

Semantic self-knowledge and the vat argument

Joshua Rowan Thorpe

Philosophical Studies

An International Journal for Philosophy
in the Analytic Tradition

ISSN 0031-8116

Volume 176

Number 9

Philos Stud (2019) 176:2289-2306

DOI 10.1007/s11098-018-1126-3

Volume 176 · Number 9 · September 2019

PHILOSOPHICAL STUDIES

AN INTERNATIONAL JOURNAL FOR PHILOSOPHY
IN THE ANALYTIC TRADITION

EDITOR-IN-CHIEF: STEWART COHEN

EDITORS: WAYNE DAVIS, JENNIFER LACKEY

 Springer

 Springer

Your article is protected by copyright and all rights are held exclusively by Springer Nature B.V.. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".



Semantic self-knowledge and the vat argument

Joshua Rowan Thorpe¹ 

Published online: 18 June 2018
© Springer Nature B.V. 2018

Abstract Putnam's vat argument is intended to show that I am not a permanently envatted brain. The argument holds promise as a response to vat scepticism, which depends on the claim that I do not know that I am not a permanently envatted brain. However, there is a widespread idea that the vat argument cannot fulfil this promise, because to employ the argument as a response to vat scepticism I would have to make assumptions about the content of the premises and/or conclusion of the argument that beg the question against the sceptic. In this paper, I show that this idea is mistaken.

Keywords Epistemology · Philosophy of mind · Scepticism · Hilary Putnam · Brains in vats

Putnam's (1981, pp. 1–22) vat argument is intended to show that I am not a permanently envatted brain. The argument holds promise as a response to what I will call *vat scepticism*, scepticism that depends on the claim that I do not know that I am not a permanently envatted brain.¹ However, there is a widespread idea that, even if it is sound, the vat argument cannot fulfil this promise, because to employ the argument as a response to vat scepticism I would have to make assumptions

¹ Putnam intended the vat argument as an attack on metaphysical realism. I will not discuss metaphysical realism in this paper, although my defence of the vat argument may be of use to anyone who does want to use it in such an attack.

✉ Joshua Rowan Thorpe
joshthop@gmail.com

¹ Centre for Logic, Epistemology and History of Science, University of Campinas, Campinas, SP 13083-859, Brazil

about the content of the argument—that is, about the contents of the premises and/or conclusion of the argument—that beg the question against the sceptic.² Thus, even among those who grant the soundness of the vat argument, it is not widely endorsed as a response to vat scepticism.

The idea that the vat argument cannot serve as a response to vat scepticism because of a problem regarding my knowledge of the content of the argument is like a bump in the carpet. You can step on the bump, but, rather than disappearing, it tends only to pop up again somewhere else. Likewise, we shall see that objections to particular manifestations of the idea that there is a problem regarding my knowledge of the content of the vat argument tend to result in it manifesting again at a different point in the argument, rather than in its disappearance. This tendency might be thought to be a sign that there is something fundamentally right about the idea that there is a problem regarding my knowledge of the content of the vat argument, even if this idea is difficult to formulate correctly.

In this paper I show that, despite its resilience, the idea is mistaken. I do this by showing that, at each step in my deployment of the vat argument, the only assumptions I need make about its content are legitimate in the context of a response to vat scepticism. Thus, the bump is, as it were, forced right off of the edge of the carpet, never to return.

1 Vat scepticism and the vat argument

Let us start with a brief sketch of vat scepticism, and of the response to it that the vat argument promises to provide.

We can distinguish between two different sorts of sceptic.³ The more demanding sort of sceptic claims that I am required to give a successful argument for the claim that I have empirical knowledge before I am entitled to conclude that I do. This is a very strong requirement, and in the absence of an argument for this requirement the more demanding sceptic is simply *too* demanding, and can therefore be dismissed. “Guilty until proven innocent” is no better as an epistemic principle than it is as a legal principle.

The advocate of vat scepticism is not the demanding sort of sceptic. Rather, she is a less demanding sort of sceptic, regarding my empirical beliefs as “innocent until proven guilty”. She is happy to concede that I am entitled to think that I have empirical knowledge, until she has provided me with a good argument that I do not. She then attempts to provide that argument. My sketch of the vat sceptic’s argument will be charitable; I will grant her points that should perhaps be questioned.

² This idea is present in MacIntyre (1984), Falvey and Owens (1994), Brueckner (1986), David (1991), Wright (1992) and Johnsen (2003). As we shall see, different authors have given the idea different forms.

It may immediately strike the reader that for a worry about knowing the contents of the premises and/or conclusion of an argument to make sense, the premises and conclusion must be conceived of as things that *have* content, rather than as things that *are* content. I address this point in Sect. 1. My thanks to an anonymous *Philosophical Studies* referee for pushing me to be clear about this.

³ Byrne (2004, pp. 300–303) makes a similar distinction.

Consequently, when we turn to the response offered by the vat argument, we have the prospect of defeating the vat sceptic on her own terms.

The vat sceptic begins her argument by asking me to imagine that I (the author of this paper) am in the following scenario:

I always have been and always will be a brain in a vat (BIV), stimulated by a supercomputer to have the sensory evidence that I actually have, whilst in fact the world contains nothing except my brain in its vat and the supercomputer.

I will need to repeat the indented block of text many times in this paper, but it would be awkward to write it out in full every time, so let us set down some conventions. From now on I will simply write 'I am a BIV', on the understanding that this phrase is an elliptical way of re-writing the indented block of text above. Similarly, I will write 'I am not a BIV' as an elliptical way of re-writing the indented text preceded by 'it is not the case that', and I will write 'if I were a BIV' as an elliptical way of re-writing the indented text preceded by 'if it were the case that'. Note that these conventions will hold for text appearing within quotation marks, and for underlined text.

The claim that I am a BIV seems to be incompatible with almost all my actual empirical beliefs about the external world. Thus, the sceptic claims, if I do not know that I am not a BIV, then I do not have any empirical knowledge. For the purposes of this discussion we shall grant this claim.⁴ The sceptic goes on to claim that I do not know that I am not a BIV. She points out that, by stipulation, if I were a BIV, my sensory evidence would be the same as it actually is, and so I cannot appeal to my sensory evidence to rule out the possibility that I am a BIV.⁵ However, that I am not a BIV is a contingent statement, and so one might think that I cannot know it without appeal to my sensory evidence, either. Thus, I do not know that I am not a BIV. The sceptical conclusion that I do not have any empirical knowledge follows.

Someone who wanted to resist vat scepticism might begin by pointing out that some contingent statements *can* be known without appeal to sensory evidence; potential examples include the claim that I exist, and the claim that I am here. The central idea of the vat argument is that the claim that I am not a BIV is just such a claim. The argument can be put as follows:

- (1) I can express the content that I am a BIV, but
- (2) If I were a BIV, I could not express the content that I am a BIV; so
- (3) I am not a BIV.⁶

⁴ The sceptic might argue for this claim on the grounds that knowledge is closed under known entailment, or on the grounds that for a belief to count as knowledge my evidence must favour that belief over any hypothesis that is incompatible with it. See Brueckner (1994a, b) and Cohen (1998) for comparison of these strategies.

⁵ One strategy for resisting the sceptical argument is to argue that this stipulation is not legitimate; see for example Williamson (2000, pp. 164–208). However, for the purposes of this discussion we will grant the stipulation.

⁶ There are many different formulations of the vat argument in the literature. Rather than considering the pros and cons of different formulations, I am simply going to give the version of the argument that I wish to defend.

I think that this argument is sound, but it is not the aim of this paper to show that it is.⁷ However, I will now talk briefly about the reasons for thinking that the premises of the argument are true, since these reasons are relevant to our discussion.

What can be invoked in support of the premises is constrained by the fact that the vat argument is intended as part of a reply to scepticism regarding my empirical knowledge of the external world. My knowledge of the premises of the argument had better not presuppose any such knowledge, on pain of begging the question against the sceptic.

At first glance, it seems that my knowledge of (1) is indeed non-empirical. I know that I can express the content that I am a BIV, because I am, right now, thinking a thought with the propositional content that I am a BIV. (One might ask what exactly I mean by 'thought', and 'content' here. These are important questions, and I will turn to them once the basic idea of the vat argument has been laid out.) That I am right now thinking a thought with the content that I am a BIV is an instance of *semantic self-knowledge*, my first-person knowledge of the content of my current mental states. Semantic self-knowledge is said to be "introspective", "groundless" or "direct" in a way that my knowledge of the content of other people's mental states is not. How exactly we should understand these claims is controversial. However, for our purposes, one broadly agreed upon implication of these claims will suffice, namely, that my semantic self-knowledge does not depend (or, at least, need not depend) on empirical knowledge of my environment. So, since my reason for believing (1) is an instance of semantic self-knowledge, it does not seem to presuppose any empirical knowledge of my environment.⁸

It may seem that the argument of the preceding paragraph is in tension with the argument I am about to give that my knowledge of premise (2) is non-empirical. This is because the latter argument depends upon externalism about thought content, the view that to have thoughts with certain content a subject must have interacted with his or her environment in a particular way, and this view is often thought to be in tension with the idea that we have non-empirical semantic self-knowledge. However, I ask the reader to ignore this tension for the moment. I will return to it in Sect. 2.

⁷ Button (2013), Wright (1992) and Wright & Thorpe (forthcoming) argue at length for their own versions of the vat argument, and much of what they have to say would apply to the version that I give here.

⁸ It has been suggested that the claim that semantic self-knowledge is introspective should be understood as meaning that this knowledge literally depends on a kind of inwardly directed observation; for example, this view seems to be implicit in Locke (1960, p. 105). I do not myself think that this is correct, but if it is then my knowledge that I am thinking that I am a BIV could be said to be empirical in the sense that it would depend upon a sort of perception. However, my knowledge of (1) would not thereby depend on the sort of empirical knowledge that the sceptic uses the possibility that I am a BIV to threaten, namely, empirical knowledge of the *external world*. Thus, the view that self-knowledge depends on inwardly directed observation does not render my assumption that I know (1) question begging as part of a response to the sceptic.

Externalism about thought content can be motivated by armchair reflection on cases such as Putnam's Twin Earth scenario. Putnam imagines a planet, Twin Earth, which is identical to earth at the macro level. However, at the micro level, wherever there is H₂O on Earth there is XYZ on Twin Earth. What concept does my doppelganger on Twin Earth express when he uses the word 'water'? Most philosophers are inclined to say that he expresses a concept of twater, which has in its extension all and only XYZ, rather than a concept of water, which has in its extension all and only H₂O. A natural explanation of why this judgement is correct is that my doppelganger has had the right kind of causal contact with twater to acquire a concept of it, whilst he has not had the right kind of causal contact with water.⁹

This explanation suggests that there is what I will call a *causal constraint on concept possession* (or *causal constraint* for short): to be able to have concepts of certain things, a subject must have been in the right sort of causal contact with those things. On the natural assumption that the content of my thoughts depends upon what my concepts are concepts of, the causal constraint amounts to a form of externalism about thought content.

It is true that until we have an account of the "right sort" of causal contact for having the concept of a particular thing we will not be able to draw a sharp boundary between cases where I can and cases where I cannot have a concept of a particular thing, and thus between cases where I can and cannot have thoughts with particular content. However, we do not need to be able to draw a *sharp* boundary to recognise that some cases certainly lie on one side or the other of that boundary. For example, we recognise that the Twin Earth scenario is a case where my doppelganger cannot have a concept of water, and, Putnam claims, we recognise that a case in which I am a BIV is one in which I would not be able to have a concept of a BIV. Thus, it seems that externalism about thought content of the sort that we get if we accept the causal constraint on concept possession implies that if I were a BIV I could not have a thought with the content that I am a BIV.¹⁰

If (2) is true, it is necessarily true. However, although my knowledge of (1) is non-empirical, it is contingently true that I can think a thought with the content that I am a BIV. I might have lacked the conceptual resources required to be able to do so. Thus, although the vat argument gives me non-empirical knowledge of (3), which says that I am not a BIV, it does not establish that (3) is necessarily true. And it isn't. There are possible worlds in which I am a BIV, and in those worlds I cannot think a thought with the content that I am a BIV (or, more cautiously, it is perfectly compatible with the vat argument that there are such worlds). But in this world I can think a thought with this content, and so I am not a BIV.

⁹ Putnam (1973,1975) used the Twin Earth case to support semantic externalism, the view that the environment of a subject plays a role in determining the referents of his or her words. Others have adapted the argument as I have here to support externalism about thought content; see for example McGinn (1977).

¹⁰ See Button (2013, pp. 118–121) for a detailed argument that if I were a BIV I could not refer to BIVs. This argument can easily be adapted, *mutatis mutandis*, to show that if I were a BIV I could not have a concept of a BIV.

Of course, although the causal constraint ensures that if I were a BIV I would not have a concept of a BIV, it would not prevent me from having a concept of a BIV if I had been envatted only recently, perhaps by some process that I am unable to detect. It is sometimes suggested that this deprives the vat argument of epistemological interest. Even if the vat argument ensures that the sceptic can't depend upon the claim that I don't know that I'm not a BIV, can't she make use of a scenario in which I have recently been envatted to the same effect?¹¹

No. A scenario in which I was only recently envatted fails to falsify large swathes of my empirical beliefs, for example my beliefs about the past and my beliefs about general scientific laws, and so it cannot be used to threaten the claim that these beliefs amount to knowledge. Thus, scepticism motivated by a scenario in which I was recently envatted would be weaker than scepticism motivated by the worry that I am a BIV, which says that I have no empirical knowledge at all. If the vat argument forces the sceptic to retreat to this weaker form of scepticism then it is still of epistemological interest.¹² In fact, as I have argued elsewhere, precisely because a scenario in which I was recently envatted could at best be used to attack some but not all of my empirical beliefs, it cannot even be used in a successful argument for the weaker form of scepticism.¹³ This is because there is nothing to prevent me from relying on the empirical knowledge of mine that the recent envatment scenario does not threaten to justify my belief that I was not recently envatted.

This concludes my sketch of how the vat argument might be used to combat vat scepticism. The rest of this paper is devoted to fending off the idea that the argument cannot be so used, because in the course of making the vat argument I must make assumptions about the content of its premises and/or conclusion that beg the question against the sceptic. However, before we start it should be noted that for a worry of this sort to make sense it seems that we cannot think of the argument as a series of entities that *are* contents, such as, perhaps, propositions. If we think of the argument in this way, then to say that I do not know the content of one of the premises or of the conclusion of the argument is to say that I do not know the content of a content. That doesn't make any sense.

Let us think of the vat argument as a series of thoughts that I think, where by 'thoughts' I mean mental representational states that *have* content, rather than things that *are* content. It will make our discussion easier if we think of thoughts as sentences (or at least as being enough like sentences that it does no harm to talk about them as if they were). Ordinarily we would assume that I know the content of the thoughts that I think when I run through the vat argument, but my opponents wish to call into question the idea that this assumption is legitimate, at least in the course of a response to scepticism.

There has already been some talk of concepts, and there will be more in what follows, so I had better make it explicit that for the purpose of this discussion I am

¹¹ This idea appears in Wright (1992, pp. 86–90), Christensen (1993, p. 314) and Brueckner (1999, p. 237).

¹² Tymoczek (1989) and Brueckner (2016) make this point.

¹³ Thorpe (2017). See also Wright and Thorpe (forthcoming).

thinking of concepts as words that occur in my thoughts (or at least as being enough like words that it does no harm to talk about them as if they were). A concept is *of* something when it refers to that thing. So, the concept 'water' and the concept 'H₂O' are both concepts of H₂O, whilst only the latter is the concept 'H₂O'. This is why I said above that my doppelganger on Twin Earth has a concept *of* twater, rather than the concept 'twater', which, plausibly, is not a concept that my doppelganger has. We shall think of content as being individuated in a relatively course grained way, such that in order to think a thought with the content that water is wet I must employ a concept of water, but it need not be the concept 'water'; the concept 'H₂O' would do equally well. Thus, it is correct to say, as I did above, that my doppelganger thinks a thought with the content that twater is wet, despite the fact that, plausibly, he does not have the concept 'twater'.

One might ask how much of what I have laid down in the preceding two paragraphs is terminological, and how much of it is substantively committing. In particular, the idea that thoughts can be thought of as sentences and that concepts can be thought of as words in those sentences might be questioned. (Although the idea that thinking in this way is necessary for our discussion might be questioned might in turn be questioned; it does not seem to be the only way of preserving a distinction between mental representational states and the content of those states.) However, I shall not pursue these issues. After all, any substantive commitments that must be made to make sense of a worry about my knowledge of the content of the vat argument are substantive commitments for my opponents, and I would rather charitably grant these commitments in order to have the prospect of defeating my opponents on their own terms.

2 Knowing the content of the premises

Let us start with the premises of the vat argument. It has been suggested that, even if the argument is sound, I cannot assume knowledge of its premises unless I assume knowledge of the content of those premises, or of the thoughts that I appeal to in support of those premises, that presupposes knowledge that I am not a BIV. If this is so, the vat argument begs the question against the sceptic, who claims that I do not know that I am not a BIV. There are two objections of this sort that can be made, focusing on premise (1) and premise (2) respectively. The first objection has received quite a lot of attention, and my response to it does not diverge greatly from responses already given. However, my response to the first objection encourages the idea that the bump arises again in the form of the second objection. So far as I know, this objection has not previously been shown to be incorrect.¹⁴

The first objection can be put as follows. It may seem that given the truth of the second premise of the argument:

- (2) If I were a BIV, I could not express the content that I am a BIV,

¹⁴ Brueckner (1994a, b) discusses the second objection, but does not offer an answer to it.

I am in no position to claim knowledge of the first premise of the argument:

(1) I can express the content that I am a BIV.

My reason for thinking that I can express the content that I am a BIV is that I am, right now, thinking a thought with that content. But, under the circumstances, it may seem illegitimate for me to assume that I know the content of this thought. Granted, I am tokening the thought 'I am a BIV'. But, if I were a BIV, this thought would not have the content that I am a BIV. So surely I must know that I am not a BIV *before* I can know that this thought has this content. If this is right, then my reason for thinking that I know premise (1) begs the question against the sceptic by assuming that I know that I am not a BIV.¹⁵

A dialectical problem for any sceptic who makes this objection is that, if I am entertaining the vat sceptic's argument, then I must be able to think a thought with the content that I am a BIV.¹⁶ Thus, if correct, the objection that I cannot know that I am thinking a thought with this content would entail that I cannot know that I am entertaining the vat sceptic's argument. This is an awkward consequence for the vat sceptic, to say the least. So the vat sceptic has a motivation to agree with me that my knowledge of (1) does not presuppose knowledge that I am not a BIV.

Dialectical problems aside, there is good reason to think that my knowledge of (1) does not presuppose knowledge that I am not a BIV. The central thrust of the objection to the vat argument that we are considering is the same as that of an old objection to externalism about thought content, of the sort that was invoked in Sect. 1 to motivate premise (2). This objection to externalism about thought content fails, and so the objection to the vat argument also fails.

When externalism about thought content was first proposed, some philosophers suggested that it jeopardises the idea that semantic self-knowledge does not rely on my empirical knowledge of my environment.¹⁷ If whether I can have a thought with the content that water is wet depends on whether I am on Earth or Twin Earth, then how can I know, without relying on my knowledge of my environment, that my current thought has the content that water is wet? Surely, my knowledge that my thought 'water is wet' has this content will depend on my empirical knowledge that I have been on Earth rather than Twin Earth. The objection to the vat argument that we are considering makes the same point regarding my thought that I am a BIV. If my being able to think a thought with the content that I am a BIV depends on whether I am a BIV, then surely my knowledge that my thought 'I am a BIV' has the content that I am a BIV depends upon my knowledge that I am not a BIV.

It is fair to say that there is now a near consensus that the criticism of externalism about thought content, at least as it has just been described, is mistaken. Various ways of responding to it have been proposed, and any of these responses might be adapted to serve as a response to the claim that I am in no position to claim

¹⁵ Jane MacIntyre (1984), Johnsen (2003) and David (1991) make very similar objections, although the details differ since they discuss different versions of the vat argument.

¹⁶ Button (2013) and Brueckner (2006) make the same point.

¹⁷ See for example Boghossian (1989).

knowledge of premise (1). For example, one response emphasises that according to externalism about thought content our environment determines the content of second order thoughts concerning the content of our first order thoughts in just the same way that it determines the content of those first order thoughts.¹⁸

Suppose that I know that I am thinking the second order thought ‘my thought “water is wet” has the content that water is wet’. I know that the concepts involved in the thought that is mentioned are the same as those used to state the content of that thought.¹⁹ Thus, I know non-empirically that what I think about the content of the mentioned thought is true. Externalism about thought content does nothing to disturb this story about how I can have non-empirical knowledge of the content of my thoughts, because it does not open up the possibility that tokens of a concept, employed by the same person at the same time, might be concepts of different things. This is because, according to externalism, my environment will have determined what all tokens of the concept are of in the same way. For example, my interactions with my environment have determined that my concept ‘water’ is of the same thing on both of its occurrences in the second order thought. Thus, it is compatible with externalism about thought content that I have non-empirical knowledge that my thought ‘water is wet’ has the content that water is wet.

The point generalises. Second order thoughts that specify the content of first order thoughts by making use of the same concepts that occur in the first order thought—that is, second order thoughts of the form ‘my thought “s” has the content that s’—are guaranteed to be correct. My recognition of the existence of this guarantee is non-empirical, and my knowledge of whether one of my thoughts is of this form is non-empirical. Thus, it is compatible with externalism about thought content that thoughts of this form can be instances of non-empirical knowledge of thought content. Let us label the non-empirical knowledge of thought content embodied in such second order thoughts *disquotational knowledge of thought content*, since it is expressed by disquotational sentences. Externalism about thought content is not incompatible with the claim that I can have non-empirical semantic self-knowledge, at least of a certain sort, for it is compatible with the claim that I can have non-empirical disquotational knowledge of the content of my current thoughts.

To return to the vat argument, my second order thought ‘my thought “I am a BIV” has the content that I am a BIV’ expresses my disquotational knowledge of the content of my thought ‘I am a BIV’. I need not know anything about my environment, and in particular I need not know that I am not a BIV, to know that this second order thought is true. This disposes of the worry that I must first know that I am not a BIV before I can know that my thought ‘I am a BIV’ has the content

¹⁸ See for example Heil (1988) and Burge (1988) offers a different argument for the compatibility of externalism about thought content and non-empirical knowledge of thought content, and we might equally well have made use of this argument as a response to the current objection to the vat argument.

¹⁹ Note that it is crucial that the second order thought is ‘my thought “water is wet” has the content that water is wet’ if I am to know this. It is my knowledge that it is *me* stating the content of one of *my* thoughts that allows me to know that my statement of the content makes use of the same concepts that are involved in the thought whose content I am stating.

that I am a BIV, and thus that I cannot know (1) without first knowing that I am not a BIV. Of course, my knowledge that I am thinking a thought with the content that I am a BIV, in conjunction with what I know about externalism about thought content, does *entail* that I am not a BIV. That is the point of the vat argument. However, the fact that one proposition that I know in conjunction with other things that I know entails another proposition does not mean that knowledge of the first proposition depends upon knowledge of the latter.²⁰

Although externalism about thought content is compatible with the claim that I have disquotational knowledge of the content of my current thoughts, it has been suggested that externalism about thought content entails that I cannot always have non-empirical knowledge of the *comparative content* of my thoughts.²¹ Knowledge of the comparative content of two thoughts is knowledge of whether those thoughts have the same content. Such knowledge is stronger than disquotational knowledge of content, in the sense that, as we are about to see, I may have disquotational knowledge of the content of two thoughts, whilst lacking comparative knowledge of their content.

Given externalism about thought content, we seem to be able to construct the following case in which I lack knowledge of the comparative content of my thoughts.²² Suppose that I am an Earthling who has the concept 'water', and that this is a concept of water, which is of course H₂O. Nonetheless, I do not know that water is H₂O, although I do have enough understanding of chemistry to grasp the basic idea of molecular structure. I meet a philosopher who tells me about a place called Twin Earth. She tells me that wherever there is water on Earth, there is something with the molecular structure XYZ on Twin Earth. By convention, we use the concept 'twater' in our discussions to refer to XYZ. As we have seen, I may have non-empirical disquotational knowledge that my thought 'water is wet' has the content that water is wet, and that my thought 'twater is wet' has the content that twater is wet. But can I have non-empirical knowledge that my thought 'water is wet' has different content from my thought 'twater is wet'? Plausibly, the answer to this question is no. In order to find out that the content that water is wet is distinct from the content that twater is wet, I would have to find out that water is not

²⁰ My defence of premise (1) involves claiming that non-empirical semantic self-knowledge is compatible with externalism about thought content. However, McKinsey (1991) claims that, jointly held, externalism about thought content and the view that semantic self-knowledge is non-empirical lead to paradox. This prompts the question of whether there is a difference between the reasoning that leads to McKinsey's paradox and the reasoning of the vat argument, and the question of whether it is possible to solve the paradox whilst preserving the vat argument. These questions are important, but this is not the place to pursue them. See however Wright (2000) and Wright and Thorpe (forthcoming) for an argument that the argument of the McKinsey paradox is vulnerable to a charge of warrant transmission failure which does not afflict the vat argument.

²¹ I take the phrase 'knowledge of comparative content', along with its definition, from Falvey and Owens (1994).

²² The point is more usually illustrated by "slow switching" cases of the sort described in Falvey and Owens (1994, pp. 111–112). However, the case I give here better serves the coming objection to the vat argument.

twater—something I would know if I found out that water is H₂O rather than XYZ—but this is something that I can only know empirically.

Note that although cases like this may show that externalism about thought content entails that I cannot *always* have non-empirical knowledge of the comparative content of my thoughts, there is no reason to think that it entails that there are *no* circumstances in which I can have non-empirical knowledge of the comparative content of my thoughts. The above case in which I seem to lack comparative knowledge of the content of my thoughts exploits the fact that the disquotational specifications of the content of the thoughts in question make use of different concepts, namely, the concept 'water' and the concept 'twater'. In the case, I know that the concept of 'twater' is a concept of XZY, but I do not know that the concept 'water' is not a concept of XYZ. It will be an empirical discovery that the concept 'water' is not the concept of XYZ. However, suppose that I ask whether my current thought 'water is wet' has the same content as my current thought 'water is wet'. In this case, I employ the same concept, namely 'water', in each thought. As noted above, externalism about thought content gives us no reason to think that 'water' will not be a concept of the same thing in both thoughts. So externalism gives us no reason to think that I cannot have non-empirical knowledge that the content of these two thoughts is the same. This point will be important later.

Falvey and Owens do not think that externalists should be worried that their position is incompatible with the claim that we always have non-empirical knowledge of the comparative content of our thoughts.²³ However, they do think that this incompatibility calls into question the legitimacy of my claim to know premise (2) of the vat argument. They start by arguing that lack of knowledge of comparative content will prevent me from using my knowledge that my thought 'water is wet' has the content that water is wet, and my knowledge that externalism about thought content is true, in a non-empirical demonstration that I am not on Twin Earth. To do so, I would have to know the conditional:

(C) If I were on Twin Earth I could not think a thought with the content that water is wet.

However, for all that I can know non-empirically, my concept 'water' is a concept of XZY because I am on Twin Earth. So, for all that I can know non-empirically, my thought 'water is wet' has the same content as my thought 'twater is wet'; that is to say, for all that I can know non-empirically, my thought 'water is wet' has the content that twater is wet. And I *can* think a thought with the content that twater is wet if I am on Twin Earth. So, for all that I can know non-empirically, (C) is false. If I were to use (C) in my attempted non-empirical demonstration that I am not on Twin Earth, I would be committed to the question begging assumption that I am not on Twin Earth. For this reason, my attempted non-empirical demonstration that I am not on Twin Earth fails.

²³ In fact, as Falvey and Owens (1994, pp. 118–112) point out, there are reasons independent of externalism about thought content to think that I do not have non-empirical comparative knowledge of the content of my thoughts. These reasons are developed in more depth in Owens (1990).

Falvey and Owens' go on to argue that it is question begging for me to assume that I know (2) when I run the vat argument for the same reasons that it is question begging for me to assume that I know (C) in my attempted proof that I am not on Twin Earth. Imagine that I am a BIV, and I token the thought, 'I am a BIV'.²⁴ (Recall that the ellipsis conventions laid down in Sect. 1 hold for underlined text.) Since I am a BIV, the causal constraint ensures that I do not have a concept of what we might call a *real* BIV, that is, a thing made up of grey matter and neurons sitting in a nutrient filled container, and so on. Thus, as a BIV my thought 'I am a BIV' cannot have the content that I am a real BIV. What content would it have?

This depends upon what sort of things I would have concepts of if I were a BIV. Putnam (1981, p. 14) suggests three possibilities: I might have concepts of collections of sensory experiences, or of collections of the electric impulses that cause my sensory experience, or of the parts of the computer program that generate the electrical impulses. It is not obvious which of these options is correct, but we need not settle this issue here. Let us say that if I were a BIV I would have concepts of *illusory* things, leaving it open which of the above three kinds of things these are. What matters for the point that Falvey and Owens wish to make is that if I were a BIV and I thought 'I am a BIV' I would express a thought with the following content:

I always have been and always will be an illusory BIV, stimulated by an illusory supercomputer to have the sensory evidence that I actually have, whilst in fact the world contains nothing except my illusory brain in its illusory vat and the illusory supercomputer.

As before, I will adopt the convention of writing 'I am an illusory BIV' as an elliptical way of rewriting the above block of text, 'I am not an illusory BIV' as an elliptical way of writing the indented text preceded by 'it is not the case that', and 'if I were an illusory BIV' as an elliptical way of writing the indented text preceded by 'if it were the case that'.

Now we can ask: when I, here and now, think 'I am a BIV', do I think a thought with the same content as that of the thought 'I am a BIV'? Falvey and Owens claim that I must know that the content of these two thoughts is not the same in order to know that premise (2) is true. For suppose that, unbeknownst to me, the content of these two thoughts is the same. In that case, even if I were a BIV I would be able to think the former content, because I would be able to think the latter content simply by thinking 'I am a BIV', and these two contents are the same. However, it seems that in order to rule out the possibility that the content of these two thoughts are the same, I would have to know that I am not a BIV. For, if I am in fact a BIV, the content that I express when I think 'I am a BIV' is the same as the content of the thought 'I am a BIV'.

Falvey and Owens conclude that, if I am to know premise (2), I must know that I am not a BIV. Thus, the vat argument begs the question against the sceptic by

²⁴ Following Wright (1992), I shall adopt the convention of underlining thoughts in the language of BIVs.

presupposing that I know (2), just as my attempted non-empirical demonstration that I am not on Twin Earth begs the question by presupposing that I know (C).

This objection is founded upon a mistake. Unlike (C), my knowledge that (2) is true does not depend upon my knowledge of comparative content in the way that Falvey and Owens' suggest. This is because (2) is true whether my thought 'I am a BIV' has the same content as the thought 'I am a real BIV', or the same content as the thought 'I am a BIV'. To see this, suppose first that when I think 'I am a BIV' my thought has the same content as the thought 'I am a real BIV'. That is, my thought 'I am a BIV' has the content that I am a real BIV. In that case, the thought that I have when I think premise (2), namely 'if I were a BIV I would not be able to express the content that I am a BIV', has the content that if I were a real BIV then I would not be able to express the content that I am a real BIV. This is true for the reasons given in Sect. 1.

Now suppose that my thought 'I am a BIV' has the same content as the thought 'I am a BIV'. That is, it has the content that I am an illusory BIV. In that case, my thought 'if I were a BIV, I could not express the content that I am a BIV' has the content that if I were an illusory BIV then I could not express the content that I am an illusory BIV. In that case, my thought would still be true. Whatever exactly illusory BIVs are, they are no more capable of expressing content than are the "people" that I see in my dreams, or when I play single-player computer games. So, if I were an illusory BIV I could not express the content that I am an illusory BIV, for I could not express any content at all.

Thus, my thought 'if I were a BIV, I could not express the content that I am a BIV' is true whether or not the content of my thought 'I am a BIV' is the same as the content of the thought 'I am a BIV'. So I do not need to know whether or not my thought 'I am a BIV' has the same content as the thought 'I am a BIV' to know that what I am thinking when I think premise (2) is true. My knowledge of the premise does not rest on comparative knowledge of the content in the way that Falvey and Owens suggest.

What is the difference between (2) and (C)? I have argued that the thought expressed by (2) is true in either of the two relevant cases: it's true if I am a BIV, and it's true if I am not a BIV. So I don't have to presuppose that I'm not a BIV to know that (2) is true. (C), by contrast, is true in one case and false in the other: it's true if I am on Earth (because in that case it would have the true content that if I were on Twin Earth I could not express the content that water is wet), and false if I am on Twin Earth (because in that case it would have the false content that if I were on Twin Earth I could not express the content that twater is wet).²⁵ So I have to presuppose that I am on Earth in order to know that (C) is true. This is why my attempted non-empirical proof that I am not on Twin Earth fails in a way that the vat argument does not.

What led Falvey and Owens to think that I must have comparative knowledge of content to know that if I were a BIV, I could not express the content that I am a

²⁵ Note that this assumes that 'Twin Earth' is not an externalist concept. This is plausible if, as in our example, 'Twin Earth' is stipulated to be a world covered in XYZ. It would not be so if, as is sometimes the case in the literature, 'Twin Earth' is stipulated to be a world without water.

BIV? My suspicion is that they are tacitly taking it to be a possibility that the content of my thought 'if I were a BIV I could not express the content that I am a BIV' is the same as that of the thought 'if I were a *real* BIV, then I could not express the content that I am an *illusory* BIV'. The latter thought is false, and so if the former thought had the same content as the latter thought it too would be false. However, whatever my situation, my thought 'if I were a BIV then I could not express the content that I am a BIV' does not have the same content as this false thought. Whether my thought 'I am a BIV' has the same content as that of the thought that I am a real BIV, or as that of the thought that I am an illusory BIV, it has the *same* content in both the antecedent and the consequent of my thought 'if I were a BIV I would not be able to express the content that I am a BIV'.

3 Knowing the content of the conclusion

So far, we have seen that I do not beg the question against the sceptic by assuming that I have disquotational knowledge of the content of the vat argument, and that disquotational knowledge of content is sufficient for me to work through the vat argument, thus securing knowledge of its conclusion. Yet the bump threatens to resurface yet again as a worry not about the cogency of the argument, but about whether I can legitimately assume the sort of knowledge of the content of the conclusion of the argument necessary if it is to block vat scepticism. As we shall see, I may.

In his aptly titled paper, 'Of Brains in Vats, Whatever Brains in Vats May Be', Johnsen (2003, p. 238) claims that so long as I do not make the assumption that I am not a BIV, when I run the vat argument it 'establishes only that ... whatever BIVs are, [I am not an entity] of *that* sort'.²⁶ This sounds like the kind of worry we are interested in. But what exactly is the worry? It had better not be that without the assumption that I am not a BIV I cannot have any knowledge of the content of the vat argument's conclusion whatsoever.²⁷ As we have seen, disquotational knowledge of the content of the premises of the vat argument does not presuppose knowledge that I am not a BIV. Likewise, disquotational knowledge of the content of the conclusion of the vat argument does not presuppose knowledge that I am not a BIV. So I do not beg the question against the sceptic by supposing that I know that the thought 'I am not a BIV', established by running through the argument, has the content that I am not a BIV.

In what sense, then, do I not know what a BIV is? Perhaps Johnsen's point can be put in our terminology by saying that, unless I assume that I am not a BIV, I cannot assume a certain sort of comparative knowledge of the content of the conclusion of

²⁶ Emphasis in original. Changed from the first-person plural to the first-person singular.

²⁷ Brueckner (1986, p. 167) raises an objection like this, claiming that all that the vat argument can give me is 'the metalinguistic knowledge that "I am a BIV" expresses a false proposition, rather than the object language knowledge that I am not a BIV'. However, this objection is raised against a version of the vat argument which explicitly aims to prove the truth of a certain sentence. I will not defend the version of the vat argument that Brueckner attacks in this paper.

the vat argument. That is, I do not know whether it establishes a conclusion with the content that I am not a real BIV, or with the content that I am not an illusory BIV. To know that my conclusion has the former rather than the latter content, I would require comparative knowledge of content of a sort I could only have if I already know that I am not a BIV.²⁸

How might Johnsen's objection affect the vat argument as a response to vat scepticism? Vat scepticism depends upon the claim that I do not know that I am not a BIV. The vat argument establishes the thought 'I am not a BIV', which seems obviously to be incompatible with the sceptical claim. But it may now seem that, in the absence of forbidden comparative knowledge of content, I cannot know whether the conclusion of the vat argument is incompatible with the claim on which the sceptical argument depends.

However, as we noted in Sect. 2, although externalism about thought content may entail that I do not always have non-empirical comparative knowledge of the content of my thoughts, it does allow for non-empirical comparative knowledge of the content of thought tokens of the same type had at the same time. So I can have non-empirical knowledge that the content of the thought that I negate when I think 'I am not a BIV' is the same as the content of the thought 'I am a BIV'. Thus, I do not beg the question against the sceptic by assuming that the conclusion of the vat argument is incompatible with the claim that I am a BIV.

The bump may now seem to arise again. It is tempting to think that, although I have succeeded in showing that I am not a BIV, this might not be the possibility that I should have been aiming to rule out in the first place. One might try to motivate this worry by describing in the third person a BIV who has all the same sensory experiences that I do, and when I think 'I am a BIV', it thinks 'I am a BIV'. Let us suppose again that it thereby expresses the content that it is an illusory BIV. However, the BIV does not have the comparative knowledge that its thought 'I am a BIV' does not have the same content as the thought 'I am a BIV'.

The BIV runs through a vat-style argument. By this I mean that it runs through an argument that it expresses with the same words that I have used to express the vat argument in this paper, including the first person pronoun 'I', whilst having the same experiences that I have when I think through the vat argument. Let us suppose that the BIV succeeds in using this argument to establish a conclusion that it expresses by thinking 'I am not a BIV'. Nonetheless, from a non-envatted perspective, this performance seems unimpressive. The BIV has done what it set out to do: it has established its thought 'I am a BIV' to be false. However, (I can say from my non-envatted perspective) this achievement is unimpressive, for it is compatible with the fact that the BIV is a *BIV*. The BIV can achieve its aims, but the aims that it can set itself are limited. Reflection on the hapless BIV may lead to the idea that I am in the same predicament. Perhaps I have achieved my aims, but my aims are limited.

²⁸ Brueckner (2006, p. 439) replies to Johnsen by asserting that the BIV mentioned in the conclusion of the vat argument is not an illusory BIV. Although Brueckner is no doubt correct, the current point is that it would be question begging for me to presume to know this in the course of a response to scepticism.

However, problems arise when I try to describe the scenario that I may not have successfully ruled out in the first person. The BIV cannot think the thoughts that I recently expressed about how its aims are limited. That is, the BIV cannot think that its thought 'I am a BIV' does not have the content that it is not a BIV. To do so, it would have to have a concept of a BIV, which the causal constraint on concept possession ensures that it does not have. This presents no problem when describing the BIVs predicament in the third person. However, I cannot describe the predicament that I am supposed to be worried that I might be in in the first person, for I cannot specify the content that I am worried that my conclusion might be compatible with. It is in fact part of the putative worry that I cannot express this content, just as the BIV cannot express the content that it has failed to rule out. So I cannot say in what way I am worried that my aims might be limited.

So far we have seen that, on the assumption that the vat argument is sound, I can employ it to gain knowledge that I am not a BIV, and so the claim on which vat scepticism depends, namely, that I do not know that I am not a BIV, is false. Moreover, we have seen that I cannot specify some other scenario that the sceptic might employ in her argument, and that the conclusion of the vat argument may be compatible with. At this point the sceptic might claim that I can still recognise the *possibility* of a (for me unthinkable) scenario that would falsify my empirical beliefs and that I do not know not to obtain. Her hope is that my mere recognition of the possibility that there is such a scenario will be enough to motivate the sceptical conclusion that I do not have any empirical knowledge.²⁹ Strictly speaking, this would not be vat scepticism, for it does not depend upon the claim that I do not know that I am not a BIV, but the conclusion would be the same.

However, it is unclear how the mere recognition of the possibility that there is a (for me unthinkable) scenario that would falsify my empirical beliefs and that I cannot rule out could be used in a sceptical argument. Remember that the vat sceptic is the less demanding sort of sceptic, who accepts that it is legitimate for me to believe that I have empirical knowledge until she provides me with an argument that I do not. If there *is* a possibility that falsifies my empirical beliefs and that I do not know to be false then perhaps it can be employed in such an argument; if there isn't, then there is no such argument. To claim that there *might be* such a scenario is, at best, to claim that there might be such an argument. But to claim that there might be such an argument does not undermine my belief that I have empirical knowledge. By the lights of the less demanding sceptic I am entitled to believe that I have

²⁹ This strategy is put forward on behalf of the sceptic in Nagel (1986) and Pritchard and Ranilli (2016). Wright (1992, p. 93) also gives voice to the worry that there might yet be some scenario that I cannot entertain when he says that the 'real basis' for dissatisfaction with the vat argument is that 'the real spectre to be exorcised concerns the idea of a thought *standing behind* our thought that we are not brains-in-a-vat, in just the way that our thought that they *are* mere brains-in-a-vat would stand behind the thought—could they indeed think anything—of actual brains-in-a-vat that "we are not brains-in-a-vat"'. However, it should be noted that the thought with which Wright is concerned is not one which would falsify most of my empirical beliefs. Rather, it is a thought according to which my empirical beliefs are mostly true, and yet I lack the capacity to think about reality as it fundamentally is. Thus, whatever the threat that the possibility of the thought with which Wright is concerned might present, it is not a threat to the claim that I have empirical knowledge, as the (putative) thought with which we are here concerned is intended to be.

empirical knowledge until an argument against this belief is given. Her motto is “innocent until *proven* guilty”, not “innocent until we recognise the possibility of a proof of guilt”.

4 Conclusion

My strategy has been to carefully describe the knowledge of the content of the vat argument that I need to employ it as a response to vat scepticism, and to show that it is legitimate for me to assume that I have such knowledge. In Sect. 2 we saw that I do not beg the question against the sceptic when I assume that I have disquotational knowledge of content. Such knowledge suffices for me to have knowledge of the premises of the vat argument. It may be true that I cannot assume a certain sort of comparative knowledge of the content of the vat argument's premises. However, knowledge of the premises of the argument does not require comparative knowledge of content. In Sect. 3 we saw that I have sufficient knowledge of the content of the conclusion of the vat argument for it to provide a response to vat scepticism. Again, it may be question begging for me to assume that I have a certain sort of comparative knowledge of the content of the argument's conclusion, but again, I do not require such comparative knowledge of the conclusion in order to use it in a response to vat scepticism.

I conclude that I need make no problematic assumptions about my knowledge of the content of the vat argument when using it as a response to vat scepticism.

Acknowledgements Thanks to Crispin Wright, Peter Sullivan, Adrian Haddock, David Horst and Bernhard Salow for helpful discussion. This paper benefited greatly from being presented at a Philosophy Work in Progress Seminar at the University of Campinas, a Knowledge Beyond Natural Science Project Seminar at Stirling University, and a Philosophy of Language Group Seminar at Edinburgh University. I would like to thank all three audiences for their comments. An anonymous referee at *Philosophical Studies* also provided comments which have greatly improved the paper. I am grateful for a research grant from the São Paulo Research Foundation (Grant ID No. 2016/03277-1) that allowed me to complete this paper.

References

- Boghossian, P. (1989). Content and self knowledge. *Philosophical Topics*, 17, 5–26.
- Brueckner, A. (1986). Brains in a vat. *The Journal of Philosophy*, 83(3), 148–167.
- Brueckner, A. (1994a). Knowledge of content and knowledge of the world. *The Philosophical Review*, 103(2), 237–343.
- Brueckner, A. (1994b). The Structure of the sceptical argument. *Philosophy and Phenomenological Research*, 54(4), 827–835.
- Brueckner, A. (1999). Transcendental arguments from content externalism. In R. Stern (Ed.), *Transcendental arguments: Problems and prospects* (pp. 229–250). Oxford: Clarendon Press.
- Brueckner, A. (2006). Johnsen on brains in vats. *Philosophical Studies*, 129, 435–440.
- Brueckner, A. (2016). Putnam on brains in a vat. In S. C. Goldberg (Ed.), *The brain in a vat* (pp. 19–26). Cambridge: Cambridge University Press.
- Burge, T. (1988). Individualism and self knowledge. *The Journal of Philosophy*, 85(11), 649–663.
- Button, T. (2013). *The limits of realism*. Oxford: Oxford University Press.
- Byrne, A. (2004). How hard are the sceptical paradoxes? *Nous*, 38(2), 299–325.

- Christensen, D. (1993). Skeptical problems, semantical solutions. *Philosophy and Phenomenological Research*, 53(2), 301–321.
- Cohen, S. (1998). Two kinds of skeptical argument. *Philosophy and Phenomenological Research*, 58(1), 143–159.
- David, M. (1991). Neither mentioning “brains in a vat” nor mentioning brains in a vat will prove that we are not brains in a vat. *Philosophy and Phenomenological Research*, 51, 891–896.
- Falvey, K., & Owens, J. (1994). Externalism, self knowledge, and scepticism. *The Philosophical Review*, 103(1), 107–137.
- Heil, J. (1988). Privileged access. *Mind*, 97(386), 238–251.
- Johnsen, B. C. (2003). Of brains in vats, whatever brains in vats may be. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 112(3), 225–249.
- Locke, J. (1960). *An essay concerning human understanding*. (P. H. Nidditch, Ed.) Oxford: Oxford University Press.
- MacIntyre, J. (1984). Putnam’s brains. *Analysis*, 44, 56–61.
- McKinsey, M. (1991). Anti-individualism and privileged access. *Analysis*, 51(1), 9–16.
- MgGinn, C. (1977). Charity, interpretation and belief. *Journal of Philosophy*, 74, 521–535.
- Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Owens, J. (1990). Cognitive access and semantic puzzles. In C. A. Anderson & J. Owens (Eds.), *Propositional attitudes: The role of content in logic, language and mind* (pp. 147–174). Palo Alto: CSLI Press.
- Pritchard, D., & Ranilli, C. (2016). Putnam on BIVs and radical scepticism. In S. C. Golberg (Ed.), *The brain in a vat* (pp. 75–89). Cambridge: Cambridge University Press.
- Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy*, 70(19), 699–711.
- Putnam, H. (1975). The meaning of ‘meaning’. In K. Gunderson (Ed.), *Language, mind, and knowledge* (Vol. 2). Minnesota: Minnesota University Press.
- Putnam, H. (1981). *Reason truth and history*. Cambridge: Cambridge University Press.
- Thorpe, J. R. (2017). Closure scepticism and the vat argument. *Mind*, published electronically; hard copy forthcoming.
- Tymoczko, T. (1989). In defense of Putnam’s brains. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 57(3), 281–297.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Wright, C. (1992). On Putnam’s proof that we are not brains-in-a-vat. *Proceedings of the Aristotelian Society*, 92(1), 67–94.
- Wright, C. (2000). Cogency and question-begging: Some reflections on McKinsey’s paradox and Putnam’s proof. *Philosophical Issues*, 10, 140–163.
- Wright, C., & Thorpe, J. R. (forthcoming). Putnam’s proof revisited. In J. Conant, & S. Chakraborty, *Engaging Putnam*. Cambridge, Massachusetts: Harvard University Press.