

## Refutation of Putnam's Argument against the Possibility that We Are Brains in Vats

### ABSTRACT

Putnam argued (1981) that we can know that we are not brains in vats on the basis of clearly semantic reasoning. I claim that his argument is wrong because some his assumptions are false. Putnam showed that the references of a born in a vat brain's everyday language does not go beyond the computer producing the delusion. So when the brain thinks about being a brain in a vat as being something that he can experience within the delusion, he would never referentially think about the vat which he would be in, but about the computer. Nevertheless, as I will argue, though being a brain in a vat he can think about being a brain in a vat, even referentially, however it requires from him something more than thinking about objects of his usual life. Nothing implies that he could not. Referential considerations are natural part of human thinking and we are not limited to simple references of our direct perceptions. For better understanding the Putnam case I consider and compare three cases: a person abducted to a vat fully aware what happened, a person abducted but unaware, and a person living in a vat since being an embryo.

### KEY WORDS

A brain in a vat, reference, meaning, thinking, evil demon, Putnam

In the present article I am going to argue that Putnam's argument (1981, 1-21) against the possibility that we are brains in vats is wrong. So his defenders are wrong (e.g. Bruckner 1992, 2003, 2004, Johnsen 2003). Putnam's significant observation is that the references of a born in a vat brain's everyday language does not go beyond the computer producing the delusion. So when the brain thinks about being a brain in a vat as being something that he can experience within the delusion, he would never referentially think about the vat which he would be in, but about the computer in some given states. Nevertheless, as I will argue, though being a brain in a vat he can think about being a brain in a vat, even referentially, however it requires from him something more than thinking about objects of his usual life. Nothing implies that he could not.

Retrace carefully the Putnam's reasoning. I will put it in a little changed version for better underlining the most important points. One of the changes is that I consequently take the external perspective on the subject. Internal (first person) perspective is logically equal but it is more easy to make a mistake using it. Especially it is hard to differ the cognitive activity of the subject from his knowing and describing himself. When a subject is described from outside

everything becomes more clear. In fact when one is to describe himself only he can do is to do it as he were outside of himself.

Thus I will consider (as it is usually done) two persons: the brain and the observer outside. The latter usually perpetrates the skeptic utterances which philosophers strive to refute. In fact his perspective will be the ours. We will discuss here several cases but the vat situation will be similar. The observer is outside the vat. The vat maintains the brain's life and the joined computer delivers to it electric impulses which are one and the same as impulses which a normal human brain obtains from the body. Briefly the impulses contain a projection of a reality. Also the levels of all the chemical substances solved in the blood which are important for the correct work and impressions of the brain are secured by the vat and computer. Consequently the brain has delusion that he lives a normal human life.

Before I come to the Putnam-like case I will consider some others for better understanding what is going on when a brain is put into such a vat.

CASE I: (abducted aware brain, AAB): a person was abducted, his brain put into the vat, his body destroyed. The projection continues his experience as if he found himself in a different place, say foreign country, but still he experiences having the same body. He knows that he is a brain in a vat. He knows that his present experience is a delusion and that in fact he has no hands. He knows that the hands which he experiences are in fact some parts of the processor in the computer with some electric charges and impulses in it.

A few words about the latter question. What is the real reference of the brain's perceptions and utterances? Putnam (p. 14) mentions three possibilities. A tree in the projection refers to:

1. The tree in the image,
2. The electronic impulses that cause tree experience,
3. The features of the program that are responsible for those electronic impulses.

The first possibility explains nothing. In fact it means that "a tree" refers to a tree. It is an obvious disquotational truth but it does not point to any particular object. The second possibility is an obvious false. In normal life we do not see electric impulses, even if our eyes deliver them to our brains. We do not see the photochemical reactions going on in our eyes, in retina. We do not see photons coming to our eyes. We see objects. So if the brain receives some impulses the question is what object he perceives. The third possibility also may not be the answer, though Brueckner (2012) calls it "the most plausible semantic/content

externalist reference assignment”. Features are not objects. Even the program is not an object. An object is the processor which is the physical source of those impulses. Physically the program is a temporary spatial distribution of charges in a part the processor. It controls the temporary spatial distribution of electric charges and impulses in another part of the processor where are gathered the data which it is processing. They contain encoded the projected reality and the result is sent to the brain. In fact the brain perceives the processor with some electric charges spatial patterns in it. When you see Angelina Jolie on the screen in the cinema, then what do you see? In fact you see the screen with a colorful moving patch on it. The brain’s situation is similar, though the moving patch is electric.

Assume that the brain says “In the delusion I am keeping a cup of coffee in my hand now”. What is the reference of the word “hand” used by him? He knows everything and is not deceived. Both he and the observer know that it is not any real hand but the processor with electric charges pattern with encoded the hand experience in it.

In fact the brain feels as he were in a cinema. He knows that there is the world outside, and he knows that now he perceives only a projection coming from a device. He know that what he really perceives are some electric patterns in the parts of the processor as if they were some patches on the screen in the cinema. Referential meaning of his thoughts and utterances is exactly the same as the observer’s. Both they could describe the brain’s situation with the same words and the same meanings.

CASE II: (abducted unaware brain, AUB): a person was abducted when sleeping, his brain put in the vat, his body destroyed, but he does not know what happened. The projection continues his experience as if he continued his normal life, and he thinks so.

When he describes his activity using the sentence “I am keeping a cup of coffee in my hand now” referentially he means a real cup of coffee in his real hand, not a processor with an electric pattern in it. He does not know what is the real reference of his perceptions. He is deceived.

In fact this deception is of the same kind as any other. The only difference is that it is an extraordinarily massive deception. He is deceived at all. However the general idea is the same. He thinks that referentially he perceives something different than in fact he perceives. The same situation you would have watching an illusionist, or when somebody cheated or lied to you and you would believe it. Say your worker looking very badly says that he is ill, when in fact he was at a party till 5.00 a. m. (work starts at 7.00). If you believe him you are deceived

and referentially your belief misses the facts. The data you obtain you refer to another situation than in fact they come from. Actually it is the case of a wrong interpretation of your data. It could be that you are not so gullible and you could consider also some other scenarios ending with your worker's bad looking.

For better similarity to the brain's situation take such an example. Your friend Mary is seriously ill and since a few weeks it could be seen on her face, but she says nothing and hides it under a perfect makeup. So you are deceived and think that she is healthy as always. Still you see the same as always and you do not know that the referential meaning of your data has changed. This is exactly the situation of AUB. Still he sees the same as always and he does not know that the referential meaning of his data has changed. The reference has been replaced, though the data are still the same.

It is obvious that as you could be less gullible in the case of your worker, so you could be more suspicious (or careful) in the case of your friend. Generally you could know that there is lots of possible interpretations of your data and this one which you chose to believe may be false. You could consider one day that even if there is no hint for this and the idea seems to be stupid your friend might be seriously ill and hide it under her makeup. For sure it is possible that you think like that.

If one of your students had replaced secretly with his twin brother which you have not known about it would be also a case of the reference replacement with the same data held. And no matter how stupid it would be it would be possible that you consider one day if something like that obtains.

The question is if the AUB could think that he is a brain in a vat in a similar manner. The answer is yes. That is only the question of quantitative range of the deception (delusion) considered by him. Your friend's face is only a little fragment of whole the data you receive, and the question of her illness is only a little fragment of whole your picture of the world, even if you like her very much. Whereas the brain's illusion includes all perceptions that he receives and it refers to all the physical surroundings which he has a causal contact with. Everything was replaced, not only your healthy friend with your ill friend. All the objects among which he had lived before the abduction got replaced with the computer. But still it is possible that he thinks one day that no matter how foolish this idea seems to be it is possible that he was abducted, and has his brain put into a vat, and though the data are exactly the same he does not perceive the usual objects of his usual surroundings as it was before but a processor of a computer with some electric charges patterns in it.

Putnam probably would say that his thinking like that would be possible because his perception before the abduction referred to object referentially more or less similar to a possible vat, even he in fact had not had any contact with such a technique and its devices. He would insist that his thoughts and words referentially would refer to this area of reality where such a vat could exist. And it would be possible because all his thoughts refer to the reality outside the vat. That is possible because when he was learning his language the referents of his words were placed outside of any vat, with the vat included in special cases of talking about vats.

CASE III: (born unaware brain, BUB): an embryo brain was cut off from a human embryo and put into a vat, the rest of the body destroyed. The vat secures conditions necessary not only for existence but also for normal growth of the brain. The projection delivers signals which are similar to signals which a normal growing human brain receives. Thus a normal human foetal brain comes into being and a virtual birth has place. Consequently a normal human brain arises. It learns to walk and talk, walks to schools, gets married etc. The projection secures him a delusion of a usual life. He does not know what is really going on.

In fact this is a Putnam-like case. It is obvious that referentially the brain's usual thoughts and words never go beyond the processor delivering signals to him. When he says "I am keeping a cup of coffee in my hand now" he refers only to the processoral events. So even if he does not know that he talks about the computer the reference of his utterances and the reference of his perceptions are the same. Thus he is not deceived. This is one of Putnam's theses, though not literally put by him. (One of advantages of Putnam's reasoning is that he had shown that even if we are brains in vats we have some adequate knowledge of external world, it is our knowledge concerning the patterns in the processor. In the case of Evil Demon it would be the part of his mind in which the deception would be produced (cf. Brueckner 2012).)

For better understanding this point recall that in fact we are in very similar position. When you look at a tree you do not know everything about it. Putnam confessed that he even could not tell elms from cheeks (p. 18). Some people do not know about the cellular structure of plants, organelles and the cytoplasmic metabolism. Another problem is the molecular and atomic structure of a tree. And atoms are not the end of the journey. Scientists still do not know the exact structure of matter. From this point of view we do not know exactly what we see looking at a tree. But it does not mean that we are deceived or deluded.

The comparison to scientific search for the real structure of matter helps to understand that BUB does not have to know that in fact he refers to a processor

for really do it. It is enough that he thinks about this what he sees, (one might add) no matter what he really sees. It is exactly as in our (normal humans) case. We may refer to trees without the exact knowledge to what in fact we refer.

A delusion occurs only when a mixture of references takes place. The necessary condition is that you have some two qualitatively identical or almost identical (so possible to confuse) perceptions which refer to different objects. A delusion appears when you take the second perception as referring to same object as the first one. It is when you take your friend's made up face as referring to a healthy person, when you take the presence of the twin of your student as referring to himself, and when you take the processor with some electric charges patterns in it as the usual reality. It is not the case with BUB. He does not mix references. For whole the time he lives only with the reference of the processor.

Putnam claims that the circumstances in which BUB has learned his language make it impossible for him to think about anything referentially going beyond the processor. For sure he is right that when BUB thinks about his normal everyday activities like having a cup of coffee or playing with children he referentially thinks only about the processor. But when the question refers to being a brain in a vat it is an extremely extraordinary topic for him, as for every human. The problem is if BUB is able at all to think about anything referentially going beyond the processor, especially could he really think about such extraordinary situations. Putnam claims no. I maintain that he is wrong.

What is important in the BUB's situation, no matter how strange the conclusion sounds, is that he does not make any cognitive mistake. He sees the reality exactly as it is. It is so because he does not mix references. The point is only that he does not know everything about the reality. He does not know that the reference of his beliefs is not simple but is a small part of a vast compound system of objects which he theoretically would be able to perceive and know. A condition is that the causal signals generated by these objects should be connected with his senses. An alternative condition is that some hints concerning the realm of reality should be present in his normal life but he would have to adequately gather and analyze them. The third possibility is that he would have to create some special extraordinary situations in which these hints would be observable.

In fact there is no limits for considerations concerning objects causally standing beyond the objects which are well known for us. Moreover this mode of research belongs to our everyday activity and science or philosophy is only a more organized, consequent and compound kind of this. When one sees apples on the table he automatically assumes that something had happened that they have been brought there. If apples could travel by themselves he would not seek

for any other referential explanation. But we know that they cannot so we know that some other objects stay beyond that appearance, some people usually. Whatever happens we assume that it happened by itself or we search for some other objects. The latter possibility is always present. If somebody falls ill, we think that it could happen because of some processes going on in his body or soul, but still the possibility is actual that some other objects are responsible for this, some germs for example, or a poison, or even magic, if somebody believes in it.

Uranus and Neptune were postulated and discovered for explaining some perturbations in other planets movements. The dark matter hypothesis is another example of postulating of some new objects for explaining some inconsistencies in other objects movements (galaxy rotation). Physicists rain us with ideas that here or there is a tunnel to a hidden domain of reality, another world, in black holes for examples, or that there are some parallel universes with which we have almost no contact but which cause some slight physical effects in our universe etc.

It seems that there is no limits in postulating, considering or thinking over some new objects which stand beyond the objects which we have already known. Of course when we perform such thoughts we extend the sphere of reference which we think about. We coin new words with new references. So it seems that there is no obstacles that BUB could think that all his perceptions refer in fact to a processor and that this processor is a part of a wide reality which is theoretically knowable for him if his brain would be connected to some senses like eyes (some real ones).

The objection could be if thinking this BUB really would refer to the vat and the world outside. Putnam would say that it is impossible because there would be no causal connection between the out-vat objects (and even the vat itself) and the brain (cf. p. 14). He is certainly right that talking about a given object requires a causal connection. To make sure consider the following cases.

You see a photograph which shows only a man's face. He looks angry, so you joke saying "Oh, I see the chair wasn't comfortable". You do not know this but in fact that man was sitting on a chair when the picture was taken. So you have guessed that there was a chair and he was sitting on it. Do you refer to it?

You go on a picnic in a grassland with a blind man. At once he reaches out his hand and says "Let's rest under that tree". Accidentally he points to a tree, the only one in the surroundings. It has to be a pure guess. Does he refer to the tree?

And the third one. Imagine that BUB says one day “Hey! Observer! Don’t look around! Yes, I am talking to you! Take me off from that damned vat! I am sick of these processor hallucinations!”. That is not even a guess, it is an assumption. However is BUB referring to the observer, the vat and the processor?

It is obvious that the witnesses of such events could have an illusion of reference. The person who have showed you the photograph could smile and think about you “Oh, that’s a pure joke, he does know nothing about that chair”. But also she could be surprised and say “How do you know about that chair?”. On the picnic you could even think that your blind friend has some extrasensory perception abilities. Thus imagine the thrill in the observer’s body if he thought “Oh, God, he knows!”. And the next thought: “How could this happen? Is it possible that he has any information from outside? There is a spy in the lab who did this?”

The character of these reactions affirms the general principle. If you believe that somebody refers to an object you assume that he knows that the object exists, and knowledge requires some causal connection. So Putnam would be right saying that it would be impossible that BUB referred to the very vat he would be inside.

Still Putnam is wrong because his conclusions go too far. To think that he is a brain in a vat BUB does not need to think about the particular vat he is actually in. It is enough that he thinks about a vat he could be in. It is not necessary that he really refers to the very vat. Considering does not require any particular knowledge nor reference. We can consider things which do not exist. It is obvious that particular reference is not necessary for such thoughts. It is enough that you think about possible objects which could be in the referential position you consider. Take a bridge which is to be build. You can think about it even if there is no causal signal coming from it to you. However you can consider this possible bridge referentially because you refer already to the river and the banks. Thus BUB refers already to the processor, and he can think referentially that all the references of his perceptions is a processor with some electric patterns in it, his brain being in a vat, and that the processor and that vat are a part of some bigger surroundings. The processor etc. is hanging referentially on the total BUB’s perception reference as the bridge on the river and the banks.

So what in real Putnam could have had on his mind? He could answer as follows. To consider something in the referential meaning requires to be able to know if it existed. You can think about the bridge which is to be build because you will be able to see it when it comes into being. BUB cannot think about his being in a vat, because he is not able to know if any of such vats exist.



This is the crucial point. Is it really impossible that BUB knows if such a vat exists? Of course it is not possible for him if you mean his powers which he has in disposal. He can only send electric impulses to the processor and obtain some processor situations as an effect. He is like a bird in a cage which cannot fly out and see what is the color of the sky today. A bird cannot (usually) free itself with its own powers. But it does not mean that it cannot see what the sky is. It is entirely possible. What is necessary the cage is to stop to keep the bird imprisoned. It is to be opened or destroyed in a natural way for the animal could fly out.

The same is the situation with BUB. He is able to see the vat. A sufficient condition is that somebody takes him from the vat and joins some eyes to his brain, or even connects a camera (with the signals suitably modified) to his eye nerves, instead of the processor. Thus Putnam is wrong. BUB can think that he is a brain in a vat because he could know if such a vat existed if some extra conditions would be fulfilled. Moreover he can even think that the possible vat he is in looks exactly as he imagines it on the basis of his processoral experiences. (Finally it could occur that projection trees are in fact electric charges patterns in a processor, cf. Brueckner 2005.)

For better understanding of this point try also such consideration. What is the real difference between AAB, AUB and BUB with respect to their ability to think if they are brains in vats? All they are causally cut off from the outer surroundings. And if any of them thought that he was a brain in a vat their thoughts would be in the same relation to the processor projection. It is, the qualities of such their imaginations would find identical counterparts in qualities in the experiences delivered by the processor. The processoral vat thought by AAB would be one and the same as the BUB's one. The difference is that AAB knows (is sure) that there is some world beyond his actual perceptions, and BUB may only consider this. Moreover AAB has seen the world, whereas BUB may only imagine it and wonder how it looks like.

The point is that AAB was once connected causally (his nerves were) to the outer reality whereas BUB has never been. AAB knows that before the abduction his perception referred to the outer reality and after it he was switched to the processor. BUB has not been switched this way, but the idea remains. The reference depends on this to which realm of reality your sensual nerves are connected to. There is no reason that BUB could not think like that, and think that his brain could be connected to another realm of reality. In particular he can think that one realm is included in another, as the processor reality is included by the reality of the observer, and that he is a brain in a vat.

Add to this the following thought. It deserves some deeper analysis but for brevity I will discuss it shortly. A problem is if trees in the projection do refer to the real trees intermediately via the minds of the designers of the vat. Putnam claims that “such a weak connection can hardly suffice for reference” (p. 11), but it is obvious that the qualitative features of the processoral kind of trees are in a strong causal relations with real trees. Thus the processoral trees could be treated in fact as moving pictures form an interactive plant atlas. You can even imagine such a science-fiction vision that in future everybody could join his brain to a machine and perceive such projections for education or entertainment.

To avoid that problem BUB’s world should be designed as much different from the our one as possible for his normal healthy activity, at least that he could consider if he is a brain in an vat. It seems that if it were to be a world for a human brain to live then some similarity and thus some problem of reference would have to remain, even in the extreme Putnam’s case, where no trees and nothing else exist but only the computers producing the delusions for brains (p. 6). Nevertheless it would be a reference only to kinds but not to particular objects. It would not help BUB to refer to the very vat he is in. Good for Putnam. But it would help him to imagine correctly a vat he could be in, because thinking that way he could that case

The last question sounds where exactly the Putnam’s mistake is? It is not because of a circularity as Brueckner claims (1986). The critical fragment of Putnam’s argumentation is (he describes the situation of brains in vats):

“... ‘vat’ refers to vats in the image in vat-English, or something related (electronic impulses or program features), but certainly not to real vats, since the use of ‘vat’ in vat-English has no causal connection to real vats... It follows that their ‘possible world’ [in which they are brains in vats] is really the actual one, and we are really the brains in a vat, then what we now mean by ‘we are brains in a vat’ is that *we are brains in a vat in the image* or something of that kind (if we mean anything at all). But part of the hypothesis that we are brains in a vat is that we aren’t brains in a vat in the image... So if we are brains in a vat, then the sentence ‘We are brains in a vat’ says something false...” (p. 14-15).

The error appears in the passage “what we now mean by ‘we are brains in a vat’ is that *we are brains in a vat in the image*”. It is true in reference to the everyday vat-English, but it is not necessary that BUB thinks only about his everyday activities. So this passage is not necessary. Putnam omits numerous possibilities which I have sketched above. It is possible that the sentence “We are brains in a

vat” in BUB’s mind referentially goes beyond the processor and reaches a real vat.

## REFERENCES

- Brueckner, A., 1986, “Brains in a Vat”, *Journal of Philosophy*, 83(3): 148–167.
- , 1992, “Semantic Answers to Skepticism”, *Pacific Philosophical Quarterly*, 73(3): 200–219; reprinted in DeRose and Warfield (eds.) 1999, pp. 43–60.
- , 2003, “Trees, Computer Program Features, and Skeptical Hypotheses”, in S Luper (ed.), *The Sceptics: Contemporary Essays*, Burlington: Ashgate.
- , 2004, “Johnsen on Brains in Vats”, *Philosophical Studies*, 129(3): 435–440.
- Brueckner, Tony, 2012, "Skepticism and Content Externalism", *The Stanford Encyclopedia of Philosophy* (Spring 2012 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2012/entries/skepticism-content-externalism/>
- Johnsen, B., 2003, “Of Brains in Vats, Whatever Brains in Vats Might Be”, *Philosophical Studies*, 112(3): 225–249.
- Putnam, H., 1981, *Reason, Truth, and History*, Cambridge: Cambridge University Press