This is the bad case

Citation for published version:
McGlynn, A 2018, 'This is the bad case: What brains in vats can know', *Proceedings of the Aristotelian Society*, vol. 92, no. 1, pp. 183-205. https://doi.org/10.1093/arisup/aky011

Digital Object Identifier (DOI):
10.1093/arisup/aky011

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the Aristotelian Society

‘***<u>This</u>*** is the Bad Case’: What Brains-in-Vats Can Know

Aidan McGlynn
University of Edinburgh

Dugald Stewart Building
3 Charles Street
Edinburgh
EH8 9AD

Abstract The orthodox position in epistemology, for both externalists and internalists, is that a subject in a ‘bad case’—a sceptical scenario—is so epistemically badly off that they cannot know how badly off they are. In her paper, Ofra Magidor contends that externalists should break ranks on this question, and that doing so is liberating when it comes time to confront a number of central issues in epistemology, including scepticism and the new evil demon problem for process reliabilism. In this reply, I will question whether Magidor’s argument should persuade externalists, whether it really engages with the orthodox view on what subjects in bad cases can know, and whether the dispute is, as Magidor insists, a significant one for contemporary epistemology

I. Introduction Externalists and internalists in epistemology don’t agree about much. They don’t even always agree on what it is they are disagreeing about, and there are internal disagreements about this within each camp too. Is the internalist’s claim one about access or supervenience on the mental (see Conee and Feldman 2002 and Bergmann 2006: chapter 3)? If the former, does one need to actually have access, or is in-principle access good enough? Does one have access to the fact that one is justified, or the justifiers themselves (see Hatcher 2018)? If the core internalist thesis is rather one about supervenience on the mental, are we talking about justification, evidence, or both? And what’s the supervenience base – is it just one’s non-factive mental states, or one’s total mental state?[1] Should internalists care about knowledge? Is externalism simply the denial of internalism, or is there a distinctive positive thesis about what kinds of ‘external’ factors are epistemically relevant, or the relationship between justification and knowledge? Should we be pluralists, combining elements of externalism and internalism?

This is, to use the proper technical vocabulary, a complete mess.[2] But amidst all the disagreement, we can find some points of consensus, at least if we set aside those with sceptical tendencies. Here is one. Suppose we want to compare the epistemic fortunes of an epistemically successful subject in the ‘good’ case, Actual-Jill, and an unfortunate

[1] Conee and Feldman (2002, 2008), and Bird (2007) hold that justification supervenes on one’s total mental state, including any factive mental states; McGlynn 2012 argues that internalists should restrict the supervenience base to non-factive mental states (though in fact I’m sceptical that there really are any factive mental states (McGlynn 2014a: chapter 8 and 2017)). Silins 2005 and Fratantonio and McGlynn 2018 weigh up the arguments for the internalist thesis that one’s evidence supervenes on one’s non-factive mental states.

[2] For a helpful overview of the debate, see Littlejohn 2012: chapter 1.

counterpart in some corresponding ‘bad’ case: for example, BIV-Jill, who is ‘a handless brain in a vat […] which is hooked up 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’ (Magidor 2018: 1).[3] Many non-sceptical internalists and externalists alike will agree that the subject in the bad case is doubly epistemically badly off. Few of her beliefs about the external world based on her experiences in the vat can amount to knowledge, since they’re probably not even true. But she cannot even know the nature of her own predicament; were she to believe that she’s is in a bad case, or correctly believe a description of her unfortunate circumstances, these beliefs would not be knowledge. Jane Austen’s Emma soothes herself with the following explanation of Mr. Elton’s presumption in taking himself to be a suitable match for her:

> Perhaps it was not fair to expect him to feel how very much he was her inferior in talent, and all the elegances of mind. The very want of such equality might prevent his perception of it. (Austen 2003)

Likewise, internalist and externalist orthodoxy alike has it that the epistemic badness of bad cases prevents those in them from knowing their predicament. As Timothy Williamson influentially puts the point:

> 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. (2000: 165)

The claim that subjects like BIV-Jill are so epistemically badly off that they can’t know how epistemically badly off they are marks a rare point of agreement between internalists and externalists. It was inevitable, then, that a philosopher would have a go at undoing this consensus. In particular, Ofra Magidor thinks that externalists of various sorts should break ranks on this question. In my reply, I’ll question whether her argument should persuade externalists, whether it really engages with the orthodox view on what subjects in bad cases can know, and whether the dispute is, as Magidor insists, a significant one for contemporary epistemology. In the next section, I examine how Magidor’s argument for the claim that BIV-Jill can know that she is envatted combines elements of both standard process reliablist treatments of cases of global deception and David Chalmers’s radical reconstrual of such cases as metaphysical hypotheses, while departing in crucial ways from both. I then turn to appraisal of Magidor’s argument and its conclusion, and I argue that it both misses its intended target (section 3), and should be unpalatable even to many externalists about justification (section 4). Finally in section 5 I’ll close by expressing some doubts about whether as much of epistemological significance hangs on these issues as Magidor’s paper advertises.

II. What Can Brains-In-Vats Know? Rather than approach Magidor’s argument directly, I’ll first sketch the issues raised by a standard process reliabilist approach to envatted subjects, and then look at Chalmers’s revisionary conception of envattment as a metaphysical

[3] I’ll have much more to say about ‘bad’ cases and what makes them bad below in section 4.

hypothesis rather than a sceptical hypothesis. Magidor's argument weaves together elements of both, and having them out on the table will enable me to offer an assessment of her argument in sections 3 and 4.

Let's start with process reliablism, according to which one's belief that P is justified if and only if it was formed by a reliable belief-forming method (and one has no defeaters for P). This leaves a fair amount of latitude in terms of how the details are to be filled out, including what kind of ratio of truth to falsity it takes for a belief-forming method to be reliable, how we should pick out the relevant types of methods, and how to understand the modality implicit in 'reliable'.[4] Each of these issues of detail is associated with one of the principal objections to the view. The first gives rise to the *threshold problem*; how high a ratio of truth to falsity is required for a process to be reliable enough to deliver justified beliefs? Second, we're interested in the ratio of true beliefs to false ones produced by repeatable process *types* rather than tokens, leaving us with the task of picking out the relevant processes-types. However, each token process is a token of many types, some of which are more reliable than others; the *generality problem* is that there seems to be no principled way to select the relevant process-type when evaluating whether a given belief meets the reliability condition (Conee and Feldman 1998). Finally, the third issue of detail raises the *new evil demon problem* (Lehrer and Cohen 1983, Cohen 1984). Your counterpart in a bad case being massively deceived by an evil demon seems, so the standard intuition runs, to be just as justified in their perceptual beliefs as you are. However, your counterpart's beliefs seem to be produced by processes that are massively unreliable in their world, and so by the reliabilist's lights are unjustified—a verdict at odds with the widespread intuition about such cases.

The last of these choice-points is most directly relevant to what reliabilists should say about envatted subjects, and so we will focus on it.[5] In response to the new evil demon problem, Alvin Goldman proposed that reliability involves a sufficiently high truth-to-falisity ratio in *normal worlds*. The demon's victim uses belief-forming methods that are utterly unreliable in the world in which she uses them, but since that's an abnormal world, this doesn't matter; her belief can be reliably-produced in the relevant sense, and so justified. Of course, we now need an account of which worlds are 'normal', and here reliabilists divide. Goldman himself offered an account of normal worlds as those that are consistent with common and general beliefs about the actual world (Goldman 1986: 107). With this characterisation in place, Goldman concluded that justification 'displays normal-world chauvinism', and it's this chauvinism that explains why we should regard the demon victim as having reliable and so justified beliefs, even though the demon world is not itself a normal world. However, Goldman's account has been widely rejected; some have questioned the epistemic significance it accords to 'common belief' rather than something more objective, while others have focused more on the lack of explanation it offers of what might justify this kind of normal-world chauvinism (Majors and Sawyer 2005, Comesaña 2010, Gerken 2013 and 2018, and Graham forthcoming). An alternative is to think of the relevant set of worlds as those that are relevantly similar to the world which fixes the content of the subject's broad

[4] I'll discuss the 'no-defeaters' condition in section 4.

[5] However, as Magidor recognises (6), we cannot entirely divorce consideration of the third issue from consideration of the second.

states, an approach that unifies content externalism and epistemic externalism (Burge 2003, Majors and Sawyer 2005, and Gerken 2013 and 2018).[6] What should such a reliabilist say about cases of mass deception, such as new evil demon cases or brains in vats? Well, it depends; as Mikkel Gerken (2018) stresses, the details of these scenarios matters, and we shouldn't simply assume that new evil demon cases and brains-in-vats should receive a unified treatment. In particular, what we should say about these kinds of bad cases depends on what we hold fixed between them and the corresponding good cases. However, the goal remains the same as on Goldman's account, namely reconciling reliabilism with the intuition that subjects in cases of mass deception can have justified beliefs. We'll come back to the new evil demon in section 5.

Let's turn now to the unorthodox conception of BIV-Jill's situation defended in Chalmers's paper 'The Matrix As Metaphysics' (2005). According to Chalmers, the brain-in-a-vat scenario (like envattment scenarios more generally) is a *metaphysical* hypotheses: 'a hypothesis about the underlying nature of reality' (2005: 136). In particular, Chalmers argues that such scenarios are equivalent to the conjunction of three metaphysical theses which we can't rule out as potentially correct accounts of the underlying nature of reality: that 'physical processes are fundamentally computational'; that 'our cognitive processes are separate from physical processes, but interact with these processes'; and that 'physical reality was created by beings outside physical space-time' (2005: 136).

However, although Chalmers holds that we can't rule out the truth of an envattment hypothesis, taken as a metaphysical hypothesis, he also argues that recognising that it's a metaphysical hypothesis shows that it is not a *sceptical* hypothesis:

> A brain in a vat is not massively deluded (at least if it has always been in the vat). Neo does not have massively false beliefs about the external world. Instead, envatted beings have largely *correct* beliefs about their world. If so, the Matrix hypothesis is not a skeptical hypothesis, and its possibility does not undercut everything that I think I know. (2005: 135)

Why think that BIV-Jill will have mostly true beliefs? Because the hypothesis is one about the underlying nature of hands, desks, Tucson, and so on—these are 'virtual objects', objects 'constituted by computational processes' (2005: 150)—but it doesn't call into question BIV-Jill's beliefs in the existence or the behaviour of these objects. Hands are, at some level, different sorts of things than probably most of us take them to be, but there are still hands:

> I still think I cannot rule out that I am in a matrix. But I think that even if I am in a matrix, I am still in Tucson; I am still sitting at my desk; and so on. So the hypothesis that I am in the matrix is not a skeptical hypothesis. (2005: 136)

[6] These philosophers wouldn't all sign on to precisely what I've said here without qualification; indeed, not all of them call themselves reliabilists. This will be important in section 4 below.

In response to the objection that if Chalmers's brain were envatted in New York, he would surely be mistaken when he believes that he's sitting at his desk in Tucson, Chalmers appeals to content externalism: when he believes he is sitting in Tucson, he is thinking about a virtual location; virtual-Tucson. Were BIV-Chalmers to learn that he is envatted, the only beliefs he need revise are a few of his metaphysical beliefs:

> If [the brains-in-vats] discover their situation, and come to accept that they are in a matrix, they should not reject their ordinary beliefs about the external world. At most, they should come to revise their beliefs about the underlying nature of their world; they should come to accept that external objects are made of bits, and so on. These beings are not massively deluded: most of their ordinary beliefs about their world are correct. (2005: 145)

With this background in mind, let's turn to Magidor's argument. Notice that she is not arguing that BIV-Jill can know that she is a brain-in-a-vat, but only that she can know that she is, in a more general sense, envatted:

> 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 they are many other examples (for example, an agent who is a brain in a bottle, or a fully-bodied agent whose brain is hooked up 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 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. (3-4)

Let me pause to register some puzzlement about this. I agree that it would be surprising if BIV-Jill could learn 'just by quick reflection on [her] sense experiences' that she is a brain-in-a-vat rather than a brain-in-a-bottle, but by the same lights, it would be surprising if she could learn on such a basis that she's envatted—'hooked up to electrical stimuli'—rather than in some other kind of bad case: being deceived by a Cartesian demon, for example. I don't detect any significant contrast.

In any case, let's examine how Magidor argues that BIV-Jill can know that she is envatted. Recall that BIV-Jill 'is a handless brain-in-a-vat […] 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' (1). In order to '[simplify] the discussion', Magidor follows Chalmers in assuming that 'the computational process that feeds stimuli gives rise to an ontology of virtual objects', though she suggests, plausibly enough, that talk of a virtual ontology can be eliminated in favour of talk of appearances of objects generated by computational processes (6). On 'the basis of her vat experiences', BIV-Jill forms the belief that she has virtual-hands, and infers from this that she is envatted.[7] Not only is BIV-Jill right that she has virtual hands, Magidor contends, her belief is formed by a reliable process—'roughly, inferring what her virtual reality is like from her experiences in the vat'—and so a good candidate to count as knowledge:

> In environments similar to that of BIV-Jill, the experiences one has in the vat are an extremely reliable guide to one's virtual reality, and thus the reliabilist has no reason to deny that BIV-Jill's belief constitutes knowledge. (6)

That's the gist of the argument. In order to make it stick, one might have thought that Magidor would need to supplement it with convincing answers to a number of seemingly pressing questions. What is the content and the phenomenology of BIV-Jane's hand-like experiences, and are these shared with Actual-Jill? When we evaluate reliability, which are the relevant worlds? Are we to follow Chalmers in taking virtual-hands to be, in some sense, what hands are in an envattment world, and to be what 'hands' refers to in the mouth of BIV-Jill? Does buying into Chalmers's take on envattment scenarios lead to a denial that such worlds are really epistemological bad cases, as he argues? If not, what is the epistemological significance of Magidor's conclusion?

For the most part, Magidor's argument proceeds by avoiding taking a stand on these questions; instead, she attempts to show that her conclusion is more or less inevitable no matter how one answers them. So she is officially neutral on: the precise details of how BIV-Jill gets her belief that she has virtual-hands in the first place, the contents of the experiences she bases that belief on, and even what it's like to be in a vat (8-14); whether Chalmers is right that virtual-hands are just hands with a particular underlying metaphysical nature and that "hands" in BIV-Jill's mouth refers to virtual-hands (17-20); and whether the truth of her beliefs about her virtual reality show that we're no longer dealing with a sceptical scenario (10fn21). Following Heller 1995, Magidor tells us that a reliable process is one that 'has a tendency to produce true beliefs in similar environments to those of the original belief' (2), but we're not told much about what kind of similarity is relevant here.

[7] Magidor's descriptions (9-14) of how BIV-Jill might come to believe that she is envatted on the basis of her vat-experiences strike me as portraying her as close to irrational, rather than knowledgeable, since she either is inexplicably disposed to form the belief that she has virtual hands on the basis of hand-like experiences, or has an equally inexplicable prior belief that she is envatted. However, this worry—that BIV-Jill's belief that she has virtual-hands isn't rationalised or explained by her evidence— seems of a piece with objections to standard versions of process reliabilism based on familiar villains such as Norman (Bonjour 1980). Norman, recall, is a perfectly reliable clairvoyant, though he doesn't have any evidence for thinking that such a thing is possible, let alone that he himself is one. He just finds himself with beliefs about the whereabouts of the President at that moment, and these beliefs are invariably true, and so the anti-reliabilist alleges that he has unjustified beliefs which are the product of a reliable process. The parallel isn't exact, since Norman lacks any kind of experiences associated with his clairvoyance, while BIV-Jill does have experiences which are relevant; on some views, this contrast might make a big difference (for example, Comesaña 2010). The danger is descending into swapping familiar internalist and externalist intuitions, and so I'll leave this kind of concern here.

Indeed, since most existing accounts of reliability have been designed to solve the new evil demon problem by rescuing the intuition that victims of mass deception can nonetheless have warranted or justified beliefs, and since Magidor expressly argues against this response to the problem (22-3—we will return to this in section 5), the one conclusion we can draw about how to understand the notion of ‘similarity’ Magidor has in mind is that it probably can’t be anything like the options we reviewed above. This latitude in how we should answer these kinds of questions is presented as a strength of the argument; I’m not convinced it really is.

It at least makes the argument rather hard to get the measure of, since we need to assess the initial line of reasoning together with Magidor’s complicated set of attempts to show that, contrary to what we might expect, it isn’t dependent on any particular stance on the issues just raised. In order to try to keep things manageable, I’m going to focus on two interrelated and crucial issues. In the next section, I’ll argue that Magidor’s conclusion that BIV-Jill can know that she’s envatted need not conflict with the orthodoxy in epistemology that I discussed at the start of this paper, and in section 4 I’ll suggest that attempts to rectify this shortcoming in Magidor’s argument face structural problems. Finally, in section 5 I’ll argue against the epistemological significance of Magidor’s conclusion.

III. Brains-in-vats and Bad Cases My first point is that there seems to be something of a bait-and-switch in Magidor’s argument. The orthodoxy that Magidor takes aim at is that *a subject in an epistemically bad case cannot know that she is in a bad case*. However, Magidor’s conclusion, that (some) brains-in-vats can know that they are envatted, looks consistent with this. Now, we need to be careful to focus in on the right point here. We might notice that the orthodoxy concerns whether a subject can know that she is *in a bad case*[8], while Magidor’s conclusion concerns whether a subject can know that she is *envatted*. That’s not what interests me here; I’m happy to concentrate on whether certain subjects can know that they are envatted. Instead, I want to stress that *the demarcation of the relevant epistemic subject varies*; the consensus concerns what subjects in bad cases can know, while Magidor’s conclusion concerns what brains-in-vats can know. If there are brains-in-vats who are not in bad cases, in the relevant sense, then Magidor’s conclusion isn’t in tension with the orthodox claim that envatted subjects in bad cases can’t know that they’re envatted.

What does it take for an envattment scenario to be a bad case?[9] Chalmers offers the following criterion: ‘a hypothesis that I cannot rule out, and one that would falsify most of my beliefs if it were true’ (2005: 134). This won’t do, for multiple reasons. It concedes too much to scepticism to insist upfront that we cannot rule out sceptical hypotheses (Williamson 2000: chapter 8). A more neutral starting point would be that BIV-Jill has the same phenomenology as Actual-Jill, or at least that any differences are ones which they

[8] For example, Williamson focuses on whether a subject in the bad case can know that they’re in the bad case, ‘presented descriptively as ‘the bad case’’ (2000: 168).

[9] I’m using ‘bad case’ narrowly here to pick out sceptical scenarios. In a different context it might be useful to recognise epistemically bad cases that are not sceptical scenarios (see Pritchard 2012: §5), though Gerken cautions against using a single label to cover a range of different kinds of cases (2018: 109-10).

would not be able to detect by reflection on their respective experiences; for example, Williamson stipulates that '[a]s far as [content] externalism permits, things appear to one exactly the same way in the good and bad cases' (2000: 165).[10] The demand that the scenario render most of our beliefs false also seems wrong. One immediately thinks of dream scenarios, but in fact, envattment scenarios are typically described in such a way as to leave open the possibility that one's beliefs are mostly true.[11] In light of this, we might instead suggest that a sceptical hypothesis would, if true, mean that most of my beliefs were either false or only true by luck.[12] Putting this together, and abstracting away from some qualifications, bad cases are ones in which things appear to the subject just as they would in the good case, but where that subject's beliefs are mostly false or only true by luck.

We can, compatibly with all of the constraints provided by Magidor's argument, understand BIV-Jill's situation so that it's relatively unsurprising that BIV-Jill can know that she is envatted: we keep almost nothing fixed between her scenario and Actual-Jill's; we have a suitably externalist epistemology and theory of content; and we help ourselves to portions of Chalmers's account of envattment. So suppose that Actual-Jill and BIV-Jill have only some really minimal phenomenology associated with their respective experiences in common: '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' (8). They differ in the contents of their experiences, and even in their visual phenomenology, once we start describing that phenomenology in richer terms. The content of their speech and attitudes varies too, given our externalsm about content. When Actual-Jill forms a belief she would express with 'I have hands', she forms a true belief about the concrete five-fingered appendages at the ends of her arms; when BIV-Jill forms a belief she would express with 'I have hands', she forms a different true belief about features of her virtual-reality. Let's stipulate that Actual-Jill and BIV-Jill both form mostly true beliefs on the basis of their experiences. Finally, we can suppose that Actual-Jill and BIV-Jill are using different belief-forming processes to each, each of which is reliable in the world it is employed.

Actual-Jill and BIV-Jill so far seem pretty symmetrically placed, epistemically speaking. If one adds that there is a route that Actual-Jill can take that allows her to know that she is not envatted, it seems very plausible that there will be an analogous route that BIV-Jill can take to come to know that she *isn't* envatted. But this symmetry seems unsurprising and unobjectionable, at least relative to the assumptions made above. Crucially, there seems to be no reasonable sense in which BIV-Jill is in a bad case, if this is how her situation is understood; her situation is metaphysically weird but epistemically good, to echo Chalmers.

[10] Thanks to Guy Longworth here. As he points out to me, we might need to weaken this constraint even further.

[11] Putnam 1981 is an (influential) exception; see Wright 1992: 69-70.

[12] I won't attempt to analyse the relevant notion of luck here, though elsewhere I've favoured a modal account (McGlynn 2013, 2014: chapter 2). One might understand such an account in terms of reliability or safety, so that (roughly) for a belief to be true by luck is for it to be produced by an unreliable process or for there to be nearby possibilities in which it is mistaken.

To be quite clear, I'm not saying that Magidor adopts or defends the assumptions made in the version of her argument just sketched, and so finds herself committed on this basis to denying that BIV-Jill can be in a bad case. The point is rather that the wide latitude left by the way Magidor presents her argument may be a liability rather than a strength; the assumptions I made above are all ones within what's permitted by that latitude, and this reveals that there are not sufficient *constraints* placed on the argument to ensure that BIV-Jill is in a bad case. Nothing in the terms of the argument ensures that the conclusion really engages the orthodox position concerning the limits of what subjects in bad cases can know.

I worry, then, that Magidor's conclusion misses the real target. Looking at the claims she cites as her target bears this impression out. Williamson, for example, assumes ('perhaps overly generously') that content externalism is 'consistent with grasping the relevant propositions in sceptical scenarios' (2000: 165). In the bad case 'things still appear as they ordinarily do, but are some other way; one still believes *p*, but *p* is false; by any standards, one fails to know *p*, for only true propositions are known'. So the bad case, given Williamson's assumptions, is one in which the subject has the same phenomenology and the same beliefs as in the good case (or as close as possible, given content externalism), but where her beliefs are false. When Williamson asserts that a subject in the bad case cannot know she is in the bad case, this is what he has in mind. The same is true of the other epistemologists who Magidor quotes echoing Williamson's claim (Silins 2005: 380 and Fratantonio and McGlynn 2018: 87). Magidor also quotes Bernhard Salow, who writes: 'the brain in a vat is presumably in no position to tell that it lacks [the evidence that it has hands]: 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' (forthcoming: 12). However, though Salow doesn't use the terminology of 'bad' cases, he stipulates that his counterpart who's a brain in a vat shares his experiences of hands but has a false belief that it has hands (forthcoming: 12); as I've characterised bad cases above, these stipulations ensure that the brain is in a bad case.

The question of whether a subject in a bad case can know that she is envatted isn't the same question as whether a brain-in-a-vat can know that she is envatted. Magidor focuses on the latter question, while it's a pessimistic answer to the former that has conditioned the philosophical debate around scepticism.

IV. I've Got It Bad (and That Ain't Good) What I've just argued is that Magidor's argument might establish that BIV-Jill can know that she's envatted at the expense of disconnecting that conclusion from debates about scepticism and bad cases. However, nothing said so far shows that Magidor's argument *depends* on an interpretation of BIV-Jill that prevents her from counting as occupying a bad case. If we assume that BIV-Jill is in a bad case and try to run Magidor's argument, can we still reach the conclusion that BIV-Jill can know that she is envatted? If so, then Magidor's argument has the power to challenge the epistemological orthodoxy around bad cases after all. However, in this section I'll contend that structural problems stand in the way of any attempt to reach the conclusion that BIV-Jill can both occupy a bad case and come to know that she's envatted.

Let's suppose that BIV-Jill is in a bad case. In particular, her vat experiences are phenomenologically as like Actual-Jill's experiences as content externalism permits, and like Actual-Jill, BIV-Jill has lots of beliefs about the external world—that she has hands, that she is in Tucson, and so on— based on her vat experiences; however, BIV-Jill's beliefs of this sort are mostly false. She also, and also on the basis of her vat experiences, forms the beliefs that she has virtual-hands, that she is in virtual-Tucson, and so on, and these beliefs are mostly true. Now, if this is what BIV-Jill's doxastic state looks like, it seems like unfertile territory for knowledge, including knowledge that she has virtual-hands.

Let's try to substantiate this suspicion. If BIV-Jill's belief she has a hand and her belief she has a virtual-hand are produced by the same process, then clearly the latter belief is not produced by a reliable process; each true belief about her virtual-reality will be matched by a false belief about concrete reality. But Magidor suggests that belief-forming processes should be more fine-grained than this, so BIV-Jill's belief that she has virtual-hands is the product of, 'roughly, inferring what her virtual-reality is like from her experiences in the vat' (6). We shouldn't be too quick to grant this favourable description of the belief-forming process employed by BIV-Jill, however, since we haven't been offered any principled reason for taking this fine-grained process-type to be the relevant one to consider, rather than a more general process-type that could include both the formation of BIV-Jill's largely true beliefs about her virtual-reality and her false beliefs about concrete reality as tokens. This is just the generality problem for process reliabilism biting again.[13] Magidor is aware of this issue, but she insists that the non-sceptical externalist should concede that BIV-Jill is employing a fine-grained, reliable method that's not available to Actual-Jill:

> …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 such 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. (7)

However, our concern isn't with the possibility that Actual-Jill might interfere with BIV-Jill's reliability, but rather with a version of BIV-Jill who shares all of Actual-Jill's beliefs about non-virtual-reality in addition to forming the belief that she has virtual-hands. What reason do we have for taking BIV-Jill, so understood, to be employing two processes rather than a single, more general, and unreliable process?

Even if we concede that BIV-Jill's belief that she has virtual-hands is produced by a different process than her beliefs about non-virtual-reality, and that the former process is reliable even if the latter is not, many externalists may be resistant to the further claim that BIV-Jill's

[13] Magidor draws on Heller 1995 in formulating her version of reliabilism, and Heller is precisely concerned to solve the generality problem. However, Heller's contextualist strategy is aimed at showing that there's nothing problematic about our lack of a general principle that determines which process-type is relevant for reliability; that doesn't help us to decide which process-type is relevant in particular hard cases, like the ones confronted in the text.

belief is knowledgeable. There are several constraints on knowledge we might invoke here, though I lack the space to try to lay them out in detail or decide between them here, and so I'll be brief. The least controversial of these constraints is perhaps a *no-defeaters* condition, and in fact, Magidor herself explicitly endorses this as a requirement on knowing (7). She doesn't try to fully formulate such a condition, but she assumes that it suffices to satisfy it that one's total belief state is consistent. It's not clear whether BIV-Jill meets this condition; I suspect it depends heavily on the details of her case. BIV-Jill's beliefs that she has virtual-hands and so is envatted are consistent with her belief that she has hands (as is shown by Neo's predicament at the outset of *The Matrix*).[14] However, if we take BIV-Jill to believe that she's not envatted, or to believe herself to know that she has hands, for example, then her total belief state isn't consistent, since these beliefs are both inconsistent with her belief that she's envatted. It doesn't follow that BIV-Jill doesn't know that she's envatted, since Magidor only specified a *sufficient* condition for meeting the no-defeaters condition, but the point is still suggestive. importantly, even if externalists grant that BIV-Jill satisfies the no-defeater condition, there are a number of other worries they might have with the idea that BIV-Jill is gaining knowledge from a fine-grained reliable process while simultaneously taking on board a host of false beliefs from a related process which takes the same inputs (her vat-experiences). Here are some salient options: one might think that there a sense in which it's merely lucky that BIV-Jill uses a reliable method to form beliefs about her virtual reality, given that she indiscriminately relies on both reliable and unreliable belief-formation methods; relatedly, one might think that knowledge requires that one's reliable or modally-robust success in believing is due to one exercising a competence or ability, and that BIV-Jill's unreliable process calls this into question with respect to her belief that she is envatted; or one might hold that BIV-Jill's reliability with respect to her virtual reality is accidental given the content of her experiences (see for example Burge 2003, Pritchard 2012b, Gerken 2013, Friend 2014 and Graham forthcoming). I won't defend any of these lines of respond here; crucially, though, these are all diagnoses of why BIV-Jill might fail to know that she's envatted despite employing a reliable process which can be motivated on *externalist* grounds.

Magidor does address a point related to this one in her paper, arguing that a variant of BIV-Jill who mistakenly believed that she's *not* envatted and that she has hands—Magidor calls her NSBJ, for 'non-sceptical BIV-Jill—would still be *in a position to know* that she's envatted, even if she doesn't in fact know that she's envatted due to her belief to the contrary (15-6).[15] NSBJ is, it is reasonable to suppose, in a bad case, so if Magidor is correct that NSBJ is in a position to know that she's envatted, then it looks like the point I've been arguing for in this section is wrong; a subject in a bad case can—is in a position to—know that she is envatted. From this perspective, the mistake in the argument I've just offered is that it assumes that Magidor needs a version of BIV-Jill who both believes lots of false things about her external environment *and* true things about her virtual reality (including that she has virtual-hands). But this is to overlook the possibility that Magidor highlights, namely that NSBJ might share the same beliefs about non-virtual reality as Actual-Jill, lack the belief that she is envatted, and yet still be in a position to know that she's envatted.

[14] As noted above, this would not be true if we focused on Putnam's envattment scenarios (Wright 1992).

[15] Thanks to Kegan Shaw here.

What does ‘in a position to know’ mean here? The idea is roughly that everything epistemic is all set for one to have knowledge, and one’s doxastic state just needs to catch up; for example, Williamson writes that if one is in a position to know a proposition P, and one does all one is in a position to do to determine whether P, then one knows that P (2000: 95). This leaves the modality of being ‘in a position to’ rather open[16], and Magidor rightly flags this as the crucial issue (16). It’s not enough for being in a position to know P that it be metaphysically possible that one knows P; one needs to have more in common with one’s knowing counterparts to count as being in a position to know. But the factors that need to be held in common are epistemic rather than ‘psychological’ (though Magidor acknowledges this distinction is not entirely sharp); one is in a position to know that P when one has the same ‘epistemic backing’ for believing P as one’s knowing counterpart, but this leaves open that one may respond differently to that epistemic backing to one’s counterpart in terms of what psychological states one forms, thereby missing out on knowledge which one’s counterpart possesses. The upshot is that so long as the difference between NSBJ and her more enlightened envatted counterpart (who knows that she is envatted) is merely psychological, Magidor concludes, we should not deny that NSBJ is in a position to know.

I don’t think that this is sufficient to show that NSBJ is in a position to know that she is envatted, in any controversial sense. I conceded that NSBJ counts as being in a bad case, and this is plausible in part because we’re assuming certain things about her psychological states: her beliefs. In particular, we’ve assumed she forms the same beliefs as Actual-Jill, and so takes herself to have hands, to be Tucson, to not be envatted, and so on. Above I argued that such beliefs interfere with her forming a knowledgeable belief that she is envatted on the basis of her vat-experiences. Suppose that’s right. Then NSBJ is in a position to know that she is envatted only in the sense that, were she not to have those interfering beliefs about the external world, she would be able to form a belief that she’s envatted via a reliable process (and meet any other conditions for knowing).

That’s not a genuine challenge to the orthodoxy concerning the limits of what subjects in bad cases can know about the predicament they’re in. The orthodox claim is that conditions in bad cases are so unconducive to the possession of knowledge that subjects in them can’t even know how badly off they are. Were such subjects not held back by these conditions, they might well fare better, and come to know things about the predicament that they’re in. We can, if we’re so minded, introduce a notion of ‘being in a position to know’ that captures this; in this sense, envatted subjects in bad cases can be said to be in a position to know that they are envatted. But we shouldn’t mistake this latter claim for a threat to the externalist and internalist orthodoxy. To challenge the orthodox view, one needs to show that the conditions constitutive of being in a bad case are not after all an obstacle to knowledge of one’s predicament. One can take that line, but I’ve argued it requires externalists to commit themselves to a version of reliabilism that places implausibly few conditions on what it takes to have knowledge, even by their own lights.

[16] As Magidor notes (25), some of Williamson’s other remarks suggest he has a demanding reading of this claim in mind.

V. Scepticism and the New Evil Demon Let's take stock. So far, I have argued that Magidor's argument may miss its intended target. It's relatively plausible that some brains-in-vats can know that they are envatted. However, the orthodoxy she intends to challenge is a claim about what subjects in bad cases can know about their own situations, and there are brains-in-vats that do not count as occupying bad cases, in the relevant sense. In the previous section, I attempted to show that the gap between Magidor's actual conclusion and one that would undermine the orthodoxy may not be readily bridgeable—that assuming that BIV-Jill is in a bad case makes it hard to see how her belief that she has virtual-hands could count as knowledge. In this final section, I want to question the significance of Magidor's argument and conclusion.

Magidor makes several related remarks on the epistemological significance of her argument and its conclusion, meant to illustrate her claim that '[d]ebates in epistemology have, on the whole, simply taken for granted that the BIV is not in a position to know that she is envatted' (21). I'm not convinced that right, nor am I convinced that epistemological debates ought to have paid more heed to this issue than they have. To reiterate the point from section 3, Magidor's presentation fails to cleanly differentiate between claims about the limits of what subjects in bad cases can know about their situation with claims about what BIV-Jill can know about hers (regardless of whether she counts as being in a bad case). And as we have seen, epistemologists who have addressed these issues have made claims about the limits of what subjects in bad cases can know about their predicaments, rather than about brains-in-vats in general. That should already make us wary of Magidor's claim that epistemology has taken the latter for granted.

I'll close by briefly considering two of Magidor's points concerning the significance of her conclusion in more detail. First, she holds that the claim that BIV-Jill cannot know she is envatted underwrites existing epistemic externalist responses to scepticism. Second, she suggests that giving up this thesis points the way to a solution to the new evil demon problem for reliabilism.

First, let's consider scepticism. Externalists seem well placed to respond to underdetermination-based sceptical arguments. Such arguments claim that Actual-Jill's evidence or justification is no better than that of BIV-Jill in a corresponding bad case, and challenge us to explain how then Actual-Jill can have justification, warrant, or knowledge on that basis. The idea is the commonalities between Actual-Jill and BIV-Jill force an epistemic symmetry between them, and since BIV-Jill is doing poorly, epistemically speaking, Actual-Jill must be doing poorly too. What's distinctive of externalist responses, as I understand them, is that they offer accounts of justification or evidence that break that symmetry: accounts on which justification or evidence do not supervene on the conditions kept fixed between Actual-Jill and BIV-Jill. Actual-Jill can avail herself of evidence or justification that BIV-Jill cannot, even if she wouldn't be able to detect any differences in her own phenomenology and BIV-Jill's. Williamson articulates the anti-sceptical force of this as follows:

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

Magidor takes this to show that externalist responses to underdetermination-based scepticism buy into the claim that BIV-Jill can't know that she's envatted, even though Actual-Jill can know that she's not envatted. But this is just to make the conflation I've already pointed out. Williamson's point concerns an epistemic asymmetry between good and bad cases. This doesn't commit him to the claim that an asymmetry remains when we consider brains-in-vats who are not in bad cases, and there's no reason to think that such a claim lies at the 'heart' of externalist responses to scepticism, as Magidor claims.

Turning to the new evil demon, Magidor offers an argument that NSBJ—who, recall, is the more standard variety of BIV-Jill who is taken in by appearances, and mistakenly believes that she's *not* envatted, that she has hands, and so on—doesn't have a justified belief that she's not envatted.[17] The argument starts from the thesis that even though NSBJ wouldn't *know* that she's envatted, she's still in a position to know that she's envatted. This is something I conceded above (though I argued that it's only true on an interpretation that renders it uncontroversial). Appealing to some plausible principles concerning being in a position to know and propositional justification, Magidor argues that it follows that NSBJ has propositional justification for the claim that she is envatted, and so lacks it for the negation of that claim: the proposition that she is not envatted. Assuming that propositional justification is necessary for doxastic justification, it follows that NSBJ lacks a justified belief that she is not envatted. Magidor concludes:

> [T]his shows that the intuition that drives the problem (namely, the claim that NSBJ is justified) is incorrect and thus the problem does not even get off the ground. (23)

One worry with this concerns a potential gap in the argument that's revealed when we pin down what the content of the internalist intuition is supposed to be. In Magidor's initial setup of the problem, the intuition is that NSBJ is 'as ***justified*** as Actual-Jill in her beliefs about the external world' (22). It's natural to read this as saying that NSBJ is as justified in believing that she has hands, that she is in Tucson, and so on, as Actual-Jill is in her analogous beliefs, even though NSBJ's beliefs are false. This accords with how the new evil demon problem is usually presented and understood. However, Magidor's argument doesn't reach the conclusion that NSBJ lacks justification for these quotidian beliefs, but rather that NSBJ isn't justified in believing that she is not envatted. However, I suspect this gap isn't too difficult to plug. If doxastic justification is constrained by an epistemic closure principle, then if NSBJ lacks a justified belief that she is not envatted, she'll lack justified beliefs in any proposition that she recognises to entail that she's not envatted; this might include familiar candidates, such as the proposition that she has hands.[18]

[17] Magidor doesn't make any distinction between the envattment scenario and the new evil demon scenario; see Gerken 2018 for relevant discussion.

[18] Roughly, such a closure principle says that if one justifiably believes P, and competently infers Q from P, then one justifiably believes Q. For a defence of a related closure principle for propositional justification as applied to the Moorean inference discussed in the text, see McGlynn 2014b.

The real objection to Magidor's treatment of the new evil demon is that it's unclear how its conclusion, that NSBJ's beliefs about the external world are unjustified, is meant to dissolve the problem. The problem is precisely a tension between the commitments of externalist views of justification and a seemingly powerful internalist intuition; by itself, simply declaring the intuition 'incorrect' on externalist grounds seems more like a restatement of the problem than a way of showing that it 'does not even get off the ground'.[19]

References

Austen, Jane. 2003 (1816): *Emma*. London: Penguin Classics.

Bergmann, Michael. 2006: *Justification Without Awareness*. Oxford: Oxford University Press.

Bird, Alexander. 2007: 'Justified Judging', *Philosophy and Phenomenological Research* 74 (1), pp. 81-110.

Bonjour, Laurence. 1980: 'Externalist Theories of Empirical Knowledge', *Midwest Studies in Philosophy* 5 (1), pp. 53-74.

Burge, Tyler. 2003: 'Perceptual Entitlement', *Philosophy and Phenomenological Research* 67 (3), pp. 503-48.

Chalmers, David. 2005. 'The Matrix As Metaphysics'. In Christopher Grau (ed.), *Philosophers Explore the Matrix*. Oxford: Oxford University Press, pp. 132-76.

Cohen, Stewart. 1984: 'Justification and Truth', *Philosophical Studies* 46 (3), pp. 279-95.

Comesaña, Juan. 2010: 'Evidentialist Reliabilism', *Noûs* 44 (4), pp. 571-600.

Conee, Earl, and Richard Feldman. 1998: 'The Generality Problem for Reliabilism'. Reprinted in Earl Conee and Richard Feldman, *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press, pp. 135-58.

Conee, Earl, and Richard Feldman. 2002. 'Internalism Defended'. Reprinted in Earl Conee and Richard Feldman, *Evidentialism: Essays in Epistemology*. Oxford: Oxford University Press, pp. 53-82.

Conee, Earl, and Richard Feldman. 2008: 'Evidence', in Quentin Smith (ed.), *Epistemology: New Essays*. Oxford: Oxford University Press, pp. 83-104.

Fratantonio, Giada, and Aidan McGlynn. 2018: 'Reassessing the Case Against Evidential Externalism'. In Veli Mitova (ed.), *The Factive Turn in Epistemology*. Cambridge: Cambridge University Press, pp. 84-101.

Friend, Stacie. 2014: 'Believing In Stories', in Greg Currie, Matthew Kieran, Aaron Meskin, and Jon Robson (eds.), *Aesthetics and the Sciences of Mind*. Oxford: Oxford University Press, pp. 227-48.

Gerken, Mikkel. 2013: *Epistemic Reasoning and the Mental*. Basingstoke: Palgrave Macmillan.

Gerken, Mikkel. 2018: 'The New Evil Demon and the Devil in the Details'. In Veli Mitova (ed.), *The Factive Turn in Epistemology*. Cambridge: Cambridge University Press, pp. 102-22.

[19] Thanks are due to Giada Fratantonio, Guy Longworth, Bryan Pickel, Lukas Schwengerer, Kegan Shaw, and Martin Smith for very helpful discussion and feedback, and to Andy Clark, Mikkel Gerken, and Duncan Pritchard for enabling me to get access to various crucial items in the references.

Graham, Peter. Forthcoming: ‘Why Should Warrant Persist In Demon Worlds?’. In Peter Graham and Nikolaj J.L.L. Pedersen (eds.), *Epistemic Entitlement*. Oxford: Oxford University Press.

Goldman, Alvin. 1986: *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.

Hatcher, Michael. 2018: ‘Accessibilism Defined’, *Episteme* 15 (1), 1-23.

Heller, Mark. 1995: ‘The Simple Solution to the Problem of Generality’, *Noûs* 29 (4), pp. 501-15.

Lehrer, Keith, and Stewart Cohen. 1983: ‘Justification, Truth, and Coherence’, *Synthese* 55 (2), 191-207.

Littlejohn, Clayton. 2012: *Justification and the Truth-Connection*. Cambridge: Cambridge University Press.

Magidor, Ofra. 2018.

Majors, Brad and Sarah Sawyer. 2005: ‘The Epistemological Argument for Content Externalism’, *Philosophical Perspectives* 19 (1), pp. 257-80.

McGlynn, Aidan. 2012: ‘Justification As “Would-Be” Knowledge’, *Episteme* 9 (4), pp. 359-74.

McGlynn, Aidan. 2013: ‘Believing Things Unknown’, *Noûs* 47 (2), pp. 385-407.

McGlynn, Aidan. 2014a: *Knowledge First?* Basingstoke: Palgrave Macmillan.

McGlynn, Aidan. 2014b: ‘On Epistemic Alchemy’. In Dylan Dodd and Elia Zardini (eds.), *Scepticism and Perceptual Justification*. Oxford: Oxford University Press, pp. 173-89.

McGlynn, Aidan. 2017: ‘Mindreading Knowledge’. In J. Adam Carter, Emma Gordon, and Ben Jarvis (eds.), *Knowledge First: Approaches in Epistemology and Mind*. Oxford: Oxford University Press, pp. 72-94.

Pritchard, Duncan. 2011: ‘Evidentialism, Internalism, Disjunctivism’. In Trent Dougherty (ed.). *Evidentialism and its Discontents*. Oxford: Oxford University Press, pp. 235-53.

Pritchard, Duncan. 2012a: *Epistemological Disjunctivism*. Oxford: Oxford University Press.

Pritchard, Duncan. 2012b: ‘Anti-Luck Virtue Epistemology’, *Journal of Philosophy* 109 (3), pp. 247-79.

Putnam, Hilary. 1981: *Reason, Truth, and History*. New York, NY: Cambridge University Press.

Salow, Bernhard. Forthcoming: ‘The Externalist’s Guide to Fishing For Complements’, *Mind*.

Silins, Nicolas. 2005: ‘Deception and Evidence’, *Philosophical Perspectives* 19 (1), pp. 375-404.

Williamson, Timothy. 2000: *Knowledge and its Limits*. Oxford: Oxford University Press.

Wright, Crispin. 1992: ‘On Putnam’s Proof That We Are Not Brains-In-a Vat’, *Proceedings of the Aristotelian Society* 92 (1): 67-94.