YOU JUST CAN'T CRISPIN BRAINS IN A VAT

David Miguel Gray
Columbia University

P.F. Strawson once wrote, "One of the marks, though not a necessary mark, of a really great philosopher is to make a really great mistake." On first reading of Hilary Putnam's "Brains in a Vat" many people think that's exactly what he did, "make a really great mistake." I hope to arrive at a defense of Putnam's argument in light of recent criticism by Crispin Wright. I will begin by examining both the examples Putnam creates and their presentation. Second, I will take a look at the conclusion Putnam draws from his proof. Next, a brief digression will lead us into a discussion of two different types of skepticism and which one applies to Putnam's proof. I will then focus on the problem of the metaphysical realist and the applicability of the brain in a vat example to its plight. Finally I will discuss Wright's criticisms of Putnam's proof and what we may learn from it.

In Putnam's refutation of the brain-in-vat skeptic, he uses his second scenario of how the 'brains in a vat' picture works. In this case, the external world consists of living and functioning brains and their nervous systems. These are all contained in a large vat. The brains' individual nerve endings are hooked up to a fantastic computer. This computer, the automatic machinery, exists as the brains and the vat do, because it just happens to be the actual world, that is, no one made the computer. So what happens?

Being that all the individuals are hooked up to the same
computer, they all suffer from a *collective hallucination* as Putnam calls it. This means that they all see computer-induced images and that none of them have privileged access to the external world; they are all being deceived. If you and I were two of these brains, let us pretend just for a moment, both of us would receive all the sensory inputs required to be able to "see trees", "touch trees", and "throw rocks". The computer is very powerful and allows us to have the impression of interacting with the objects that we see; it also allows us to rearrange them in our environment. Similarly it allows me to communicate with the other brains in the vat. When I come up with thoughts, the computer processes them so that I am deceived into thinking that I hear myself speaking and feel my lips moving and feel sounds coming through my throat. Likewise, the brains listening to me are deceived into thinking that they are hearing my voice instead of just receiving my thoughts, *via* the computer, and may respond to me as adequately as I spoke to them. This is merely a more roundabout process than what we think is actually going on. It is clear that the brains may refer to images in their *sensual world*, but can they, as Putnam asks in the first chapter of *Reason, Truth and History*, "refer to external objects at all? (As opposed to, for example, objects in the image produced by the automatic machinery)" (12).

By 'sensual world' I mean the world of sensory perceptions that is created by the machine sending the brains electrical impulses and not the reality of their external world, which they do not know. When I refer to *people's* sensual world, assuming of course that we are not being deceived, I mean the world of sensory perceptions that has a corresponding external reality. For instance a brain in a vat's image of a coffee cup is actually induced by a computer where a person's image of a coffee cup corresponds to an actual coffee cup. It is important to try to realize what the brains are doing
when they are speaking or thinking about things. A brain's external world consists of itself and other brains, a vat filled with nutritive fluids, and a fantastic computer, none of which it can see nor can it know that it exists as such. In addition to this, the brains have always been in this environment and have no more of a reason to think that they are brains in a vat than we do. What the brains do see is their sensual world, which resembles the external world we live in (we are no longer pretending to be brains in a vat). They can only make references to things they see in their world. For instance, if a brain in a vat says, "There is a lamp in front of me" what it is really saying is, "There is an image of a lamp in front of me."

I would like to expand upon Putnam's ideas concerning the mental images of brains in a vat. For example, the context of the brains' dreams would be mental images based on images presented to them by the computer, and, likewise, when they imagine things not in their sensual world, this is simply done by manipulating a collection of the images that they have seen. For instance, they would obtain the mental image of a brain in a vat hooked up to a computer because they have seen a brain, a vat, and a computer. Like their language, their mental images also center around their sensual world. This means that they cannot possess what I will call a transcendental imagination, that is, they cannot cross the boundaries of their sensual world and refer to or accurately contemplate things in their external world. Even if the brains could correctly imagine what their situation is, this would be a fluke. (For instance, they would have no way of knowing whether they were brains in a vat or brains in a swimming pool, and the proper guess wouldn't demonstrate actual knowledge of the situation.) Furthermore, since they don't have a transcendental imagination, their imaginations couldn't even mistakenly refer to their situation in the external world because their
referential capabilities are limited to the sensual world.

I will briefly digress to discuss why Putnam doesn't use his first scenario, the original scenario, of the brain in a vat for his argument. The first case consists of a brain in a vat filled with nutrient fluids which is hooked up to a computer built by an evil scientist. With this first case, an externalist conception is possible. One may say that the images that the brain perceives are not non-referential images. The image of a lamp may be given to the brain by the computer, but the evil scientist, who has actually interacted with a lamp, programmed the computer to give such an image. Therefore, the brains are just perceiving a lamp in a round-about way, and to say that this is not an actual perception is just a prejudice towards computer received perceptions. And while the brains don't perceive these objects that we perceive in the correct 'order', so to speak, as they are in the world external to the vats, the objects were perceived by the scientist and then fed into the computer, and so much of the brains knowledge of the external world is both referential and true. This is a very important point. Crispin Wright, in his lecture "Putnam's Proof that We Are Not Brains in a Vat," elaborates on this by suggesting that, "The standard brain-in-a-vat fantasy is, whereas Putnam's is not, consistent with the truth of most of my beliefs about the material world" (Reading Putnam, 218). Wright continues:

The difference, in other words, is this: a skeptical argument which works with Putnam's fantasy can directly transmit our (putative) lack of warranted assurance that the fantasy is false into lack of warranted assurance that most of our ordinary beliefs about the material world are true. But no such direct transmission is possible if the argument works with the standard fantasy (219).
It is most likely because of these easy refutations that Putnam created the second case to avoid the possible criticisms that the evil scientist allows brains to make indirect causal references to external objects. But to solely address the first case and treat it as a way out of the skeptic's conception of the mind is to avoid the full strength of Putnam's article.

One of the points Putnam is able to demonstrate with the second case is that the brains' language is only referring to images and not actual things. By holding this position, we may not develop any externalist conception of the brains' sensual world. There is no causal connection between the images that the brains 'see' in their sensual world and real objects. If one of the brains thinks it sees an image of a lamp, there is not an actual lamp because only electrical impulses that the computer sends allow the brain to create an image of a lamp. The electrical impulses given by the computer which allow the brain to create this picture are simply part of the computer's nature. There is no actual lamp, just a program that allows the brain to create an image of it.

The question Putnam raised earlier of whether or not brains can refer to external objects at all allows him to come to the conclusion that it is impossible to say or think that we are brains in a vat. Putnam's self-refuting argument, as he calls it, is basically this:

although the people in that possible world [the brains] can think and 'say' any words we can think and say, they cannot (I claim) refer to what we can refer to. In particular, they cannot think or say that they are brains in a vat (even by thinking 'we are brains in a vat') (Reason, 8).

Putnam's idea of a self-refuting argument differs slightly from
traditional conceptions of self-refuting arguments in that the conclusion is a little harder to reach because of implicit meanings. However, like all self-refuting arguments, Putnam’s argument is one in which a part of the hypothesis is contradicted. Putnam argues that, “part of the hypothesis that we are brains in a vat is that we aren’t brains in a vat in the image” (15). This may be simplified by saying that it is implicit within the concept of being brains in a vat that we can not think of ourselves as brains in a vat in the image, that is, we do not see ourselves as brains in a vat. However, whenever supposedly envatted brains refer to something, they are referring to that something in the image, for example, there is an image of a lamp in front of us. Following from this, whenever an individual brain would say ‘I am a brain in a vat’ it is actually saying ‘I am a brain in a vat in the image’. This contradicts the part of the hypothesis that says that if we are brains in a vat, we cannot be brains in a vat in the image. Because of this, the statement ‘we are brains in a vat’ made by the brains is necessarily false. It should also be said that all of this also holds for a brain thinking that it is a brain in a vat, though this is a minor point. Much like the language the brain uses to refer to its sensual world, its ‘linguistic thoughts’, or thoughts in the form of words, refer to the sensual world as well. Furthermore, as I stated earlier in the paper, the brain gets its mental images, and hence many of its thoughts from the sensual world. Perhaps it is most easily stated by saying that regardless of whether a brain says or thinks ‘we are brains in a vat’, it still contradicts the hypothesis.

The major question still remains: how does Putnam’s objection to the brain-in-a-vat skeptic help us with the problem of skepticism. The answer to this question actually lies in the previous sentence and is perhaps the greatest misinterpretation of Putnam’s article. The brain-in-a-vat skeptic is not what we might think it to be.
In order to explain this I would like to describe two skeptical positions Putnam describes as Infinitely Regressive and Internal.

An infinitely regressive skeptic is of the type that no matter what premises you present the individual with, you will most likely get the response "And how do you know that?" or "Can you prove it?". This sort of skeptic is virtually impossible to please unless presented with the simplest of arguments such as whether or not we exist. Putnam had no intentions of refuting infinitely regressive skepticism.

Putnam states in his "Comments and Replies to Crispin Wright" that, "the aim of the internal skeptic is to convince us, on the basis of assumptions we ourselves hold, that all or a large part of our claims about the empirical world cannot amount to knowledge" (Reading, 284). Putnam's brain in a vat argument was geared towards disproving this type of skepticism.

Let's go back to infinitely regressive skepticism for a moment. Let us imagine the following scenario: You are on a spaceship exploring the outer limits of the galaxy when you and your crew happen upon a space station which wasn't made by human beings. Let us, for the sake of the example, suppose that no one made it, the station just happens to be another world of sorts. Upon entering the space station you see a large vat filled with billions of brains and an enormous computer that appears to be connected to each one. After examining the brains and running some tests on them, they appear to be isomorphic to human brains. Also at the computer is a terminal with a video screen and a video camera and microphone that allows you to view the brains world. The computer also takes images of you on the outside world and transforms the image of you into an image of a regular individual in the world of the envatted brains. After using this terminal to talk to other brains and explore their world,
you find that it is identical to earth, a twin-earth of sorts. Eventually you happen upon a brain who happens to be a philosophy student and is familiar with Putnam's article (even Putnam has a brain counterpart on this space station). After trying to inform the student of its situation the brain presents you with Putnam's argument, thereby proving that it is not a brain in a vat. After repeatedly trying to convince the brain that it is in fact a brain in a vat you give up and return to your home planet.

This scenario is the product of an infinitely regressive skeptical position. Putnam's argument keeps us from saying and thinking 'we are brains in vats' in response to the hypothesis. Thus we may not speak or think of the issue. Nevertheless, the possibility remains that we are brains in a vat, even though we may not speak or think on the matter and come to the conclusion that this is true. Although we may not speak of it, we could possibly accept such a position since we have not been convinced that such a position is impossible.

My reply to this scenario, and I imagine Putnam would take a similar stance, is that of course this example can't be refuted: it has already been verified that the possibility is true, that is, the envatted brains do exist, and there is nothing we can do to argue against this. This is what infinitely regressive skepticism is capable of: presenting a possibility that no matter what argument is used against it, it can always answer with, "What if?"

We may conclude that Putnam's argument doesn't even entirely rule out the brain-in-a-vat skeptic, it merely shuts it up. While Putnam's essay does make some very good points and does present a useful skeptical scenario in a materialist framework, his solution to the brain in the vat problem may only be considered a small contribution to the battle against skepticism.
While most of the things the infinitely regressive skeptic claims are correct, none of it matters! As said earlier, Putnam had no intention of refuting infinitely regressive skepticism. Furthermore, his goal was not to deliver a crushing blow to epistemological skepticism. Rather, as Crispin Wright correctly claims, "Putnam's real project is, as so often, to embarrass the metaphysical realist" (Reading, 217). And what is this metaphysical realist? To put it vaguely, since I've never seen it put otherwise, the metaphysical realist sees the world as set, a place where things exist objectively apart from our conception of them. The metaphysical realist believes that we can only hope to develop a view of the world that actually corresponds, at least in its basic suppositions, with the way the world really is. This barrier between our thoughts of the world and how the world really exists is what prevents us from developing an indubitable picture of reality. As Wright adds in his concluding remarks concerning metaphysical realism, "This is what commits the metaphysical realist to the possibility that even an ideal theory might be false or seriously incomplete" (Reading, 238). Wright continues:

Once one thinks of the world in that way, one is presumably committed to the bare possibility of conceptual creatures naturally so constituted as not to be prone to form concepts which reflect the real kinds that there are. The real character of the world and its constituents would thus elude both the cognition and comprehension of such creatures (238).

These metaphysical realists are exactly the types of creatures Putnam intended the brains in a vat to be. They, like the envatted thinkers, have no actual link with their external reality that they can be sure of. However, they do believe in semantic externalism, which
Putnam describes as "current philosophy-of-language jargon for the idea that meaning and reference are subject to causal constraints" (Reading, 285). The metaphysical realists must believe this, for how would they even attempt to describe the world if they didn't believe that they were somehow causally related to it? This is the strength of the argument from the position of internal skepticism. The internal skeptic is taking the possibilities that the metaphysical realist believes in as premises: First, that we may be brains in a vat, and second, that we have some sort of causal relations to the world. As we have already seen, the belief that they are brains in a vat and the belief that they can refer to brains in a vat is what got them into trouble.

Crispin Wright feels that in addition to being of little epistemological significance, Putnam's proof misses the mark, or marks, when dealing with metaphysical realism. The central problem Wright has with Putnam's proof is that it seems to answer specific types of metaphysical realism that can be discussed instead of metaphysical realism in general. Wright clarifies this position by saying that we "convict metaphysical realism" of something similar to the idea of \( \Omega \)-inconsistency:

An \( \Omega \)-inconsistent system of arithmetic, recall, is one which, for some arithmetic predicate \( F \), both contains a proof that there is an \( x \) such that not \( Fx \) and proofs of each statement of the form, \( \text{Fn} \), "\( n \)" being a numeral. Simple inconsistency is avoided only because the recognition that each \( \text{Fn} \) is provable cannot be accomplished via means formalizable within the system (239).
This means that metaphysical realists must accept that they are in a general sort of condition, like $F_n$, which is impossible to refute. To flesh this out a bit more, we may be in a general state, "a cognitive predicament of a certain very general sort" as Wright states. However, when presented with a specific example of what that state might be, like $F_x$, we find that such specific situations are often refutable. Wright claims that this is the case with Putnam's example.

Wright defends this position by clarifying the predicament of the metaphysical realist. In reference to Putnam's conclusion concerning the unthinkableability of the brain in a vat scenario:

... the sort of dislocation whose possibility is arguably implicit in metaphysical realism does not involve that its victims can conceptualize their predicament; quite the contrary - their predicament consists in part precisely in the fact that they are debarred from arriving at the concepts necessary to capture the most fundamental features of their world and their place in it (Reading, 239).

This provides support for Wright's claim that "Putnam's proof does not represent a general method of disproving any specific version of the relevant kind of possibility; at best, it represents a general method for disproving any specific version which we can understand" (Reading, 239).

An obvious problem just presented itself. The treatment of the metaphysical realist à la Crispin Wright does something Putnam never intended. Postulating that we still may be brains in a vat or something else which we cannot understand, is to expect our thoughts to transcend the reality that they must subscribe to. We would have to have a transcendental imagination. For metaphysical realists to transcend their linguistic constraints would be nothing short of
divine and to claim that doing this is a possibility would be to leave the realm of internal skepticism and enter into infinitely regressive skepticism. Putnam never intended to defend against this line of argument.

Putnam sums this up by saying, in his reply to Wright, "Wright's paper . . . seems to waver between pointing out that infinitely regressive skepticism has still not been refuted, and attempting a reply on behalf of an internal skeptic. But if the latter is Wright's intention, it is not clear what the reply is" (Reading, 285).

The only question left then is, "What may we salvage from Crispin Wright's argument?" Is there any way to Crispin brains in a vat? It seems from what we have seen that Wright's argument does not apply to brains in a vat. What I believe Wright does do is present an important aspect of the metaphysical realist. I don't believe that the brain-in-a-vat argument applies to all metaphysical realists, as I believe Putnam would also hold. What Wright has introduced is that complete satisfaction for some metaphysical realist would involve something that is perhaps unattainable. Certain metaphysical realists might hold that they must answer the doubts of infinitely regressive skeptics in addition to internal skeptics. This shows us that whether or not such a conception of reality could be true, large portions of this conception are unthinkable. So while we may be able to Crispin the metaphysical realist, we just can't Crispin brains in a vat.
WORKS CITED

