Could i conceive being a brain in a vat?
John D. Collier
University of Calgary,

To cite this Article Collier, John D.(1990) 'Could i conceive being a brain in a vat?', Australasian Journal of Philosophy, 68: 4, 413 — 419
To link to this Article DOI: 10.1080/00048409012344411
URL: http://dx.doi.org/10.1080/00048409012344411

PLEASE SCROLL DOWN FOR ARTICLE
COULD I CONCEIVE BEING A BRAIN IN A VAT?

John D. Collier

Hilary Putnam [11, 12] considered whether, in the light of contemporary theories of meaning, the theory that we are brains in vats could be true. His conclusion that it could not has caused consternation among realists, since his conclusion implies that an ideal theory satisfying all of our methodological and theoretical constraints could not be false. So far there has been no really effective reply. Most attempted refutations have either tried to show that Putnam's argument is unsound, or else that one or more of his assumptions is unjustified. I will accept here that Putnam's argument is sound, but argue that even then his conclusion allows a robust realism. This realism is relativistic in much the same way as Putnam's internal realism, but it allows the possibility that an ideal theory could be radically wrong about the world.  

My strategy will be to distinguish between our ability to state a theory and our ability to conceive its possibility. I will argue that even given Putnam's assumptions, and despite Putnam's arguments, we can conceive being a brain in a vat, and that this is all we need to retain a robust realism. Conceivability, not statability, is required by metaphysical realism.  

Putnam's conclusion is part of his larger argument against metaphysical realism. 2 Metaphysical realism involves 1) a correspondence theory of truth, 2) bivalence of properties, and 3) the existence of a world independent of our representations. Its important consequences are that truth is radically non-epistemic, and that an ideal theory might be false. Although traditional metaphysical realism has opposed relativism, there is nothing intrinsic to Putnam's characterisation of it that rules relativism out. In particular, the correspondence theory of truth allows that different representations can correspond to the same reality, as long as the mode of their correspondence is different.  

Conceivability versus Statability  

Conceivability is a second-order capability that allows us to consider possible

---

1 I am grateful to Ron Clark, Dave Davies, Charlie Martin, Bill Demopoulos, Steve Savitt, Dick Grandy and an anonymous referee for helping me to avoid oversimplifying this issue. Final work on this paper was done during the tenure of an SSHRCC Canada Research Fellowship at the University of Calgary.

2 The rest of the argument considers the alternative to a causal theory of meaning that supposes that the structure of our concepts determines their reference. I will not deal with this part of the argument here. For a more complete account, see J. D. Collier [3].
Could I Conceive Being a Brain in a Vat?

evidence that might lead us to change our minds about the truth of certain sentences that we would have no grounds to question without the evidence. In some cases, the evidence would require us to revise our views about the truth conditions of these sentences. Conceivability does not require that we can formulate precisely the statements we would arrive at, or their truth conditions. It merely requires that we are able to recognise that new evidence might lead us to re-evaluate the meaning of the sentences in question. This typically happens in revolutionary changes in science in which evidence that is anomalous in the context of the old theory forces us to consider the possibility that our understanding of the world is fundamentally flawed. Recognition of this possibility need not lead directly to resolution of the anomaly. The resolution, when it does come, often involves reconceptions of fundamental concepts, e.g., simultaneity in the transition to relativity theory, or location and measurement in the transition to quantum mechanics.

A theory is conceivable if there is possible evidence that would justify us (or at least allow us) to formulate the theory. It is conceivable that a theory is false if there is possible evidence that could lead us to reject the theory if we were presented with that evidence.\(^3\) One way a theory could be conceivably false is if there were a conceivable theory that, if accepted, would displace the original theory. A theory is inconceivable only if there is no possible evidence that would lead to its acceptance. Only irremediably inconsistent theories or theories that are immune to supporting evidence are inconceivable. The falsity of a theory is inconceivable only if there is no possible evidence that could lead us to reject it and there are no conceivable theories whose acceptance would lead to its rejection.

Putnam's rejection of metaphysical realism hangs on the interpretation of the phrase 'our best possible theory might be false'. If this phrase means that the theory can be stated in circumstances in which it is false, Putnam is correct. If the phrase means merely that we can conceive that our best possible theory is false, Putnam has not established his case against metaphysical realism. The possibility of conceiving being a brain in a vat defeats Putnam's argument, and re-opens the door to metaphysical realism.

Putnam's Argument

Putnam argues that if we were brains in vats, totally isolated from any causal interaction with the rest of the world except for computer generated stimulation producing experiences phenomenally identical to what we do experience, and if meaning is determined by causal relations (or, more generally, any concrete particular relations with the world), the only possible candidates for the reference of our words would be computer generated vat-tokens. Consequently, we could at best express the proposition that we are vat-brains in vat-vats, which of course we would not be.

\(^3\) I use words here like 'allow' and 'could' because the Quine-Duhem thesis and Kuhnian considerations suggest that a theory can be held in the face of any evidence. 'Would' is too strong in the face of this possibility.
An essential part of Putnam's argument is his thesis that referring is something that we do: 'We interpret our language or nothing does' [11, 482]. There is nothing that the vat-brain can do that allows it to detect that it is in anything but a normal world, nor is there anything that allows it to refer to the external world. Consequently, whatever a vat-brain refers to, if anything, must be within the vat-world. Accepting both Putnam's thesis and a causal theory of meaning or some analogue, we encounter the following dilemma: if we were brains in vats, the theory 'I am a brain in a vat' would have to be false. If we are not brains in vats, on the other hand, the theory 'I am a brain in a vat' is clearly false. Given whatever we could mean by the phrase 'brain in a vat', any hypothesis we can formulate that we could really be a brain in a vat leads to an absurdity. It is impossible for me to state truly that I am a brain in a vat, so whatever the statement means, it is not true. The sense in which I cannot be a brain in a vat is the sense in which it is impossible for me to make a true statement that I am. We can be sure, Putnam concludes, that any theory that we can formulate that we are brains in vats is not true.

Putnam appears to draw the further conclusion that we could not be brains in vats. At least, many authors have taken him to have concluded this, and have argued that the conclusion is unwarranted. Whether or not Putnam does draw the stronger conclusion, his arguments don't support it, and he doesn't need it.

Metaphysical realism implies that an ideal theory could be false. For example, a metaphysical realist might hold that we might all really be brains in vats, even if we could never determine this nor have evidence for it. Putnam needs only to establish that any theory we can formulate of the form 'I am a brain in a vat' could not conceivably be true. Since any theory like 'I am a brain in a vat' that we can formulate cannot be true, he seems to have done this.

A Gap in Putnam's Argument
By generalizing Putnam's argument to all other similar cases, it follows directly from the fact that we cannot have a true theory that we are brains in vats that an ideal theory cannot be false. It seems to follow directly from

Lewis [8] holds that the conclusion of Putnam's line of reasoning is so unacceptable that his premise that referring must be something that we do must be false. I suggested elsewhere [3], that Putnam implicitly assumes that an understanding of the world is within our grasp. I would advocate a naturalistic account of meaning that would in no way ensure this. Nonetheless, for the purpose of this paper, I will go along with Putnam's thesis.

See, for example, Jane MacIntyre [9], who argues that Putnam equivocates on different senses of being a brain in a vat. Other authors who argue similarly are Peter Smith [13], J. Harrison [4], Anthony L. Brueckner [1] and Gary Iseminger [5]. See J. D. Collier [3] for further discussion.

David Lewis [8], in a sophisticated if dogmatic paper, tries to nullify Putnam's argument by attacking the causal theory of reference, but the difficulties he raises do not take into account the workings of Putnam's internal account of references. Lewis must come up with something that would undermine Putnam's account of reference.
this that metaphysical realism is false. There is a gap in the argument, however. In order to fill the gap, it must be established that everything that can be conceived could be truly stated in some possible situation. If everything that is conceivable can be stated so that the statement could be true for at least some possible set of circumstances, then it is inconceivable that I am a brain in a vat. If so, the possibility that I am a brain in a vat is absurd.

Putnam's argument fails because we can conceive possibilities that we cannot truly state no matter what situation we are in. Putnam is correct that we cannot truthfully say that we are brains in vats; nonetheless we can conceive being brains in vats. It is therefore conceivable that our best theory of the world could be false, even though we might not be able to truthfully assert that it is false. My argument for this conclusion depends on the contingency of the conditions that limit what we can truthfully assert, and that no empirical theory can conclusively rule out possible evidence that we do not encounter.

The isolation of the brain in the vat is contingent. It could, if circumstances differed, communicate with 'outsiders' via the computer, which would serve as a kind of prosthetic device. It is conceivable that a brain in a vat could be kept in isolation up to a certain time, and then be told that it is taking part in an unusual experiment. We can imagine what it would be like to be such a brain. We can, then, conceive what it would be like to discover that we are a brain in a vat. Although Putnam's hypothetical brain in a vat never makes this discovery, it could if circumstances were different. It cannot formulate the hypothesis that it is a brain in a vat, but it can conceive of evidence that would lead it to radically reinterpret its experience.

The possibility involved is analogous to the possibility of a Kuhnian shift in a paradigm [7], except that it is even more global. Theories accepted in the course of normal science are not corrigible on their own grounds. If they encounter an anomaly—a piece of evidence that resists accommodation by the theory—the only satisfactory way to accommodate the anomaly is to adopt a new paradigm whose hypotheses are absurd within the context of current science. The anomaly allows a transition that would previously have been rejected as inconceivable, or, at the very least, false.

For a previously isolated brain in a vat, being contacted by the 'outside world' would present a remarkable anomaly. Even if this anomaly never occurs, it is conceivable that it could. Accommodation of this anomaly would permit a reinterpretation of the brain's experience to allow it to formulate the hypothesis that it was a brain in a vat. Likewise, I can conceive of circumstances which would force me to radically reinterpret my experience, and allow me to express propositions I have no good reason to believe I can express with my current language. Even if I have all the evidence I could possibly have, and have the best theory I could to account for that evidence, it is still conceivable that there might have been other evidence which, had I been presented with it, would have required me to radically revise my theory of the world. Therefore, it is conceivable that an ideal
theory of the world could be false.

Since the statement

(1) I am a totally isolated brain in a vat

would not be true under any circumstances in which it is considered, it seems strange that it is conceivable that its negation could be false. To resolve this paradox, it is necessary to distinguish between the circumstances that make a sentence true and the circumstances that make the proposition it expresses true.

David Kaplan [6] has pointed out that there are at least two distinct types of logical necessity: analyticity and metaphysical necessity. The two forms can come in all possible combinations. He gives the following example of an analytic but metaphysically contingent sentence:

(2) I am here now.

Kaplan distinguishes between character and content. The character is the standard use of the terms in the statement, whereas its content is the proposition it expresses in its context of utterance. The character is a function from terms and contexts to contents, and the content is a function from possible situations to truth values. Because of the characters of the terms involved, (2) is true whenever it is uttered (barring the possibility that unconventional meanings have been assigned). Yet (2) is clearly not necessarily true, since I can conceive of circumstances under which I might not have been here now, for example, if I had been called to a meeting.

The analyticity of (2) together with its metaphysical contingency depends on the facts that its components are indexicals: their interpretation depends on the context of utterance. Sentences with no indexicals show no distinction between character and content, i.e., they are univocal. Although the terms in (1) (other than the personal pronoun) do not appear to be indexical, it is central to Putnam’s argument that the meanings of the terms and of the whole statement depend on the circumstances in which (1) is uttered. Contrary to appearance, the terms are indexical.

Putnam’s argument establishes that it is analytic that (1) is not true. The necessity of its untruth would follow from its analyticity only if the proposition expressed by (1) is not true in any possible circumstance. In order to determine whether this is so, it is necessary to determine what proposition (1) expresses. Here we are faced with some choices.

Stalnaker [1976, 1978] has argued that the proposition expressed by an indexical statement is normally what he calls the diagonal proposition. This is the proposition that takes as its value in each possible world the situation in that world referred to by the statement if it is uttered in that world. If this is the correct interpretation, then neither (2) nor (1) is contingent. For the case of (2), at least, this flies in the face of common sense; surely I might have been elsewhere. Stalnaker’s rationale for adopting the diagonal proposition comes from three pragmatic conditions he places on rational discourse. If these conditions are violated, the context of the discourse is
Could I Conceive Being a Brain in a Vat?

defective. Non-defective contexts are ones in which the presuppositions of
the statement are the same for all interpreters. Stalnaker asserts that dis-
cussions will usually converge on non-defective contexts. Only the third
condition is relevant here. Stalnaker requires that an utterance must express
the same proposition relative to each possible situation. This condition is
satisfied only if the diagonal proposition is adopted for indexical statements.

If Stalnaker's third condition is violated, the meaning of a given utterance
would depend on the truth of other statements, violating the positivist
condition that truth and meaning must be separate for the purposes of
communication. While this is a reasonable constraint for most situations,
I have argued elsewhere [2] that situations involving anomalies generally
involve defective contexts. This is because the anomalous evidence leads
to (at least the possibility of) a revision in our understanding of the meanings
of our sentences. In this sort of case, the rationale for the positivist constraint
on communication is undermined: the truth of a sentence depends on other
facts. The possibility of conceiving (1) to be true involves just this sort of
situation.

Although the metaphysical possibility of (2) being false seems mundane,
in fact it would be quite shocking to find out that one was not where one
intended by 'here'. Any evidence that gave rise to doubt about the content
of what one intends in uttering (2) would surely be anomalous. A defective
context is created by the anomalous evidence because the presuppositions
in effect when the statement was uttered are not compatible with those
implied by the anomalous evidence. The use of 'here' in (2) would come
to be reinterpreted in the face of the anomalous evidence. I might say
something like 'I am here, but here isn't where I thought it was at all'.

It is my contention that anomalous evidence would create a similarly
defective context for (1), thus it is inappropriate to take the proposition
expressed by (1) in such circumstances to be the diagonal proposition. So,
there are conceivable circumstances in which (1) is analytically untrue, but
metaphysically true. I conclude that it is conceivable that (1) is true.

Although if I am an isolated brain in a vat I cannot explicitly represent
my condition, I can conceive circumstances that would force a reinterpretation
of (1) so that I could come to rightly believe it to be true. In these circumstances
I might say something like 'Well, I am in a normal world, but the normal
world isn't anything like I thought it was'. The conceivability of these
circumstances is not affected by whether they will occur, or even if they
are nomologically possible. No matter what my actual circumstances, I can
conceive being an isolated brain in a vat, though, if I am, I cannot state
this possibility explicitly. Trying to do this would be like trying to express
the idea that I am not where I think I am by saying 'I am not here now'.
What is peculiar about Putnam's case is that almost all terms are indexical,
and we would have to be able to step outside of our circumstances in order
to describe it. This, of course, we cannot do. But this does not mean that
we cannot conceive the possibility.

Metaphysical realism depends not on the possibility of the best possible
theory being false, but on the conceivability of its being false. Since scientific theories are always falsifiable, even if true, they are always open to anomalous evidence. This is true even of our ideal theory, that has satisfactorily accommodated all the evidence available to us. We can never completely close off the possibility of discovering a new anomaly. Since anomalous evidence can make the context of a theory defective, even our ideal theory is open in principle to radical modification. Although we cannot rationally believe that our ideal theory is false, because there are no conditions under which we could consider it in which it would be false, it is conceivable that it could be false. The correct interpretation of the phrase ‘our ideal theory might be false’ is that it is conceivable that our ideal theory is false. Metaphysical realism is not defeated by Putnam’s brain in a vat argument.

Both Putnam and many of his critics went wrong in assuming that in order to conceive the possibility that a theory could be false we must consider the truth or falsity of the theory. If this is required, then Putnam’s argument is without fault. Fortunately for metaphysical realists, it is possible to consider what might lead one to find a theory false, without considering directly whether or not the theory is false. Therefore it is possible to conceive of the ideal theory being false, even though direct consideration of whether or not the ideal theory is false forces the conclusion that it must be true. This is all the realism that a metaphysical realist could want.

References