

Does Putnam's Argument Beg the Question against the Skeptic?

Bad News for Radical Skepticism

Prof. Dr. Olaf L. Mueller, Humboldt University, Berlin (www.GehirnImTank.de)

TABLE OF CONTENT.

1. The Rules of the Game
 2. The Skeptical Scenario
 3. Neither Disquotation nor Externalism were Introduced for Repudiating Skepticism
 4. A Failing Version of the Argument
 5. Neither Disquotation nor Externalism Presuppose the Existence of Tigers
 6. Two New Premisses
 7. Let's Talk about Computers Instead
 8. Causal Contact to Computers
 9. Externalist Constraints on the Reference of the Term 'Computer'
 10. What if Computers, Cables, Tigers, and Kangaroos are Bits and Bytes?
 11. Logic: Either—Or
 12. Generalizing the Strategy
 13. Metaphysically Impossible or Epistemologically Impossible?
- Appendix
Acknowledgments
Notes
References

ABSTRACT. Are we perhaps in the *matrix*, or anyway, victims of perfect and permanent computer simulation? No. The most convincing—and shortest—version of Putnam's argument against the possibility of our eternal envattment is due to Crispin Wright (1994). It avoids most of the misunderstandings that have been elicited by Putnam's original presentation of the argument in *Reason, Truth and History* (1981). But it is still open to the charge of question-begging. True enough, the premisses of the argument (disquotation and externalism) can be formulated and defended without presupposing external objects whose existence appears doubtful in the light of the very skeptical scenario which Putnam wants to repudiate. However, the argument is only valid if we add an extra premiss as to the existence of *some* external objects. In order to avoid circularity, we should run the argument with external objects which must exist even if we are brains in a vat, e.g. with computers rather than with trees. As long as the skeptic is engaged in a discussion of the brain-in-a-vat scenario, she should neither deny the existence of computers nor the existence of causal relations; for if she does, she is in fact denying that we are brains in a vat.

NOTE. This electronic version of the paper is published here with kind permission of the publishing house Kluwer. It differs from the printed version stylistically, typographically, and with respect to lay out. Section headings and one note are added. The arguments are unchanged.

Does Putnam's Argument Beg the Question against the Skeptic?

Bad News for Radical Skepticism^{*)}

1. *The Rules of the Game*

Whenever you introduce a new philosophy student or a layman to Putnam's celebrated argument that we could not always have been brains in a vat,¹ she will spontaneously reply that the argument begs the question against the skeptic. This charge can take many different forms.

One particularly uninteresting version of the charge relies solely on challenging whatever premiss is put forward by the anti-skeptic, cf. Putnam (1994a, 284). A more interesting charge of begging-the-question would have to satisfy two constraints.

First, the skeptic should specify, and keep constant, the skeptical scenario about which she is talking, and which she expects us to repudiate. (For example, in a discussion about the existence of the *past*, it would not be fair to suddenly charge the anti-skeptic only because his reasoning depended on, say, the assumption that there exist external things *right now*).

Secondly, the skeptic should explain in precisely what way the anti-skeptic's premisses beg the question against an argument invoking the very skeptical scenario under discussion. So it is not enough to reply to a given premiss, *How do you know it to be true?* Rather, the skeptic is obliged to show that the anti-skeptic is not entitled to use his premiss within the dialectical situation engaging the two opponents. In order to show this, the skeptic must make plausible the charge that it is the very skeptical scenario under discussion which prevents us from knowing the accused premiss.

2. *The Skeptical Scenario*

Let us see how the general requirements from the preceding section look when the skeptic wishes to convince us that Putnam, in his argument, begs the question against skepticism. As

to the first of our two requirements, the skeptic has little choice. Putnam's argument is designed to prove that we could not have been envatted from the beginning of our lives. As Putnam's argument does not concern itself with permanent dreams, evil demons and the like, the skeptic cannot use these more traditional (Cartesian) hypotheses to launch her charge against Putnam's argument. As long as Putnam's argument is being criticized for begging the question, there is but one skeptical hypothesis we need fear: that is, the hypothesis that we have always been brains in a vat.²

Often overlooked is that it already follows that, while evaluating Putnam's argument, we need not worry about the *existence* of the external world: According to the brain-in-a-vat scenario, there are external objects, if only a few, cf. Putnam (1998, 256). According to the scenario there are, for example, brains and vats. Thus, in obeying our first requirement, the skeptic should not complain that Putnam's premisses presuppose the existence of an external world.

Let us, for the remainder of the paper, restrict our attention to the following scenario:

(S) The external world is almost empty. There exist only four distinct objects (and, of course, their parts plus the mereological sums thereof): One computer, one brain, one vat of nutrients, and one cable. These four objects do not overlap. They are arranged as follows: The brain is placed in the vat and connected to the computer by means of the cable. The subject's sensory impressions are identical to (supervene on / are nomologically related to—or what have you³) brain processes caused by the connected computer.

Our second requirement, then, leads to the following question: Does Putnam's argument succeed in repudiating (S) without appealing to premisses that are doubtful, as long as (S) has not been ruled out beforehand?

3. Neither Disquotation nor Externalism were Introduced for Repudiating Skepticism

In order to answer the question from the preceding section, we need to have a closer look at Putnam's argument. The argument employs two philosophical principles that must be checked: disquotation and externalism. Before going in detail, however, I'd like to point out why the skeptic has no good *a priori* chance of convincing us that an appeal to disquotation and externalism amounts to begging the question against the skeptic. The reason for this has to do with the history of the two principles. Neither of them was designed specifically for the

discussion with the skeptic. Rather, they have been employed for a number of years without any anti-skeptical motive. Let me explain this first for the older of the two principles.

It was Alfred Tarski who first formulated disquotation principles for truth and reference in the course of his epoch-making explication of truth. He explicitly stated that his explication is metaphysically neutral.⁴ But if the disquotation principles are metaphysically neutral, they won't tell us anything about the situation in which we are stuck. They won't tell us, for example, whether our sensory experiences derive from Putnam's famous computer or, less exotically, from ordinary sense organs. Thus it will be quite hard for the skeptic to make us believe that we presuppose too much when we appeal to Tarskian disquotation principles.

A similar point can be made with externalism. The externalist doctrine was also not designed for the battle with the skeptic, but has emerged from a debate about the reference of proper names and natural kind terms, from a debate, that is, which originally did not seem to have any impact on the issue of skepticism.⁵

The history of disquotation and externalism being thus, it came as a surprise when in 1981 Putnam conjoined the two principles for his well-known attack on the brain-in-a-vat scenario. Putnam seemed to have *discovered* an anti-skeptical potential in the two principles which no one had expected to be there. I think this discovery was a surprise for Putnam, too. But while the entire history of the argument does not look much like the history of an argument designed to beg any question, the argument could still beg the question, as it were, by mistake. In the end, I am going to show that this is not so. Let us first, however, consider a version of the argument which does indeed fail because it presupposes too much.

4. A Failing Version of the Argument

In the version I have alluded to at the end of the preceding section, the argument runs thus:

- (1) In the language I am actually speaking right now, the word 'tiger' refers to tigers.⁶
- (2) In the language of an envatted brain, the word 'tiger' does not refer to tigers.
- (3) Therefore, the language which I am actually speaking right now is different from the language of any envatted brain. (From (1) and (2)).
- (4) Therefore, I am not a brain in a vat. (From (3)).⁷

It is true, the two premisses in this argument do not depend on philosophical principles other than disquotation (premiss (1)) and externalism (premiss (2)). But while it seems initially plausible that we can formulate and defend disquotation and externalism without begging the question against the skeptic (see below), it is obvious that in the above argument these principles are used in a way which does presuppose too much. As it stands, the argument presupposes the existence of tigers.

To see this, let us imagine there are no tigers. In such a world we would have to face difficult philosophical problems as to the truth values of our premisses. You might say that they would have to be either false or senseless; but for the argument to succeed we need true premisses. Premiss (2), being a negative statement, might even be true with no tigers. (It would, then, parallel sentences such as "The word 'four' does *not* refer to witches" or "There are *no* witches"). But this line of thought cannot be extended to the first premiss, which does not contain a negative statement.

It doesn't help to read premiss (1) without existential import. That is, it doesn't help to allow for its truth, even in a tigerless world, by way of interpreting the premiss as follows:

(1') In the language I am actually speaking right now, the word 'tiger' refers to all and only tigers,

which would be true if there were no tigers (because then we could say: "In the language I am actually speaking right now, the word 'tiger' refers to all and only tigers, *namely to nothing*"). This move might save the argument's first premiss, but it cannot save the argument. It cannot save the argument because under this interpretation, a word like 'tiger', which does *not* refer to tigers in a given language (of an envatted brain, for example), might also refer to all and only tigers, *namely to nothing*. (We are still assuming that there are no tigers). Therefore, our second premiss no longer suffices in distinguishing my own language from the language of an envatted brain: in both languages the word 'tiger' might lack referents. And thus, the third step in our argument might fail to be correct—which means that it does not follow from the two premisses.

To be sure, as soon as we have a tiger on our hands, the two premisses (1) and (2) guarantee a difference between my own language and that of an envatted brain. But this shows only that

for the argument to go through we must assume the existence of tigers. And while such an assumption might be readily granted in a discussion among, say, biologists, the assumption is not innocent in our discussion with the skeptic. This illustrates nicely how the skeptic can fulfill the second of our two requirements for launching a successful charge to the effect that the question has been begged (section 1): If the skeptical scenario (S) were to be true, we would, sadly enough, inhabit a tigerless world; therefore we cannot assume the existence of tigers in our attempt to repudiate (S); but precisely this we had to assume in the course of our reasoning. Therefore, the argument we have been considering does indeed beg the question, as the skeptic wants to have it.

5. Neither Disquotation nor Externalism Presuppose the Existence of Tigers

The argument from the preceding section fails in a manner that is instructive, revealing that we shouldn't despair of the two philosophical principles we have been using in the course of the argument. As we'll see in a moment, both disquotation and externalism can be applied to the word 'tiger' without any presupposition as to the *existence* of tigers. If this is so, the only problem with the argument is having to appeal to an extra-premiss which says that there are tigers. This extra-premiss has nothing to do with disquotation and externalism. Before we see how to run the argument without the extra-premiss, I want to demonstrate that it is not built in to disquotation or externalism, even when we apply these doctrines to the word 'tiger'.

As we have seen, there is a reading of premiss (1) without existential import; but because it is all too tempting to smuggle in unwarranted assumptions as the argument goes on, we should explicitly exclude any zoologico-ontological commitment:

(D) If, in the language I am actually speaking right now, the word 'tiger' refers to something, then it refers (in that language) to tigers.

Obviously, such a principle can be maintained without any ontological commitment as to the existence of tigers. What is more, the principle can be known *a priori*. We need reflect only on our notion of reference, and on the way we use quotation marks, in order to see that (D) must be true.⁸

Let us follow a similar line for externalism:

- (E) For a word in a given language to refer to tigers, its user must have been in appropriate causal contact with at least one tiger.⁹

When articulated in such a way, externalism does not presuppose too much. For instance, (E) could be true even if there were no tigers (and had never been any). In this case, the principle says that no speaker can refer to tigers. Thus the ontological commitments of the principle's *truth* are small; but what must we presuppose in *defending* the principle? Again, not much. We can, I submit, defend the principle on purely *a priori* grounds. For this we first need to reflect what kind of objects we take tigers to be: *If* they exist, then they must be concrete objects located in space and time, rather than abstract objects like sets or numbers. Second we should realize that there are causal constraints on any reference to concrete objects; in other words, we have to rule out magical theories of reference. Such theories can be, as we know, ruled out by performing thought experiments alone. (Just *imagine* several twin-earth scenarios, and you'll see that our *doppelgangers* from twin-earth cannot refer to what we call 'water'...¹⁰)

One may protest that in the course of such a defense we must assume at least that there are material objects, capable of standing in mutual causal relation. But this is not so, for we can defend our principle hypothetically. We can say: No reference to material objects without causality. This leaves it open as to whether there really is such a thing as causality.¹¹ (It also leaves open whether we can and do refer to material objects). Admittedly, there may be no conceivable alternative to a world of material objects which is causally organized.¹² But even if there is no conceivable alternative, our defense of the principle still does not *depend* on the assumption of a causally structured world.¹³

6. Two New Premises

Let us cast light on the failure of our original argument by way of putting it into a more explicit form.

- (D) If, in the language I am actually speaking right now, the word 'tiger' refers to some thing, then it refers (in that language) to tigers.
(E) For a word in a given language to refer to tigers, its user must have been in appropriate causal contact with at least one tiger.
(-1) There are tigers.

- (0) Envatted brains lack appropriate causal contact with tigers. (Follows from how the situation of an envatted brain was introduced in (S), cf. section 2).
- (1) In the language I am actually speaking right now, the word 'tiger' refers to some tigers. (From (D) and (-1)).¹⁴
- (2) In the language of an envatted brain, the word 'tiger' does not refer to tigers. (From (E) and (0)).
- (3) The language which I am actually speaking right now is different from the language of any envatted brain. (From (1) and (2)).
- (4) I am not a brain in a vat. (From (3)).

When put in this form, our argument suffers from two defects. The first one is harmless and can be fixed by way of introducing an additional premiss; the second one is insuperable in the discussion with the skeptic: it will force us to reorganize the whole argument (section 7). Let us inspect the defects in turn.

The first problem with our argument in its present form is that the transition from (D) and (-1) to (1) is not deductively valid; it is not fully explicit. In order to render the transition valid we have to appeal to another instance of disquotation:

- (A) There are tigers if and only if, in the language I am actually speaking right now, the word 'tiger' refers to some thing.

The reasons for accepting (A) are of course the same as those which convinced us of (D), see section 5 above; premiss (A) also follows from reflections on our notion of reference and on the way we use quotation marks.

Still, one might want to criticize (A) along the following lines. On the left-hand side of the biconditional in (A) we speak about a certain *ontological* issue, viz., the existence of tigers; and on the biconditional's right-hand side we speak about *linguistic* issues, viz., the reference of a certain word from our language. But how can the zoologico-ontological issue depend upon facts about language? After all, the existence of tigers does not and should not presuppose people who can speak. (There could be tigers in a world without speakers).

That's right. But our additional premiss does not imply any such presupposition. The premiss states an innocent (merely material) biconditional, whose truth value can be calculated according to the method of truth tables. Its truth does not require any metaphysically or

ontologically *necessary* connection between its subsentences. What I claim is merely that the biconditional's truth can be seen on purely *a priori* grounds.

To defend this claim we have to do two steps. First, let us assume that there are tigers. As I have *formulated* this assumption by way of making use of my own language, I have used a certain word, that is, the word 'tiger'. Using this very word I have spoken about tigers (which exist, according to the assumption). Therefore, I have used a word which refers to something. So, on purely *a priori* grounds we have seen the truth of the following conditional:

(A_→) If there are tigers then, in the language I am actually speaking right now, the word 'tiger' refers to some thing.

Second, let me assume that there exists some object to which the word 'tiger' refers (in the language that I am actually using right now). So when I say 'tiger', then there exists something which bears that biological name. Let me do this now, let me ask (for example): *Are there tigers?* The answer is to the positive, of course, because in formulating the question I have used the very word whose referents do exist, according to our assumption. So the following conditional is an *a priori* truth, too:

(A_←) There are tigers if, in the language I am actually speaking right now, the word 'tiger' refers to some thing.

Taking the two steps together we arrive at the apriority of premiss (A).¹⁵

7. Let's Talk about Computers Instead

Let us add premiss (A) to the list of our premisses. This makes the argument valid; and the only reason we cannot be sure about the truth of its conclusion lies in the fact that premiss (-1) is dubious in the light of the skeptical scenario which the argument is supposed to repudiate. How can we run the argument without that premiss? Well, *tigers* are not essential for the argument. If we run the whole argument with kangaroos, the argument is no less valid. Thus we can avoid the dubious appeal to tigers' existence—appealing to kangaroos instead. But while it is encouraging to realize that we can do without tigers, it is certainly not better to rely

on kangaroos. The whole zoo's existence is dubious in the light of the skeptical scenario we wish to refute.

We must run the argument with entities which are not to be doubted because of (S). What entities could those be? They would have to be material objects, because otherwise there wouldn't be externalist constraints upon reference to them. And they would have to exist even under the conditions of our skeptical scenario (S). This gives us the clue. Even according to (S) there are some material objects: a brain, a vat, a computer, and a cable. So let us try the argument with computers in place of animals:

- (A*) There are computers if and only if, in the language I am actually speaking right now, the word 'computer' refers to some thing.
- (D*) If, in the language I am actually speaking right now, the word 'computer' refers to some thing, then it refers (in that language) to computers.¹⁶
- (E*) For a word in a given language to refer to computers, its user must have been in appropriate causal contact with at least one computer.
- (-1*) There are computers.
- (0*) Envatted brains lack appropriate causal contact with computers. (Follows from how the situation of an envatted brain was introduced in (S)).
- (1*) In the language I am actually speaking right now, the word 'computer' refers to some computers. (From (D*), (A*), and (-1*)).
- (2*) In the language of an envatted brain, the word 'computer' does not refer to computers. (From (E*) and (0*)).
- (3) The language which I am actually speaking right now is different from the language of any envatted brain. (From (1*) and (2*)).
- (4) I am not a brain in a vat. (From (3)).¹⁷

Of course, the skeptic remains unimpressed. Once more she wants to accuse us of having begged the question. But while it was easy to direct such a charge against the predecessors of our present argument, now the skeptic has a more difficult task. Clearly the existence of tigers (or kangaroos) is incompatible with scenario (S), which we are out to refute. But the existence of computers is more than just compatible with the scenario; computers *must* exist if the scenario is to hold.

8. Causal Contact to Computers

Still, there is a number of objections the skeptic might come up with in order to maintain her charge. Let us inspect her objections and begin with the less convincing ones.

First objection. While an envatted brain indeed lacks causal contact with tigers (as claimed by (0) in the preceding argument), it does have a causal connection to computers; after all, it is a computer that generates the brain's impressions of an external world. Thus it would seem that premiss (0*) fails to hold.

It only seems so. The premiss calls for *appropriate* causal contact with computers. True, there is *some* causal connection between the envatted brain and the simulating computer. But clearly this does not suffice for reference.¹⁸ The brain does not, and cannot possibly, perceive or manipulate the computer. (For example, the brain cannot tell the computer's colour; neither can it destroy the computer). Rather the brain interacts with certain configurations of bits and bytes within the computer. (And according to externalism, then, the brain can refer to those bits and bytes).

9. Externalist Constraints on the Reference of the Term 'Computer'

Second objection. Even if we were allowed to presuppose the existence of computers, they could still be utterly different from what we take them to be. In particular, it is not clear that reference to computers stands under externalist constraints. This is an objection against (E*).¹⁹ The objection does not deny that we *assume* computers to be material objects and that reference to material objects is causally constrained. Rather, it is worried that by mistake we might have introduced the term 'computer' without supplying the proper *kind* of referents. (In this case, reference to what is called 'computer' need not stand under externalist constraints). What if, for example, our notion of computers derived from hallucination alone? If this were true, our witnessing the ceremony of baptism for the term 'computer' would have been hallucinatory. Even then the term could perhaps refer to *something* (maybe to certain aspects of our hallucinatory images?); but clearly it would not refer to material objects. Doesn't such a possibility cast doubt on our premiss (E*)?

Convincing as this objection may sound within a general discussion of skepticism, it is out of place in a discussion merely concerned with envatted brains. To substantiate her charge, the skeptic resorted to another skeptical scenario (viz., the scenario of consistent hallucination). This violates the first requirement we set up at the beginning of our discussion. As Putnam does not seek to prove impossible that we are merely hallucinating, his premisses should not

be criticized for failing in the case of hallucination. In a discussion of Putnam's argument, the only troublesome circumstances are those which involve envatment.

Now, what if the term 'computer' had been introduced under such circumstances? Well, then it would still refer to material objects; not to real computers, to be sure. It would refer, I think, to those configurations of bits and bytes (within the simulating computer) that are responsible for the computer images the envatted brain enjoys. And the reason for this lies precisely in the externalist doctrine. Thus externalism is applicable to the term 'computer' as long as we confine our attention to the brain-in-a-vat scenario.²⁰

10. What if Computers, Cables, Tigers, and Kangaroos are Bits and Bytes?

Third objection. A moment ago we said that in the language of an envatted brain the word 'computer' refers to certain bits and bytes. And we said—repeating step (2*) from our argument—that this word (in the language of an envatted brain) does not refer to computers. But that presupposes that those bits and bytes aren't computers themselves.²¹ How do we know they are not? In any case, of course, those bits and bytes are proper parts of a given computer. Now, if those bits and bytes were themselves computers, this would imply that some computers are proper parts of some other computers. Why should this not be so? Some mountains are proper parts of some other mountains; some mouthfuls of honey are proper parts of some other mouthfuls of honey. On the other hand, no shoe is a proper part of another shoe; and no tiger is a proper part of another tiger.²² It seems that an empirical investigation is called for before we can tell for sure whether, in this respect, computers are closer to mountains and honey or closer to shoes and tigers. But we cannot be sure of the result of any empirical investigation without first ruling out the brain-in-a-vat scenario. And this means that we cannot maintain step (2*) from our argument without begging the question—says the skeptic in her third objection.

The objection does not rescue the skeptic, however. We know, without empirical investigation, that those bits and bytes which are responsible for the brain's enjoying computer images either are or are not computers. If they are not, Putnam's argument works (or anyway needs to be blocked by a different objection). So let us assume that those bits and bytes indeed are computers, i.e., that they are not only correctly called 'computers' in the brain's language but that they simply *are* computers. By reformulating our reasoning from (D*) to (4) with

cables (instead of computers), we can force the skeptic to apply her objection to the notion of a cable. She must claim, then, that certain configurations of bits and bytes (those responsible for cable images experienced by the brain) indeed *are* cables. Similarly, we force the skeptic into admitting that certain configurations of bits and bytes are brains, and that certain other configurations are vats. A slippery slope: Why stop here? Certain configurations of bits and bytes will turn out to be transistors; others will turn out to be nerve fibers; and still others will turn out to be tigers. Each material object which the envatted brain seems to have only an image of, will really exist (namely in form of certain bits and bytes). Whenever the brain enjoys an image of a kangaroo, there really is a kangaroo (that is, certain bits and bytes responsible for kangaroo images). And if, still impressed by the kangaroo image, the brain exclaims: "Lo, a kangaroo!" it will be right.²³ And this is not an accident: Most of the brain's observation sentences will turn out true.²⁴ It's a little bit odd to call such a situation a skeptical scenario. If kangaroo simulations are kangaroos, then the whole subject of simulation loses its point, and radical skepticism fades away.

It is true, Putnam's argument does not work when we follow the skeptic's third objection so far. But this doesn't matter; the objection turns the whole scenario into something which is epistemically harmless. If the skeptic wishes to draw radical skeptical conclusions she had better come up with another objection.

11. Logic: Either—Or

The skeptic's fourth objection is an attempt to attack premiss (-1*). Admittedly, it is slightly odd to start an anti-skeptical argument with the premiss:

(-1*) There are computers.

Now, what exactly is wrong with this premiss in our dialectical situation? As we saw, the skeptic shouldn't say that the *truth* of the premiss is incompatible with the skeptical scenario under discussion. (She shouldn't say this because that scenario calls for at least one computer). But the skeptic could say: For an anti-skeptical argument to succeed it is not enough to start from *true* premisses, because it is not enough to arrive merely at a true conclusion. What we need and want is *knowledge*; so we must start from premisses known to be true. But we cannot know that there are computers so long as we haven't ruled out the very skeptical

scenario under discussion. The brain-in-a-vat scenario threatens *every* claim to knowledge of external things; it even threatens our ability to gain knowledge of those claims whose truth is called for by the scenario itself.

This is a serious objection which forces me to reorganize the whole argument. It is right, we shouldn't *start* with premise (-1*). We should start with something less controversial, preferably from some logical truth. How about this:

(L) Either there are computers or there are no computers.²⁵

Let us deal with the two possibilities in turn. If there are no computers, then I cannot be an envatted brain whose sensory impressions result from some computer processes; in this case, (S) is clearly false. *If*, however, there *are* computers, then the whole argument we have been considering gains momentum, and (S) is false again. Taking the two cases together we can conclude from (L) that (S) must be false. Even without presupposing the existence of computers, I know that I am not envatted from the beginning of my life.²⁶

12. Generalizing the Strategy

Let us generalize the strategy that has worked so neatly against the skeptic's fourth objection. Suppose the skeptic comes with the objection that Putnam's argument presupposes externalism, which in turn presupposes causality, and thus begs the question. (The skeptic might ask: How do we know about causality?) To this we reply (in addition to what we've already said in section 5): Either there is causality—or there is not.²⁷ If there is no causality, (S) cannot be true because according to that very scenario my sense impressions are *caused* by certain cybernetic processes. But *if* there is causality (second case), then with the argument we sketched above, we get the conclusion that (S) is false.

Or again, suppose the skeptic accuses us of having used premises like (E*), which are doubtful because we might have been merely dreaming from the very beginning.²⁸ Then we reply: Either I have or I have not been merely dreaming from the very beginning.²⁹ In the case of the former, there are no computers, and (S) must be false. If, however, I have not been dreaming all along, I am allowed to employ the premises from our argument and can conclude again that I haven't always been envatted.

To be sure, this does not show that I haven't been *dreaming* all along. It only shows that the skeptic cannot appeal to the hypothesis of false dreams in order to convince me that I don't know whether I have been forever *envatted*. That the skeptic, in our discussion, cannot appeal to false dreams shouldn't surprise us. From the beginning we have made clear that the skeptic should not switch to another skeptical scenario in the middle of the discussion—compare our first requirement (section 1). Now we know more. Even if the skeptic were to ignore the requirement and try to come up with new skeptical scenarios, that are incompatible with the original scenario (S) and with some of our premisses, she could still not maintain her charge that Putnam's argument begs the question. Without any circularity we have been able to show that we haven't been envatted for all time. While this, admittedly, is not enough against the whole bunch of scenarios the skeptic might have in reserve for a more general discussion of skepticism, it is more than nothing.³⁰ It is bad news for the skeptic.

13. *Metaphysically Impossible or Epistemologically Impossible?*

But is it really? The skeptic might try a last rescue. She might be willing to grant that we were able to prove the *truth* of the sentence:

I am not an envatted brain (from the beginning of my existence).

But she might ask: Shouldn't we demand more? Shouldn't we demand a proof that we *cannot possibly* be envatted? —Much depends on how the modal notion from this question is interpreted. If you interpret it metaphysically then I must confess that I am unable to prove that I *cannot* be a brain in a vat. The reason for this is that at least two of our premisses fail to hold by metaphysical necessity. The disquotation principles in (A*) and (D*) exhibit *contingent* features of my language. I *could have been* brought up in a speech community where the word 'tiger' does not refer to tigers but to kangaroos. In fact, of course, I have not been brought up in such a counterfactual (but possible) speech community, *and this I know a priori!* (See again sections 5, 6).

But now we are led to another —epistemological—understanding of the modality involved in what the skeptic wants us to prove:

I *cannot* be a brain in a vat.

The modality involved is an epistemological modality, not a metaphysical one; it is the modality of *a priori* knowledge.³¹

And I claim to have proved my conclusion ("I am not a brain in the vat") on purely *a priori* grounds: The proof is (as I hope to have shown) deductively valid, and all its premisses are *a priori* truths. But because aprioricity is closed under deductive inferences, my conclusion is an *a priori* truth, too. (Its status can be compared to that of the contingent *a priori* truth "I exist").

The skeptic was not interested in metaphysical modalities in the first place. She was busily asking an *epistemological* question: How do you *know* you are not a brain in the vat? My reply is: I know it *a priori*.

Appendix

Here is a version of the argument that accounts for the complication mentioned in note 9 (section 5).³² As premiss (E*) does only hold for terms that haven't been introduced via description, a more accurate version of the premiss runs thus:

(E") A word that is not introduced via description to a certain language can only refer to computers if its user has been in appropriate causal contact with at least one computer.

As we have decided in section 11, we won't use (-1*) as a premiss; instead we introduce it as an assumption:

(-1*) There are computers. (Assumption).

The other premisses remain unchanged:

(A*) There are computers if and only if, in the language I am actually speaking right now, the word 'computer' refers to some thing.

(D*) If, in the language I am actually speaking right now, the word 'computer' refers to some thing, then it refers (in that language) to computers.

(0*) Envatted brains lack appropriate causal contact with computers.

Our first step is not new:

(1*) In the language I am actually speaking right now, the word 'computer' refers to some computers. (From (D*), (A*), and (-1*)).

Now from (E'') and (0*) we *cannot* infer:

(2*) In the language of an envatted brain, the word 'computer' does not refer to computers. (From (E*) and (0*)).

Rather, we have only the following:

(2'') If a word is not introduced via description to the language of an envatted brain, then that word does not refer to computers. (From (E'') and (0*)).³³

But from this and (1*) we cannot immediately conclude that my language is different from that of an envatted brain (as was concluded in the earlier argument); we have to employ an additional premiss:

(B) The word 'computer' has not been introduced via description to the language which I am actually speaking right now.

This premiss might be known introspectively (it expresses structural knowledge about my own language rather than substantial knowledge about the external world); but even if the speaker is not sure about such a premiss when applied to the word 'computer', she can still count on an analogous premiss being applied to some other word denoting material objects different from bits and bytes within the computer.³⁴ For she can infer from (-1*), by *a priori* reasoning, that there have to be some words for material objects (different from bits and bytes within the computer) which are not introduced via description: Even if the word 'computer' is introduced via description, the description itself can ultimately rely only on words not introduced descriptively. And some of the final constituent words in the descriptive definition of the word 'computer' must refer to material objects different from bits and bytes within the computer.³⁵ (The computer cannot merely consist of bits and bytes, and thus, cannot be described merely in terms of bits and bytes).

As soon as we have something like (B) we can go on as usual:

- (3) The language which I am actually speaking right now is different from the language of any envatted brain. (From (B), (1*) and (2)).
- (4) I am not a brain in a vat. (From (3)).

As this conclusion rests on the *assumption* of (-1*), we have so far no more than this:

- (5) If there are computers, then I am not a brain in a vat.

But because the brain-in-a-vat scenario cannot hold without computers, we also know:

- (6) If there are no computers, then I am not a brain in a vat.

From logic we have:

- (L) Either there are computers or there are no computers.³⁶

And this, finally, gives us what we want:

- (7) I am not a brain in a vat. (From (5), (6), and (L)).

Quod erat demonstrandum.

Acknowledgments

The research for this paper has been made possible by a generous grant from DFG (*Deutsche Forschungsgemeinschaft*). I wish to express my gratitude to James Pryor and Hilary Putnam for their patience in numerous discussions about envatted brains during my stay at Harvard. I am grateful to David Hyder for linguistic advice and to an anonymous referee for criticism that has helped me to clarify my argument. I am also very much indebted to Oliver Wachsmuth, Sven Rosenkranz, Tatjana Tarkian, and Wolfgang Carl for years and years of discussion and quarrel over Putnam's argument. Without intention on their part, I believe, it was they who finally convinced me that Putnam is right. As they do not agree, the flaws in this paper should not be blamed on them.

References

- *) After publication of this article, I have formulated a more detailed defense of the proof presented here, see Mueller (2003, chapters III–VI). [Note added in 2011].
- 1 For the details of the brain-in-a-vat scenario and Putnam's argument against it cf. Putnam (1981, 1-21).
 - 2 To be sure, radical skepticism remains an open possibility until we reply to the more traditional (Cartesian) skeptical hypotheses. But if the skeptic wishes to maintain her skepticism because of this, she should not look for hidden circularities in *Putnam's* argument; she should instead enter a different discussion. Putnam's argument might also play a prominent role in this more general discussion. One might want to argue that any skeptical hypothesis either (i) fails to be conceivable, or (ii) lacks the generality needed for yielding *radical* skepticism, or (iii) can be repudiated by an argument parallel to Putnam's. In this paper we shall not be concerned with evaluating the prospects of such an overall approach to skepticism. (But compare Putnam (1998, 256) and Wright (1994, 237); the former contrasts (i) with (iii); and the latter (ii) with (iii)). The original argument, with its limited scope, is interesting enough by itself.
 - 3 Here you may substitute what follows from your preferred answer to the mind-body problem.
 - 4 Cf. Tarski (1944, 363-4). It is true, he wished his explication to be in accordance with a classical correspondence view of truth, cf. Tarski (1935, 5)—which might be interpreted as a certain bias towards realism on Tarski's part. But if Tarski's T-sentences is all we can expect from a correspondence theory of truth, then this theory does not give the realist much. Indeed, a T-sentence like "The sentence 'It snows' is true iff it snows" goes nicely even with idealism!
For the purposes of this paper we need not wonder whether Tarski's definitions succeed in explicating truth or reference. You may well maintain that Tarski's story leaves too much out, cf. Mueller (1998, 47-72). What we need only agree on here is that everything we can infer from Tarski's definitions is indeed true and can be known *a priori* (when, of course, the object language is contained in the meta-language). That much seems uncontroversial.

- 5 Cf. Kripke (1980, 91ff., 122ff. *et passim*); Putnam (1975, 223ff.)—To tell the truth, with regard to Putnam, the history is a bit more complicated. Originally, he combined his externalist reasoning with scientific realism (1975, 235ff.), which does not seem to be a doctrine agreeing well with radical skepticism. However, externalism is not married to scientific realism. After all, Putnam is not a scientific realist any longer; but he is still an externalist.
- 6 Here is how I am going to talk about reference of general terms: A general term refers to an entity x iff x belongs to the term's extension. Thus, the word 'tiger' does not refer to an abstract object like tigerhood.
- 7 This version of the argument is close to the careful reconstruction of Putnam's argument given by Wright (1994, 224, 236-7). Wright does not run the argument with tigers though; on p. 224 he uses the term 'brain in a vat' (which might be good and well as we shall see in sections 7-12); on pp. 236-7 the argument is generalized by way of using any appropriate predicate F . So my appeal to tigers in the argument above seems to be in the spirit of Wright's version of Putnam's argument. As far as I remember, Putnam himself ran the argument with trees when he presented it at the *First Goettingen Philosophy Colloquy* in December 1994; he credited that version to Crispin Wright.
- 8 Notice, however, that the truth of (D) does not hold by metaphysical necessity: I could have been brought up in a speech community in which the word 'tiger' refers to kangaroos. Compare section 13.
- 9 "Causal contact" should be broadly understood here so as to allow for indirect causal contact which is mediated by other speakers who have stood in direct causal contact with tigers. Notice also that the principle provides no more than a necessary condition for reference to tigers; (E) neither defines nor reduces this relation. For simplicity's sake I am going to neglect that (E) does not hold for words that are introduced via description. This complication has been brought to my attention by James Pryor; I shall take care of it in the appendix.
- 10 Cf. Putnam (1975, 223ff.) and Putnam (1981, 3-5, 17-21).
- 11 For another reply to the worry concerning causality see note 13 and section 12 below.
- 12 Cf. Putnam (1998, 239, 244, 247-8).
- 13 And even if it depended on such an assumption, the principle could not be accused of begging the question against the brain-in-a-vat scenario. According to this scenario, the world is causally organized, too. Therefore, the scenario cannot cast doubt on our confidence in causality (cf. section 12).—Notice, in addition, that if there were no conceivable alternatives to a world of material objects organized by causality, then the more traditional (Cartesian) forms of skepticism would not be conceivable either, cf. Putnam (1998, 244).
- 14 The transition from (D) and (-1) to (1) is not deductively valid. We shall see soon that the transition can be rendered valid by means of an additional premiss (A).
- 15 Notice that this does not amount to (A)'s being necessarily true. (Compare section 13 as well as note 8).
- 16 There is an important objection against premisses such as (D*) and its predecessors (D) or (1), an objection which rests upon arguments challenging the assumption that there is a definite reference scheme for each language. Among others, Quine, Davidson, and (with reservations: see end of this note) Putnam are known to have launched such a challenge, appealing to so-called "proxy functions" that make it possible to assign any referent whatsoever to a given predicate. If you want to do this and want to avoid wildly paradoxical consequences, you must of course reassign

suitable referents to the other expressions of the language concerned. (Cf. Quine (1992, 31-33); Davidson (1984, 229); Putnam (1981, 217-8); the roots of such arguments go back to Jeffrey (1964, 82-84)).

The trouble is that those non-standard reference schemes are not only possible for languages other than our own but that the problem "begins at home", cf. Quine (1969, 46-7). To the skeptic this seems to give new ammunition against our premises. However, on a closer look it becomes clear that Quine's, Davidson's and Putnam's arguments against the definiteness of reference schemes need not worry us much. These arguments may be of interest within the philosophy of language; within epistemology they are not. Their main force comes from the observation (first made by Frege (1977, XXII)) that the primary vehicle of meaning is the complete sentence rather than parts of sentences. According to a radical understanding of this context principle we should always try to get along without semantic notions applicable to expressions shorter than sentences. In most of the cases (outside the philosophy of language anyway) this is possible in principle but not advisable in practice. Now there is a way to reformulate Putnam's argument in terms of semantic notions which work on the level of complete sentences: the argument can be run, for example, with the notion of truth instead of reference, or alternatively, with that of sentence meaning. As we are not mainly concerned here with subtleties from the philosophy of language but with epistemological skepticism, I shall not attempt to spell out these alternative versions of Putnam's argument in the present paper. (In *Reason, Truth and History* (1981), admittedly, Putnam seems to come close to a self-contradiction: Whereas in the first chapter he sometimes speaks as if reference (being *definitely not* subject to magical accounts) had to be definite, he gives strong reasons against its being definite in the second chapter and in the appendix of this book; compare for example (1981, 13-4) on one hand with (1981, 32-38 and 217-8) on the other hand. But there is no real contradiction here. In later writings Putnam has made clear that he had been employing the permutation arguments not against the definiteness of reference but as a *reductio ad absurdum* of metaphysical realism and its views on reference, cf. Putnam (1994b, 280-1)).

- 17 If I am right that this version of the argument (using any item from scenario (S)) is okay while its predecessor (using tigers) either isn't valid or starts with a false premise or begs the question, then this casts doubt on Crispin Wright's claim that an argument à la Putnam can show not only that semantically *auto*-disruptive skeptical hypotheses must be false but also that semantically disruptive skeptical hypotheses must be false. Cf. Wright (1994, 236).
- 18 For a similar point see Putnam (1981, 14-5).
- 19 The objection is credited to James Pryor (oral communication).
- 20 This prompts the following question: What if an envatted brain repeated the whole argument as we have reconstructed it so far? The obvious answer is that then the whole argument would have to be reinterpreted along externalist lines. It can be shown that its premises are true in the language of an envatted brain and that they imply literally the same conclusion (i.e., (4)—interpreted in the brain's language, of course). While some people might feel uncomfortable with such an answer, we should, I think, be satisfied with it. It opens the way to a deep understanding of what Putnam's internal realism is all about. Alas, we cannot go into this here.
- 21 This is, I think, what Falvey and Owens might want to say by way of applying their objection to the version of Putnam's proof considered here, see Falvey / Owens (1994, 123-137).

- 22 Let us neglect complications due to pregnancy.
- 23 A similar point is made by Putnam (1981, 14).
- 24 There are some exceptions though. Suppose the brain has the impression of being lost in the Sahara. The brain might say then "There's no fluid wide and far"—which would be wrong, as the brain is still located in a vat of nutrient fluid.
- 25 Admittedly, this is only uncontroversial within classical logic. But even if the skeptic does not want to grant us the law of excluded middle, we can still run the argument. See next note.
- 26 If the skeptic insists on a version of the argument that does not invoke the law of excluded middle (see previous note), we can run the proof as follows. In a first step, we introduce two assumptions (-1*) and (S) (that is, we assume that there are computers and that we are envatted from the beginning of our lives). From these assumptions we derive a contradiction by way of reasoning along the lines of section 7. As the skeptic wishes to maintain (S) as long as she can, she has to deny (-1*). Thus, in a second step, we start with (S) and the negation of (-1*). This is even more easily seen to be a contradiction. ((S) says, among other things, that there is a computer, which the negation of (-1*) denies). Taking the two steps together, intuitionistic logic allows us to conclude that (S) must be false.
- 27 See previous note for a line of thought that does without the law of excluded middle.
- 28 Reminder: *Merely* dreaming that p is defined as dreaming that p while p is false. This is a stronger hypothesis than one in which the dream might give a true picture of reality; and it opens a more convenient way towards skepticism (which could be called "direct strategy", cf. Wright (1994, 218-9, 235)).
- 29 See note 26 in section 11 for a line of thought that does without the law of excluded middle.
- 30 See note 2 (section 2).
- 31 In his original presentation of the argument, Putnam does not make sufficiently clear that what matters is epistemological rather than metaphysical necessity. He complains about "the tendency in our culture [...] to take *physics* as our metaphysics" (1981, 15, Putnam's italics) and contrasts this tendency with his claim: "What rules out [the brain-in-a-vat] possibility is not physics but *philosophy*" (1981, 15, Putnam's italics). A little earlier on the same page he says that the brain-in-a-vat hypothesis is "(necessarily) false". These quotations may sound as if Putnam had wished to prove a metaphysically necessary conclusion. But he also says that he has come to his externalist premiss (i.e., (E*) from our argument) "*by reasoning a priori*" (1981, 16, Putnam's italics). And in his later writings he distanced himself from appeals to metaphysical necessity (such as Kripke's), cf. Putnam (1990, 55, 67 *et passim*).
- 32 The complication itself and the way around it have been brought to my attention by James Pryor, although he might not agree with the details of my presentation. This appendix owes a lot to his subtle refusals to accept Putnam's argument.
- 33 Notice that this step does without any mention of specific words from the envatted brain's language; unlike its predecessor (2*), (2") does not talk about the word 'tiger'. Thereby we avoid a problem which has thus far been ignored: If we wish to differentiate my language from that of an envatted brain by way of spotting referential differences between my word 'tiger' and the brain's word 'tiger', then we presuppose that the words themselves are the same. But this is dubious; an envatted brain can neither write nor talk. Therefore, it is more attractive to differentiate the two languages by way of comparing my word 'tiger' to every word from the brain's language. (2")

- helps us to do this.—The problem and its solution are due to James Pryor (oral communication).
- 34 Then, of course, she would have to reformulate the whole argument using those objects rather than computers.
- 35 You may wonder: Couldn't the word 'computer' be defined in terms of sense-data? To this question, very much in the spirit of Carnap's *Aufbau* (1979), we could reply that Carnap's project was repudiated on purely *a priori* grounds. But even if this is not granted, we still have an answer. If Carnap's project were to succeed some happy day, each sentence describing the state of the external world could be (in principle, anyway) tested by checking sense-data alone. Radical skepticism is not an option in such a picture. So the skeptic had better not revive good old phenomenalist reductionism.
- 36 See note 26 in section 11 for a line of thought that does without the law of excluded middle.

References

- Carnap, R.: 1979, *Der logische Aufbau der Welt*, Ullstein, Frankfurt.
- Davidson, D.: 1984, 'The Inscrutability of Reference', in D. Davidson (ed.), *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford, 227-241.
- Falvey, K. and J. Owens.: 1994, 'Externalism, Self-Knowledge, and Skepticism', *The Philosophical Review*, **103**, 107-137.
- Frege, G.: 1977, *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung über den Begriff der Zahl*, second reprint, Georg Olms Verlag, Hildesheim.
- Jeffrey, R. C.: 1964, 'Logic, Methodology, and Philosophy of Science: Three Reviews. – Sections IV-V', *The Journal of Philosophy*, **LXI** (2), 79-88.
- Kripke, S. A.: 1980, *Naming and Necessity*, Harvard University Press, Cambridge, MA.
- Mueller, O.: 1998, *Synonymie und Analytizität: Zwei sinnvolle Begriffe. Eine Auseinandersetzung mit W.V.O. Quines Bedeutungsskepsis*, Schöningh Verlag, Paderborn.
- Mueller, O.: 2003, *Hilary Putnam und der Abschied vom Skeptizismus oder Warum die Welt keine Computersimulation ist. Wirklichkeit ohne Illusionen, Band 1*, Mentis, Paderborn.
- Putnam, H.: 1975, 'The Meaning of "Meaning"', in H. Putnam (ed.), *Mind, Language and Reality. Philosophical Papers, Volume 2*, Cambridge University Press, Cambridge, 215-271.
- Putnam, H.: 1981, *Reason, Truth and History*, Cambridge University Press, Cambridge.
- Putnam, H.: 1990, 'Is Water Necessarily H₂O?', in H. Putnam (ed.), *Realism with a Human Face*, Harvard University Press, Cambridge, MA, 54-79.
- Putnam, H.: 1994a, 'Crispin Wright on the Brain-in-a-Vat-Argument', in P. Clark and B. Hale (eds.), *Reading Putnam*, Blackwell, Cambridge, MA, 283-288.
- Putnam, H.: 1994b, 'Realism Without Absolutes', in H. Putnam (ed.), *Words and Life*, Harvard University Press, Cambridge, MA, 279-294.

- Putnam, H.: 1998, 'Skepticism', in M. Stamm (ed.), *Philosophie in synthetischer Absicht – Synthesis in Mind*, Klett-Cotta, Stuttgart, 239-268.
- Quine, W. V. O.: 1969, 'Ontological Relativity' , in W. V. O. Quine (ed.), *Ontological Relativity and Other Essays*, Columbia University Press, New York, 26-68.
- Quine, W. V. O.: 1992, *Pursuit of Truth* , revised edition, Harvard University Press, Cambridge, MA.
- Tarski, A.: 1935, *Der Wahrheitsbegriff in den formalisierten Sprachen*, *Studia Philosophica*, Leopoli.
- Tarski, A.: 1944, 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophical and Phenomenological Research*, **4**, 341-376.
- Wright, C.: 1994, 'On Putnam's Proof that We are not Brains in a Vat', in P. Clark and B. Hale (eds.), *Reading Putnam*, Blackwell, Cambridge, MA, 216-241.