not only reliable, but also satisfies a relevant no defeater condition. However, it is worth keeping in mind just how minimal no defeater conditions typically are: for many externalists, all that is required is a 'no believed

¹² See Sudduth (2008) for an overview.

defeater' condition –namely that the agent in question does not believe that the belief in question is epistemically inappropriate.¹³ For those who interpret no-defeater condition in such a minimal manner it will be easy to meet the challenge: it suffices to stipulate that BIV-Jill does not take her beliefs to be inappropriate. Other externalists seem to require a slightly stronger condition. For example, according to some externalists, certain cases where an agent *should* have formed but did not in fact form the belief that their first-order belief is inappropriate are sufficient to bar the first-order belief from counting as knowledge.¹⁴ A lot more needs to be said about how the detail of such views are to be developed, and particular care needs to be taken to ensure that such views do not just collapse into run-of-the-mill internalism.¹⁵ Developing non-defeater-externalist positions in detail falls well beyond the scope of this paper. However, in what follows, I will assume that any total belief state which is (in some informal and fairly liberal sense) internally coherent, will count as satisfying the 'no defeater' conditions.

Let us consider now how the description of BIV-Jill's belief forming process can be fleshed out in a way that allows these three challenges to be addressed. We start by considering what the phenomenology of BIV-Jill's experience might be like. In line with how the BIV thought experiment is typically set up (and focusing, for simplicity of discussion, only on the visual aspect of her experience), I will hold fixed the following: the electrical stimuli present BIV-Jill with a visual scene which has precisely the same arrangement of colours, shapes, and other 'low-level' properties as Actual-Jill's visual scene. As debates in the philosophy of perception show, however, this constraint doesn't necessarily fully determine the nature of the visual phenomenology of BIV-Jill.¹⁶

I consider three possibilities: according to the first, BIV-Jill's visual phenomenology is neutral concerning whether she is experiencing virtual-objects or ordinary concrete objects. According to the second, her phenomenology represents (or presents) virtual-objects. According to third, her phenomenology represents concrete objects.

¹³ See e.g. Nozick (1981: 196) and Bergmann (2006), ch. 6.

¹⁴ See Goldman (1986): 62-3, and Bergmann (2006), ch. 6 (though Bergmann takes these restrictions to fall in the purview of a different necessary condition on knowledge, on requiring proper function).

¹⁵ For example, if one assumes that whenever an agent believes something without justification (traditionally construed), they should take their belief to be inappropriate, then one's view of knowledge effectively demands the same sort of justification as any internalist view and is subject to similar worries (such as the regress problem).

¹⁶ See e.g. Siegel (2016)

Consider the first possibility: BIV-Jill's phenomenology is neutral on whether she is experiencing virtual objects or concrete objects. This can be either because her experiences don't have any content at all, or because her experience represents a very minimal content, one that is neutral between the BIV-world and the actual-world.

How might BIV-Jill form the belief that she has virtual-hands based on such neutral phenomenology? One possibility is that BIV-Jill starts out believing that she is envatted and this belief causally entails that, when presented with the neutral phenomenology, she forms the perceptual belief that she has virtual-hands.¹⁷ Another possibility, is that BIV-Jill forms the perceptual belief that she has virtual-hands using a similar causal process, but one which originates from a positive doxastic attitude which falls short of outright belief towards the proposition that she is envatted. For example, BIV-Jill might have a high prior credence or a mere disposition to believe that she is envatted. A third possibility is that BIV-Jill is simply directly disposed to believe that she has virtual-hands when presented with such neutral phenomenology without having any prior substantive doxastic attitude towards the proposition that she is envatted.¹⁸ (Her belief forming mechanism in that case would arguably be analogous to that which is often attributed to chicken-sexers, where upon seeing a chick, one reliably forms beliefs regarding the chick's sex, without having any background beliefs about which features of the chicks indicate their sex.)

Do these suggested belief forming processes address the three challenges raised above? In response to the first challenge, I have just described some psychological background facts that would lead BIV-Jill form the relevant belief. No doubt, these brief descriptions do not fully prove that these belief forming processes are compatible with human psychology, but I do not see any obvious reason why they would not be. (Indeed, given that some people in the actual world seem to believe that we are living in a computer simulation, it seems implausible to deny that such belief-patterns are possible.¹⁹) Suppose it is possible for BIV-Jill to form beliefs using one of the processes described. It seems that in that case, the beliefs in question are formed by a reliable process: if BIV-Jill responds to all (or most) experiential inputs she

¹⁷ Note that this causal process doesn't need to be inferential. To borrow an example from Littlejohn (2009): 134, if you know that your neighbour drives a Harley you might, upon hearing the sound of his motorcycle, form an immediate non-inferential perceptual belief that there is a Harley outside, even if the perceptual experience doesn't represent such high level properties as being a Harley.

¹⁸ One fairly easy way this might be the case is if due to content externalism, BIV-Jill only possesses the concept of virtual-hands and not that of hands, or at least has a much more immediate concept of virtual-hands. - I return to this issue in §III.i below.

¹⁹ See e.g. Bostrom (2003).

receives in the vat by forming the corresponding belief about her virtual reality, then her belief forming processes would tend to yield true beliefs, thus addressing the second challenge.

Finally, is BIV-Jill's total doxastic state coherent in way that satisfies a non-defeater condition? Again, it is hard to see why not. After all, on the envisaged scenario she does not believe that she is *not* envatted (on some versions, she positively believes that she is envatted), and has no other beliefs that suggest her perceptual stimuli are not reliable indicators of her virtual reality.²⁰ Moreover, beliefs about the nature of her phenomenology do not threaten to introduce any incoherence. For one thing, by the very definition of neutral phenomenology, it is one that (as far as the phenomenology itself goes) is compatible with either the claim that BIV-Jill is experiencing virtual-hands or that she is experiencing hands. To reinforce the point, let me make two additional observations. First, I suspect many non-sceptical epistemologists accept that Actual-Jill's phenomenology is similarly neutral. But then if Actual-Jill's doxastic state is taken to satisfy the no-defeater condition and allows her to count as knowing that she is *not* envatted, then presumably, combining such neutral phenomenology with a substantive belief about whether or not one is envatted is similarly coherent when BIV-Jill combines it with the positive belief that she *is* envatted. Second, even if we deny that Actual-Jill's phenomenology is neutral (as disjunctivists, for example, would), the following consideration suggests that BIV-Jill's doxastic state is internally coherent. Call the agent in the scenario we are discussing (namely, where BIV-Jill believes she has virtual hands and is envatted) '**SBJ**' (for 'sceptical BIV-Jill').²¹ Contrast her, with '**NSBJ**' (for 'Non-Sceptical BIV-Jill'), the agent in the more commonly discussed scenario, namely the one where Jill is a BIV but falsely believes that she is not envatted and that she is experiencing hands. I think it's extremely plausible that *if* the phenomenology of SBJ is neutral on whether she is experiencing hands or virtual-hands, then so is the phenomenology of NSBJ. But it is standardly accepted that NSBJ is – if not fully justified – at least internally coherent in her beliefs (that is, after all, a large part of what motivates the sceptical puzzle, as well as the related New Evil Demon problem). However, if NSBJ is internally coherent when she positively believes that she has hands based on her neutral phenomenology, then SBJ is internally coherent when she positively believes that she

²⁰ I have ignored here issues about evidence from other members of the community – but one can suppose that BIV-Jill either doesn't interact much with other agents, or else interacts with agents with similarly sceptical beliefs.

²¹ Note that 'sceptical' is merely a label for the case where BIV-Jill believes that she is envatted – it is not intended to take a stance on whether Chalmers (2005) is right that the hypothesis in question should not really be treated as a form of scepticism.

has virtual-hands based on similarly neutral phenomenology. The upshot, then, is that there is no reason to think that BIV-Jill's belief is defeated.

So far I have argued that in the case where BIV-Jill has neutral phenomenology, there is a psychologically possible, knowledge conducive belief forming process on which she forms the belief that she has virtual-hands. Exactly the same holds for the belief that she is envatted. As I suggested above, this belief can be formed inferentially from the belief that she has virtual-hands (by either a deductive or an inductive inference process), and parallel considerations to those above suggest that this belief is not defeated. One might worry that at least on one of the ways of fleshing out the story, BIV-Jill starts out with a prior belief that she is envatted which is causally relevant to her forming belief that she has virtual-hands, and this in turn is the inferential basis for her knowledge that she is envatted. But note that even on this way of fleshing out the story, my argument for why BIV-Jill knows that she has virtual-hands did not assume that she has prior *knowledge* that she is envatted (at best, we assume she has a prior belief), and thus there is nothing to block us from supposing that having acquired the knowledge that she has virtual hands, BIV-Jill can later rebase the belief she is envatted, now in a way that constitutes knowledge.

Assume next that the phenomenal character of BIV-Jill's experience is such that it represents (or presents) the world as containing virtual-hands. How could Jill have such a phenomenology? On the most plausible way of filling out the details, BIV-Jill has, just as in the previous case, either a prior positive doxastic attitude towards the claim that she is envatted, or a direct disposition to believe, when faced with a visual scene containing hand-like shapes, that she has virtual-hands. Moreover, these attitudes or dispositions influence the nature of her experience via a mechanism of cognitive penetration. Can BIV-Jill's belief constitute knowledge in this scenario? Again, as above her belief is true and reliably formed. Moreover, even if we add a no-defeater condition, this time her phenomenology *supports* her belief, so if she had knowledge in the previous scenario we certainly shouldn't deny her knowledge in the current one.²²

Clearly the most difficult case is one where the phenomenal character of BIV-Jill's experience represents her as having (concrete) hands. But I think even here the reliabilist can maintain that

²² Some might worry that the fact that this support comes via cognitive penetration renders it circular. Remember, though, that this supporting role is playing no role at all in *justifying* the belief (the belief counts as knowledge simply because it is reliably formed). I am merely mentioning the support as an argument for why the phenomenology does not provide a defeater.

BIV-Jill is able to have knowledge. I'd like to propose two different models for how BIV-Jill can come to know that she has virtual-hands in this case.

According to the first model, when BIV-Jill has experiences of virtual-hands, she is directly disposed to form the belief that she has virtual-hands. Of course, by stipulation those experiences have a phenomenal character that represents having hands, but on this model, BIV-Jill does not realise that her experiences represent having hands.²³ I maintain first, that her belief forming process is reliable: when she has experiences of virtual-hands she consistently forms the true belief that she has virtual-hands (and *mutatis mutandis* for experiences of other virtual objects). Of course, on any plausible way of how BIV-Jill might have such a disposition she would also respond to experiences of hands by forming (false) beliefs that she has virtual-hands. But as should by now be clear, this is not sufficient to render her belief forming process unreliable: BIV-Jill is embedded in an environment where she is not at risk of having any experiences of hands, and hence this remote possibility does not undermine her knowledge. More difficult is the question whether BIV-Jill's belief satisfies the no-defeater condition, but plausibly, the answer is positive: after all, BIV-Jill does not realise (and arguably is not easily in a position to realise) that her experience represents the property of having hands, and thus her doxastic state is arguably internally coherent.

A good analogy for the first model is the following case, discussed in detail in Littlejohn (2009). Suppose that Amy is colour blind, so that she experiences both red things and green things the way most of us experience red things. However, Amy grows up an environment containing no red objects, and where she couldn't have easily encountered any red objects. She learns to apply the label 'green' to green objects just as well as anyone else in her community who is not colour blind, and is disposed, when confronted with green objects, to form beliefs that she expresses by saying 'That is green'. At least arguably, when Amy has such a belief, it simply is the belief that the relevant object is green.²⁴ Moreover, since her belief forming process is as reliable as that of anyone else in her community, it is highly plausible that many such beliefs constitute knowledge. This case is analogous to my first model for BIV-Jill above: when Amy is presented with a green apple in normal lighting conditions, the phenomenal character of her

²³ As discussed above, a very natural way to develop this line of thought is to assume that BIV-Jill simply does not possess the concept of hands, but this assumption is not necessary. She might just be directly disposed to interpret these experiences as ones of virtual-hands.

²⁴ One might argue that Amy does not possess the concept of 'green' and hence cannot form such beliefs, but I don't think it's obvious that in order to possess a colour concept one must be acquainted with the relevant phenomenal character.

experiences arguably represents the property of being *red*. But Amy does not realise that it represents the property of being red, and based on her experience she forms the reliably true belief that the apple is green. If one accepts (as I think we should) that she knows the apple is green, then we should attribute similar knowledge to BIV-Jill.

According to the second model, BIV-Jill fully realises that her experience represents hands rather than virtual-hands. However, she has a prior belief in (or another positive doxastic attitude towards) the proposition that she is envatted, and moreover she predicts that envatted people experience of virtual-hands via phenomenology that represents hands. Thus, when she is confronted with an experience that represents having hands, she forms the belief that her experience is phenomenologically misleading and that she is experiencing virtual-hands. Again, her belief forming process is reliable: in her environment, she consistently and reliably responds to experiences of virtual-hands (albeit ones that represent hands) by forming the true belief that she has virtual-hands. Moreover, it does not seem that her doxastic state is internally incoherent: of course, the very content of her experience is in tension with the belief she forms on its basis, but her total doxastic state accounts for this divergence.

This second model is thus similar to a somewhat different scenario involving colour perception. Suppose that Amy is colour inverted so that she experiences green things the way most of us experience red things, and vice versa.²⁵ Suppose also that this time Amy fully realises that she is colour inverted in this way. When she looks at a green apple, her experience represents the apple as red and she fully realises it represents the apple as red. However, she (rightly) assumes that if an apple looks to her to be red it must, due to her colour inversion, actually be green so she infers that the apple is green. Her belief in this case seems to be a clear case of knowledge.

One might object that in the latter case Amy only knows that the apple is green if she has prior *knowledge* (or at least justified belief) that she is colour inverted in this way. After all, consider an agent Sammy who is just like Amy, except that rather than knowing that he is colour inverted, he randomly forms the belief that he is. One might worry that even though Sammy's colour-inversion belief happens to be true, if he form his belief that the apple is green on this basis, the belief does not constitute knowledge. This is a problem, according to the objection, because in so far as we assume that BIV-Jill comes to know that she is envatted by inferring

²⁵ Alternatively, she experiences both colours as typical agents experience red objects, but is not at risk at encountering any red objects.

this from the claim that she has virtual-hands, we do not want her belief that she has virtual-hands to require prior knowledge that she is envatted.²⁶

My intuitions on Sammy's case are not entirely clear, but even assuming that Sammy does not know that the apple is green, there are peculiarities of this case that do not generalise to the case of BIV-Jill. First, Sammy's belief about colour-inversion is a very specific one: if Sammy randomly formed the (true) belief that he is seeing green objects as red, then presumably he could have easily formed, using the same random process, the (false) belief that he is seeing blue objects as red and thus concluded that the apple in question is blue. But then the total process that Sammy used to form the belief that the apple is green is not after all reliable, which explains why he lacks knowledge. By contrast, BIV-Jill's belief that she is envatted systematically influences how she interprets all her visual experiences in both her scenario and nearby worlds. Her belief forming process is thus much more reliable than Sammy's.

Second, and more tentatively, our intuition about Sammy's case might be clouded by some natural ways of filling out the details of the case, ones that arguably entail that Sammy's beliefs are defeated – not merely by the representational content of his experience, but by other features of his total doxastic state. For example, we might be assuming that Sammy knows that colour-inversion is extremely rare so that his belief is unlikely to be true. BIV-Jill, as we're construing the case, though, has no such defeaters for her belief that she is envatted. If we really hold fixed Sammy has no prior evidence either way about the likelihood of colour-inversion, then at least those with sufficiently externalist leanings might be more open to attribute to him knowledge that the apple is green.

I conclude, then, that whatever view one takes on the phenomenal character of BIV-Jill's experience, there are ways to fill out the details of the case so that BIV-Jill's belief forming process constitutes knowledge.

iii *The Argument expanded.* My discussion thus far has focused on Process Reliabilism, but similar considerations apply to other externalist views. Consider the safety theoretic account of knowledge. Assume that there are no nearby world where BIV-Jill is not envatted and that BIV-Jill's 'sceptical disposition' (her positive attitude towards the claim that she is envatted or her disposition to interpret her hand-shaped experiences as experiences of virtual-hands) are

²⁶ It is worth noting, though, that not all externalist views would accept that BIV's knowledge that she is envatted can only be obtained in this inferential manner. For example, safety theorists might maintain that if BIV-Jill directly forms the belief that she is envatted and does so in a modally stable way, then since there are no nearby worlds where her belief is false, it constitutes knowledge.

modally stable. BIV-Jill's belief that she has virtual-hands is safe, because nearby worlds are either ones where she is still receiving the hand-shaped-stimuli and (truly) believes that she has virtual-hands, or ones where she is receiving alternative stimuli (e.g. ones simulating her hands getting cut-off in an accident) and ceases to believe that she has virtual-hands. Her belief that she is envatted is also safe, because all nearby worlds are ones where she is envatted and still believes that she is envatted (either because she starts out with this default belief or because she infers this from beliefs about virtual-objects). Of course, this result will only hold in those cases where our assumptions are true (namely, that both BIV-Jill's sceptical disposition and her envattedness are modally stable). But this is hardly surprising – it was never my argument that every case of a BIV who believes that she is envatted is a case of knowledge (e.g. it will clearly not be knowledge if she forms her belief by guessing or reading the tea leaves...). Moreover, these assumptions are precisely parallel to those that safety theorists make in accounting for how Actual-Jill can know that she is not a BIV: if there are nearby worlds in which Jill is a BIV but believes that she is not or alternatively, if there are nearby worlds where she believes that she is a BIV even though she is not, her actual belief would be unsafe.

Similar remarks apply to the externalist evidentialist view. The details depend (as before) on one's precise views on what counts as evidence and evidential support, but to get a feel: BIV-Jill's evidence might consist of the very sense experience of virtual-hands or of the proposition that she is experiencing virtual-hands. This evidence arguably supports her belief that she has virtual-hands, which can in turn support her belief that she is envatted.

A final note is in order: my discussion has focused on a case where BIV-Jill responds to her stimuli by forming the belief that she has virtual-hands and that she is envatted. One might worry, however, that when it is said that the BIV is not in a position to know that she is envatted, the intention is to focus on the more typical case of NSBJ, who does not respond to her stimuli in this way.

The simplest response to this worry is to note at the very least my discussion shows that at least *some* BIVs are in a position to know, merely on the basis of their vat-experiences, that they are envatted – a surprising conclusion in its own right. However, I do want to endorse the stronger claim that even NSBJ is in a position to know.²⁷ Does the fact that SBJ responds to her stimuli differently than NSBJ show that her case is irrelevant to whether NSBJ is in a position to know?

²⁷ Note that many of the implications discussed in Section IV depend on this stronger claim.

The issue depends on how the modality ‘in a position to know’ is interpreted. Clearly, it is not sufficient for S to be in a position to know that p, that it be metaphysically possible that S knows that p. (For one thing, as the notion is standardly interpreted, one is only in a position to know that p if it is true that p.) On the other hand, it is also clear that the modality does not require holding all psychological facts about the agent fixed. After all, it is entirely uncontroversial that one can be in a position to know that p even when one does not believe that p, but comparing a situation where one does not believe p to a possibility where one knows (and hence believes) that p, requires changing at least some facts about the agent’s psychology! Moreover, saying that the agent is in a position to know that p is standardly used to mark an epistemic rather than a psychological property (though I admit the distinction between the two is not entirely sharp). For example, consider the popular formulation of knowledge-closure which states that when one knows that p and knows that q, one is in a position to know that p and q. This principle is taken to mean that when one knows each of the conjuncts, one possesses all the epistemic backing which is required for the formation of a knowledgeable belief in the conjunction, rather than that one is psychologically disposed to form a belief in the conjunction. I conclude that the psychological differences between NSBJ and SBJ should not lead us to deny that NSBJ is in a position to know that she is envatted.

III. *Supplementary Externalist Commitments*

So far, I have argued that standard epistemic externalist theories predict that BIV-Jill is in a position to know that she is in envatted. In this section, however, I consider some additional externalist-friendly commitments that have played an important role in the scepticism debate, and offer some brief comments on their impact on my line of argument: Putnam’s semantic externalism (i), Nozick’s sensitivity condition on knowledge (ii), and epistemic contextualism (iii).

i. *Semantic Externalism.* In his seminal paper ‘Brains in a vat’ (Putnam (1981)), Putnam argues that a BIV cannot believe that she is not a BIV because to have that belief, her mental representations must be capable of referring to brains and to vats. But as she has had no causal contacts with such concrete objects, her mental representations cannot refer in this way.²⁸ In Putnam’s original discussion, semantic externalism was wheeled-in as a response to scepticism: if BIVs cannot even form the belief that they are not a BIV, then anyone who

²⁸ As is often pointed out, this argument might not apply to an agent that acquired her language and concepts in the concrete world before being transferred to a vat. I will focus on the (more difficult for my purposes) case where BIV-Jill has always been envatted.

believes that they are not a BIV is not at risk of having a false belief. My interest in this section is not, however, in semantic externalism as a response to scepticism. For the purposes of this paper, I am assuming that the sceptical challenge is addressed via standard epistemic externalist theories (reliabilism, etc.). Nevertheless, many philosophers of language are attracted to semantic externalism independently of its potential to address scepticism. And a natural worry is that accepting semantic externalism (alongside standard epistemic externalist theories) would threaten my conclusion in §II. After all, if BIV-Jill cannot form the (false) belief that she is *not* a brain in a vat because she lacks the relevant concepts, then presumably she also cannot form the (true) belief - let alone know - that she *is* a brain in a vat.

Recall, however, that the question I am focused on isn't quite whether BIV-Jill is in a position to know that she is a BIV, but rather on whether she is in a position to know that she is *envatted*. Does adopting semantic externalism still allow BIV-Jill form that belief, and moreover do so in a knowledge conducive way?

I think it does. Let us distinguish between terms that we might call 'semantically anchored' and those that we can call 'semantically neutral'.²⁹ Roughly, semantically anchored terms are ones that have a different meaning when used by Actual-Jill and by BIV-Jill. The term 'brain', I'll assume following the semantic externalist, is semantically anchored: Actual-Jill but not BIV-Jill uses it to refer to concrete brains. But even if we accept semantic externalism there are arguably still some semantically neutral terms, for example 'friend' or 'number'.

I consider three hypotheses, and show that each of them is potentially compatible with the claim that BIV-Jill is in a position to know that she is envatted. The three hypotheses vary according to their response to the following two questions. First, are terms for particular types of virtual objects such as 'virtual-hand' semantically anchored? On one view, such terms are semantically anchored: when Actual-Jill uses the term 'virtual-hand' she is referring to a virtual object that is constituted by a concrete (or concretely-realised) computational process, while when BIV-Jill uses it, she is referring to an object that is constituted by a virtual simulation of a computational process (a concrete computer that is simulating a computer which is simulating a hand). However, even those who accept that the term 'hand' is semantically anchored might nevertheless think that 'virtual-hand' is not semantically anchored – perhaps 'virtual-hand' refers to any kind of object that looks or feels like a hand but is constituted by some kind of

²⁹ I focus on the case of language here, but the discussion is meant to carry over to the corresponding mental representations.

simulation (whether it is a first or second-order simulation). Second, is the more general term ‘envatted’ semantically anchored? On one view it is semantically anchored: when actual-Jill uses the term she is referring to the predicament of receiving one’s input from a concrete computational process while when BIV-Jill uses the term she is referring to the predicament of receiving one’s inputs from a virtual simulation of a computational process. However, it is also possible to think that ‘envatted’ is a semantically neutral term: it refers to the predicament of receiving one’s inputs (whether directly or indirectly) from a computational simulation.³⁰

According to the first hypothesis (NEUTRAL), both ‘virtual-hand’ and ‘envatted’ are semantically neutral. According to the second hypothesis (ANCHORED) both ‘virtual-hand’ and ‘envatted’ are semantically anchored. According to the third hypothesis (MIXED), ‘virtual-hand’ is semantically anchored, but ‘envatted’ is semantically neutral.³¹

Assume first that NEUTRAL is right. In that case, Putnam’s semantic externalism will have little bearing on my arguments in §II: BIV-Jill can form beliefs using the concepts of virtual-hand (which she refers to as ‘virtual-hand’) and of being envatted (which she refers to as ‘being envatted’), so the argument will not be affected. Next, assume that ANCHORED is right. I will assume, following Putnam, that the anchored term ‘hands’ refers in BIV-Jill’s mouth to virtual-hands. Moreover, I take it that if claims such as ‘I am envatted’ are semantically anchored then plausibly, so are claims such as ‘I am in the concrete world’. Thus, on a natural way of developing ANCHORED, BIV-Jill can say to herself things such as ‘I have an experience of hands. Therefore, I must be in the concrete world’ but her thought would express the claim that she is having an experience of virtual-hands, and is therefore envatted.³² Finally, consider MIXED (i.e. ‘virtual-hands’ is anchored, but ‘envatted’ is neutral). This time I think the most plausible way to fill out the details is by assuming that BIV-Jill has an explicit metaphysical hypothesis about the underlying nature of the objects she labels as ‘hands’.³³ Thus, she might say to herself ‘I have hands’ (which expresses the belief that she has virtual-hands), and then infer, due to her metaphysical commitments, that she is envatted (a condition she refers to using the term ‘being envatted’).

³⁰ See Chalmers (2005) for a defence of the claim that ‘envatted’ is a semantically neutral term.

³¹ I don’t think anyone would be tempted by the view that ‘virtual-hand’ is neutral but ‘envatted’ is anchored – more general and abstract terms such as ‘envatted’ are better candidates for being neutral.

³² Putnam’s discussion suggests that he would be sympathetic to ANCHORED and this diagnosis of its implications.

³³ See Chalmers (2005) for a development of this idea.

The upshot, then, is that either way, BIV-Jill can form the beliefs that she has virtual-hands and is thereby envatted. Semantic externalism merely affects how she would phrase these beliefs. My conclusion that BIV-Jill is in a position to know that she is envatted is not threatened by Putnam's semantic externalism.

ii. *Nozick's Sensitivity Condition.* In my discussion in §II, I have focused on showing that BIV-Jill is in a position to satisfy the relevant externalist necessary condition on knowledge (e.g. that one's belief is formed by a reliable process or that it is safe). However, this is not quite sufficient that to show that she is in a position to know, for there might be additional necessary conditions on knowledge. One such condition, is Nozick's sensitivity condition.³⁴

According to Nozick, a necessary condition for S to know that p is that S's belief is *sensitive*: if p were false, S would not believe that p (on a standard possible-worlds account of counterfactuals: in the closest not-p worlds, S does not believe that p). This condition is compatible with the previously discussed epistemic externalist conditions, and is most naturally coupled with the safety condition.

Consider then a view of knowledge which requires both Safety and Sensitivity. Actual-Jill's belief that she is not envatted satisfies both these requirements: it is safe (because there are no close worlds where she falsely believes that she has hands), and it is sensitive (because the closest worlds where she doesn't have hands are ones where, for example, she loses her hands in an accident, realises this has happened, and hence no longer believes she has hands). Her belief that she is not envatted, on the other hand, does not satisfy these requirements: while it is safe, it is not sensitive because the closest worlds where she is envatted, are presumably ones where she nevertheless still believes that she is not envatted.³⁵

The exact parallel applies, I think, to the case of SBJ. Suppose that BIV-Jill believes, in the manner suggested in my discussion, that she has virtual-hands and that she is envatted. Her belief that she has virtual-hands is both safe and sensitive: we have already seen that it is safe, and it is sensitive because the closest worlds where she does not have virtual-hands are ones where the vat-computer simulates a handless person, and presumably in those worlds she does not believe she has virtual-hands. On the other hand, her belief that she is envatted is safe but

³⁴ See Nozick (1981): 179.

³⁵ Does my line of argument in this paper threaten this claim? I don't think so. Of course, it is a tricky issue to determine the relevant measure of similarity for assessing such counterfactuals, but it is not implausible to think that a world where BIV-Jill believes that she is not envatted are closer to the actual world than those where she believes that she is envatted.

not sensitive: in the closest worlds where she is not envatted, she presumably still responds to her indistinguishable experiences by forming the belief that she is envatted.

The upshot is that on a sensitivity account of knowledge, BIV-Jill is in a position to know that she has virtual-hands, but not in a position to know that she is envatted. But, of course, on that account, Actual-Jill is also not in a position to know that she is not envatted and hence this case presents no epistemic asymmetry between the situation of Actual and BIV-Jill.

iii. *Epistemic Contextualism.* Another theory which has been widely discussed in the context of the sceptical challenge, and which is compatible (and indeed often combined) with epistemic externalism, is epistemic contextualism. According to contextualists, the word 'know' is context sensitive: in different contexts, 'S knows that p', has truth-conditions that require different epistemic standards. Contextualist theories typically offer the following kind of response to the sceptical challenge: in some contexts ('low standard' ones), 'knows' expresses a relation that Actual-Jill stands in with respect to the proposition that she has hands and that she is not envatted. By contrast in other contexts ('high standard' ones), it expresses a relation that Actual-Jill fails to stand in with respect to those same propositions. This response is thus partially concessive to the sceptic: concessive because in high-standard contexts 'Actual-Jill does not know that she has hands' is true, but only partially so, as in other contexts such claims are false.

Epistemic contextualism can be naturally combined with epistemic externalism. For example, Heller (1995) argues for a version of Process Reliabilism, on which the generality problem is addressed by maintaining that processes are individuated in different ways relative to different contexts. Compare Actual-Jill who, in light of her perceptual experience of hands, forms the belief that she has hands with the case of NSBJ who, following her hand-like experiences in the vat, also forms the belief that she has hands. In some contexts, 'knows' is interpreted in such a way that construes Actual-Jill and NSBJ as utilising the same process, rendering the attribution 'Actual-Jill knows that she has hands' false, while in other contexts, it is interpreted in a way that construes Actual-Jill and NSBJ as using different processes, thus allowing the 'knowledge'-attribution to be true.

What does a view of this sort predict regarding my claim that BIV-Jill is in a position to know that she has virtual hands or is envatted? I think it makes exactly parallel predictions to the ones concerning Actual-Jill. Consider SBJ, who upon having a hand-shaped experience in the vat forms the (true) belief that she has virtual-hands. There are remote worlds where she is not

envatted and has indistinguishable experiences of real hands, but nevertheless forms the (false) belief that she has virtual-hands. In some contexts ‘knows’ expresses a relation that is based on an individuation of processes where in both worlds Jill counts as using the same process, and thus the attribution ‘BIV-Jill knows that she has virtual-hands’ is false, while in other contexts it expresses a relation where processes are individuated more narrowly, rendering the ‘knowledge’ attribution true. Similar observations apply to BIV-Jill’s ‘knowledge’ that she is envatted. Thus, according to the contextualist reliabilist, BIV-Jill is still in a position to ‘know’ that she is envatted, as long as we’re in a context where ‘know’ expresses a sufficiently low-standard relation. Again, this places BIV-Jill in an epistemically symmetric situation to that which the contextualist attributes to Actual-Jill.

§IV Implications

Debates in epistemology have, on the whole, simply taken for granted that the BIV is not in a position to know that she is envatted. So far I have argued that, given the relevant externalist theories, this assumption is wrong. In this final section, I will point out several ways in which debates in epistemology would be affected if we accept the following:

PTK: BIV-Jill is in a position to know that she is envatted

i. *Chauvinism and Epistemic Credit.* The first point I want to make is merely an observation about the dialectical impact of accepting PTK.

Externalist responses to the sceptical challenge are often met with unease. At least one source of unease concerns, I think, the impression that the response implies a kind of chauvinism. Compare Actual-Jill to NSBJ: the both respond to indistinguishable stimuli by forming analogous beliefs and thus seem equally reasonable, and yet only Actual-Jill gets full epistemic credit, i.e. knowledge. PTK does not deny that (non-sceptical) Actual-Jill gets more credit than NSBJ in this sense. However it does show that there is nothing special about the actual-world (or for that matter, about anti-sceptical beliefs): after all, there are other pairs of agents – SBJ in the vat-world and sceptical-Jill in the actual world, who both believe they are envatted, and yet this time it is BIV-Jill who is credited with knowledge. Getting epistemic credit isn’t reserved to beings in the actual world: it’s a matter of forming beliefs in a way that reliably matches one’s environment, and agents in both the Actual-world and the vat can equally succeed or fail in this task.

ii. *Propositional Justification, Doxastic Justification, and the New Evil Demon*. An increasingly discussed issue in epistemology is the New Evil Demon problem.³⁶ Consider the typical BIV, NSBJ. Defenders of the New Evil Demon problem maintain that even if epistemic externalists can account for why NSBJ lacks knowledge about the external world, they should nevertheless accept that NSBJ is as *justified* as Actual-Jill in her beliefs about the external world. This raises a problem, however, for various externalist views of justification which entail that NSBJ is not justified in her beliefs.

To see how PTK bears on this debate, recall a common distinction in the epistemology: that between propositional and doxastic justification.³⁷ The proposition that p is doxastically justified for an agent if and only if they have a justified belief that p. The notion of propositional justification is more controversial, but roughly the thought is that the proposition that p is propositionally justified for an agent if and only if the agent is in a position to form a justified belief that p. For example, if a detective on the scene possess excellent evidence that Jones is the murderer but fails to form the belief that Jones is the murderer, then the proposition that Jones is the murderer is said to be propositionally but not doxastically justified for the detective.

Here are three plausible principles about these notions:

- (i) If an agent S is in a position to know that p, then the proposition that p is propositionally justified for S.
- (ii) If the proposition that p is doxastically justified for an agent, then it is also propositionally justified for that agent.
- (iii) It is never the case that both the proposition that p and the proposition that not p are simultaneously propositionally justified for an agent.

Principle (i) is supported by how ‘being in a position to know’ and ‘is propositionally justified’ are standardly interpreted. If an agent is in a position to know that p, then they are in a position to form a knowledgeable belief that p. But assuming that any belief which constitutes knowledge is also a justified one, the agent is also in a position to form a justified belief that p,

³⁶ See Littlejohn (2008) for an overview.

³⁷ See Goldman (1979)’s distinction between *ex ante* vs. *ex post* justification and the discussion Turi (2010). As Turi notes, the distinction is also sometimes labelled as ‘justified vs. justifiable belief’ or ‘justifiably believing vs. being in a position to justifiably believe’. Also relevant is Smithies’ (2012) distinction between doxastic and epistemic justification.

and hence p is propositionally justified for that agent.³⁸ Principle (ii) is entirely straightforward: one can only have a justified belief that p if one has the relevant epistemic support for that belief.³⁹ Principle (iii) is standardly assumed by many frameworks for thinking about evidence and justification. For example, if a proposition is propositionally justified just in case it has a high probability on one's evidence (with 'high' being set to any value greater than 0.5), then it cannot be the case that both p and not p are simultaneously justified.⁴⁰

Suppose, then, that we accept these three principles. Now consider NSBJ, who believes that she is not envatted. According to PTK, NSBJ is in a position to know that she is envatted. But then by (i), the proposition that she is envatted is propositionally justified for NSBJ; By (iii) the proposition that she is *not* envatted is not propositionally justified for NSBJ; So by (ii) the proposition that she is not envatted is not doxastically justified for NSBJ – or in other words, her belief that she is not envatted is not justified.

The upshot is that accepting PTK has substantive implication to the debate on the New Evil Demon problem and to how we think about propositional and doxastic justification more generally. Either we should accept that NSBJ is not justified in her belief that she is not envatted, thus undermining the very intuition driving the problem in the first place. Alternatively, we must incur a controversial commitment concerning the nature of justification: reject one of principle (i)-(iii) –for example, accept that one can know without being justified or that one can be propositionally justified in both a proposition and its negation.⁴¹

³⁸ Turi (2010): 322 suggests an objection to a related principle: if an agent possesses the means to form a belief that p which would be doxastically justified, then the proposition that p is propositionally justified for that agent. The objection concerns cases like this: suppose that I possess the means to prove a certain mathematical theorem T , but I haven't yet proven it. It seems that I possess the means to form a doxastically justified belief that I have proven T , even though the claim that I have proven T isn't propositionally justified for me yet. However, note that the way 'in a position to know' is standardly interpreted, it wouldn't count as being in a position to know that I have proven that T in this scenario (for one thing, being in a position to know is taken to be a factive operator – see Williamson (2000): 95).

³⁹ Turi (2010) objects to attempts to analyse doxastic justification in terms of propositional justification, but I am not suggesting such an analysis here – I am merely claiming that propositional justification is a necessary condition for doxastic justification.

⁴⁰ One framework on which (iii) is arguably not supported is a purely coherentist view of justification, but I think the whole concept of propositional justification does not sit well with such a view. For example, we might maintain that p is propositionally justified just in case there is some coherent total doxastic state available to the agent which contains p . This would allow cases where both p and not- p are propositionally justified, but on this definition any (or almost any) internally consistent proposition would count as propositionally justified as well.

⁴¹ It is worth pointing out in this context that it is far from trivial that the combination of accepting an externalist view of knowledge and the claim that knowledge entails justification entail that NSBJ is not justified in her belief: this combination still allows that one can be justified without knowing.

iii. Epistemic principles and being in a position to know. A wide range of epistemic principles are phrased using the notion of ‘in a position to know’.

Consider, for example, principles of epistemic transparency. On one construal, a condition C is transparent, if and only if, one always knows whether or not it obtains. Many, however, find transparency principles phrased in terms of actual knowledge hopelessly strong: any condition is plausibly such that one can simply fail to have any opinion on, and hence fail to know, whether it obtains. This caveat often leads scholars to phrase such transparency principles using the weaker notion of *being in a position to know*. Thus, for example, Williamson defines a condition C to be *luminous* if and only if ‘For every case α , if in α C obtains, then in α one is in a position to know that C obtains.’ (Williamson (2000):95), and Smithies (2012):725 extends this by saying that a condition is *strongly luminous* or *transparent* if, when it obtains then one is in a position to know that it obtains, and when it fails to obtain then one is in a position to know that it fails to obtain.

My discussion in this paper should, I think, lead us to be more cautious about the import and interpretation of such principles. Being envatted, for example, is often taken to be a paradigmatic case of a condition that is uncontroversially *non-transparent*: the BIV, the thought goes, is envatted, but is not in a position to know that she is envatted. Relatedly, the situation of the BIV is standardly taken to be a clear case against negative introspection of knowledge (the claim that if one fails to know, one is in a position to know that one fails to know). The BIV, the thought goes, does not know that she is not a BIV, but is in the unfortunate situation of not being in a position to know that she does not know this.

If PTK is right both of these alleged counterexamples to transparency need to be reconsidered. According to PTK, a typical BIV *is* in a position to know that she is envatted and hence the case isn’t a counterexample to the transparency of *being envatted*. Moreover, the BIV is arguably thereby also in a position to know that she does not know that she is envatted, entailing that the case isn’t a counterexample to the transparency of *knowledge* either.⁴² To be clear: I am not saying that PTK shows that being envatted or knowing *are* transparent conditions. Other counterexamples plausibly show that they are not: for example, a less typical BIV, one that is envatted but could very easily have been non-envatted, might not be in a position to know that

⁴² The simple thought is that in the situation where BIV-Jill knows that she is envatted she is also able to infer that she does not know that she is not envatted. Admittedly, this simple thought might be a bit too quick, as there is a delicate issue of whether this possibility holds fixed enough facts about BIV-Jill to entail that NSBJ is in a position to know that she does not know. I will not pursue the issue any further here.

they are envatted (even if they were to form a true belief that they are envatted, that belief would arguably not be safe), and thus would also not be in a position to know that they do not know that they are not envatted. My point, however, is that care needs to be taken in showing what kind of counterexamples refute transparency and luminosity principles.

Conversely, I think more care needs to be taken in understanding what transparency principles entail if and where they do hold. Both proponents and opponents of transparency seem to often assume that if a condition *C* is strongly transparent, then agents have a fool-proof recipe or algorithm for forming true beliefs concerning whether or not *C* holds.⁴³ But if I am right, this picture is misleading. Assume, hypothetically, that the only two possible scenarios are those of Actual-Jill and BIV-Jill (i.e. there are no two close situations where in one Jill is envatted and in the other she is not). Roughly, this translates to saying that being envatted is a transparent condition for Jill: in the actual-scenario she is not envatted and is in a position to know that she is not envatted; in the BIV-scenario she is envatted, and is in a position to know that she is envatted. Still, none of this implies that in either scenario she has a fool-proof algorithm for determining whether or not she is envatted. In each scenario, the fact that she is in a position to know depends on her being lucky enough to form the right beliefs in the right environment. Thus even if there are conditions for which transparency holds, this may prove much less than either proponents or opponents take it to show.

Another interesting point demonstrated by PTK is that the notion of being in a position to know is not closed under (multi-premise) closure, in the following sense: one might be in a position to know that *p*, and in a position to know that *q*, but not in a position to know that *p* and *q*. For consider NSBJ. She does not believe, and hence does not know, that she is envatted. Moreover, we may assume that she knows that she does not believe, and hence that she does not know, that she is envatted. But clearly, if one knows that *p*, one is in a position to know that *p*. Thus NSBJ is in a position to know that she does not know that she is envatted. According to PTK, she is also in a position to know that she is envatted. However, she is not in a position to know that she is envatted and does not know that she is envatted. For suppose that there is a possibility where she knows the conjunction: 'I am envatted but do not know that I am envatted'.

⁴³ Thus consider for example the following from Williamson (2000): 95 'If one is in a position to know *p*, and one has done what one is in a position to do to decide whether *p* is true, then one does know *p*. The fact is open to one's view, unhidden, even if one does not yet see it...The picture is that a luminous condition always shines brightly enough to make its presence visible.'

Assuming knowledge of a conjunction entails knowledge of each of the conjuncts, this would entail that she knows that she is envatted, and (due to the facticity of knowledge of the second conjunct) that she does not know that she is envatted. Since this is impossible, it follows that BIV-Jill is not in a position to know the conjunction. Given that ‘being in a position to know’ is a kind of possibility operator, the rejection of the relevant closure principle is perhaps not entirely surprising (possibly p and possibly q, does not entail possibly p and q), but the lesson is worth taking on board when applying the notion of being in a position to know.⁴⁴

Another place where the notion of being in a position to know sometimes plays a central role is in discussions of (possessed) normative reasons. On one standard view, the fact that p is a reason you possess to *phi* only if you know that p.⁴⁵ Others, however, have defended a slightly weaker version of this constraint: perhaps it is only required that you be *in position to know* that p.⁴⁶ Moreover, this view applies in particular to reasons for belief, which in turn might be taken to be what justifies a belief. If we accept this view, then the fact that BIV-Jill is in a position to know that she is envatted has crucial implications to what she is justified in believing.

Consider Lord’s discussion of the New Evil Demon problem (Lord (*forthcoming*)). NSBJ believes that she has hands, and Lord argues that she has a very good reason to believe this: namely, that it appears to her that she has hands. If reason possession requires knowledge, though, it is far from clear that NSBJ possesses this reason: after all, she might not form the belief that it appears to her that she has hands (she might only form the stronger false belief that she *has* hands). But this objection is dissolved on the view on which possession merely requires being in a position to know. Interestingly, Lord argues that NSBJ’s reasons to believe that she has hands are just as weighty as Actual-Jill’s reasons to believe that she has hands. Roughly, the idea is that while Actual-Jill possesses more reasons for believing that she has hands (for example that she sees her hands), she also possesses more potential defeaters for that belief than NSBJ (for example, Actual-Jill knows that sometimes she perceives things in a misleading way – a defeater that isn’t available to BIV-Jill). But whether or not Lord is

⁴⁴ The point I make in this paragraph is very much related to the rejection of closure for ‘in a position to know’ argued for in Haylen (2016).

⁴⁵ See Hawthorne and Magidor (*forthcoming*), §2 for an extensive defence of this principle.

⁴⁶ See Gibbons (2013) and Lord (*forthcoming*)

otherwise correct, PTK entirely undermines this analysis.⁴⁷ For if, as I have argued, NSBJ is in a position to know that she is envatted, then this counter-claim constitutes a very strong counter-reason for NSBJ's belief that she has hands, rendering her overall belief unjustified.⁴⁸

References

- Bergmann, M. (2006): *Justification without Awareness*, Oxford: Oxford University Press.
- Bostrom, N. (2003), 'Are you living in a computer simulation?', *Philosophical Quarterly* 53: 243-255.
- Chalmers, D., (2005): 'The Matrix as metaphysics', in C. Grau (ed.), *Philosophers Explore the Matrix*, Oxford: Oxford University Press, pp. 132-176.
- Fratantonio, G. and McGlynn, A. (*forthcoming*): 'Reassessing the case against evidential externalism', in V. Mitova (ed.), *The Factive Turn in Epistemology*, forthcoming in Cambridge University Press.
- Gibbons, J. (2013): *The Norm of Belief*, Oxford: Oxford University Press.
- Goldman, A. (1979): 'What is justified belief?', in G. Pappas (ed.), *Justification and Knowledge*, Boston: D. Reidel, pp. 1-25.
- Goldman, A. (1986): *Epistemology and Cognition*, Cambridge, MA: Harvard University Press.
- Hawthorne, J. and Magidor, O. (*forthcoming*), 'Reflections on the ideology of reasons', in D. Star (ed.), *The Oxford Handbook of Reasons and Normativity*, forthcoming with Oxford University Press.
- Haylen, J. (2016): 'Being in a position to know and closure', *Thought* 5:63-67.
- Heller, M. (1994): 'The simple solution to the problem of generality', *Nous* 29: 501-515.
- Littlejohn, C. (2008): 'The new evil demon problem', *Internet Encyclopaedia of Philosophy* (version accessed on 05/12/2017).
- Littlejohn, C. (2009): 'On the coherence of inversion', *Acta Analytica* 24: 127-137.

⁴⁷ That is, assuming that possession requires only being in a position to know. Lord's line of argument can still be run if possession requires knowledge and NSBJ knows that it appears to her that she has hands.

⁴⁸ Thanks to Jeremy Fantl, Anil Gomes, John Hawthorne, Alex Kaiserman, Maria Lasonen-Aarnio, Guy Longworth, David Liebesman, Scott Sturgeon, Levi Spectre and audiences in, Barcelona, Bristol, Haifa, London, and Oxford for helpful discussion of this paper. Thanks also to the Leverhulme Trust for financial support of this research.

- Lord, E. (*forthcoming*): ‘Defeating the externalist deamons’, in F. Dorsch and J. Dutant (eds.), *The New Evil Demon*, forthcoming with Oxford University Press.
- Mitova, V. (2015): ‘Truthy psychologism about evidence’, *Philosophical Studies* 172: 1105-1126.
- Neta, R. (*forthcoming*), ‘Why must evidence be true?’, in V. Mitova (ed.), *The Factive Turn in Epistemology*, forthcoming in Cambridge University Press.
- Nozick, R. (1981): *Philosophical Explanations*, Cambridge MA: Harvard University Press.
- Putnam, H. (1981): ‘Brains in a vat’, in Putnam H., *Reasons, Truth, and History*, Cambridge: Cambridge University Press, chapter 1.
- Pritchard, D. (2005), *Epistemic Luck*, Oxford: Oxford University Press.
- Salow, B. (*forthcoming*): ‘The externalist’s guide to fishing for complements’, forthcoming in *Mind*.
- Schellenberg, S. (2013): ‘Experience and evidence’, *Mind* 122: 699-747.
- Schellenberg, S. (*forthcoming*): ‘Perceptual capacities, knowledge, and Gettier cases’, in R. Borges, C. de Almeida, and P. Klein (eds.), *Explaining Knowledge: New Essays on the Gettier Problem*, Oxford: Oxford University Press.
- Sudduth, M. (2008), ‘Defeaters in epistemology’ in *Internet Encyclopaedia of Philosophy* (version accessed on 2/08/2017).
- Siegel, S. (2016), ‘The contents of perception’, in E. Zalta (ed.), *Stanford Encyclopaedia of Philosophy*, Spring 2016 edition.
- Sillins, N. (2005): ‘Deception and evidence’, *Philosophical Perspectives* 19: 375-404.
- Sosa, E. (1999): ‘How to defeat the opposition to Moore’, *Philosophical Perspectives* 13: 141-153.
- Weisberg, J. (2012): ‘The bootstrapping problem’, *Philosophy Compass* 7: 597-610.
- Williamson, T. (2000): *Knowledge and Its Limits*, Oxford: Oxford University Press.