Brain in a vat theory by hilary putnam | analysis

In Hilary Putnam’s Brain-in-a-vat (BIV) example, a world exists in which brains, a neuroscientist, a supercomputer running simulations of brains contained in a vat, and the vat itself are the only objects. These items have either always existed or appeared completely randomly with each item in the same state (i. e. computers running simulations, brains are in vats, etc.) Understanding this, imagine the following scenario:

You are the evil scientist who monitors BIV’s and the experiences they receive from the computers. You ensure that all the BIV’s connected believe they are living a functional life in Springfield, Illinois. One of your BIV’s is code-named “ The Chancellor.” After some time passes, the Chancellor virtually utters the phrase “ I know I am just a brain-in-a-vat,” which he believes to be true, and then continues on with his programmed functions.

The BIV scenario Putnam presents is one such argument positing the skeptical hypothesis. Much like the Descartes’ Evil Genius, the skeptical hypothesis calls into doubt one’s knowledge of the external world. While the Evil Genius relies on a supreme deceiving deity, Putnam’s BIV ponders the effects of a mad scientist using computers to induce illusory perceptions and experiences. Traditional skeptics contend we are unable to discern the BIV hypothesis as false; if we were to grant the BIV premises as true, then our experiences would appear just they presently do (Stanford 2009). Consequently, skeptics maintain that we lack the ability to know anything about the world external to us.

Putnam applies his semantic externalism and consequently deems the scenario with the Chancellor impossible. Semantic externalism is a form of

externalism where “ meanings and truth conditions of one’s sentences, and the contents of one’s intentional mental states, depend upon the character of one’s external, causal environment” (Stanford 2009). More precisely, he focuses his concern to the first-person sentence, “ I am a brain-in-a-vat” to demonstrate that an instance in which the Chancellor verbally uttered said phrase is necessarily false. Consequently, Putnam deduces that we mustn’t be BIVs. I will argue that Putnam misuses the definition of “ vat” in his brain-in-a-vat experiment, thereby decreasing his likelihood of disproving the skeptical hypothesis. In this essay, I will define semantic externalism, followed by Putnam’s use of it against three different BIV scenarios.

If we were to accept semantic externalism, then we would necessarily acknowledge that how we define a term is not the sole factor in deciding what the word means to us. A common example would be the examination of a familiar substance (e. g. water) and how its meaning would remain constant even before encountering it. More precisely, those who adhere to semantic externalism would view the word “ water” as a term ascribed to a substance with a chemical composition of H2O before scientists had discerned the molecules comprising it; however, the composition of this substance we had labeled “ water” did, to some extent, contribute to our meaning (DeRose 102). For Putnam, interaction with things in the world represents the external factor. For example, consider two people who have the same mental states and then start interacting with substances which are cosmetically similar while still composed of different molecules. Perhaps one individual interacts exclusively with Ag (silver), and the other interacts only with ABC, but both learn the word “ silver” to refer to each of their respective

substances. As a result, each individual would have the same mental states (desires, beliefs, volitions, etc.), but with differing in what they reference; “ silver” would mean Ag for one, and ABC for the other.

In order for Putnam to connect semantic externalism to his BIV experiment, he begins by noting that it is necessary to acknowledge that any articulation of the sentence “ I am a brain-in-a-vat” appears to be self-refuting. More precisely, if we were to maintain this sentence to be true, the sentence would still derive a false conclusion since we couldn’t say “ I am a brain-in-a-vat” and know that I am in such a state. Take for example the statement “[t]here is no definite statement.” If you view this statement as true, it would have to be false because it is a definite statement. If you posited the statement to be false, then the answer is still false.

To demonstrate how this relates to Putnam’s BIV’s, first assume that we actually exist in the tangible world (T) just as commonly held, instead of in vats (instance T, where T is the Tangible world.) We have the following:

1. (T1) If I live in a Tangible world, I am not a “ Brain-in-a-vat.”
2. (T2) In instance T, I live in a tangible world.
3. (TC) I am not a “ Brain-in-a-vat” (True)(T1, T2)
4. I am not a BIV. (TC)

Next, consider we are now the brains in the vats a different instance (instance BIV). In this instance BIV, an intricate computing system continually feeds us phenomenal experiences. So, we now have the complicated computing system sending us signals for us to construct our experiences. Now we have a different problem in instance BIV:

1. (BIV1)If I am an actual BIV, a computer is sending me phenomenal experiences.
2. (BIV2) I am an actual BIV.
3. (BIV3)A computer is sending me phenomenal experiences. (BIV1, BIV2)
4. (BIV4)If I utter " I am a BIV", I am a brain-in-a-vat. (False, BIV3)
5. (BIVC) " I am not a brain-in-a-vat" (True) (BIV1, BIV2, BIV3, BIV4)
6. Thus, " I am not a brain-in-a-vat" (BIVC)

To clarify, semantic externalism implies that the subject will never interact with the Tangible vats in the world. So, when the Chancellor iterates " vat," he does not mean tangible-vats, but instead the origin of these electric impulses. More precisely, when he says " vat" he means artificial-vats because he actually interacted with a computer program. If the Chancellor was referring to a brain in the instance BIV, that brain would be referring to the electrical impulses sent from the computer in the form of artificial vats. Therefore, we arrive at the following scenario:

1. (BIV1) " I am a brain-in-a-vat" (False);
2. (BIV1) implies (BIV2) " I am not a brain-in-a-vat" (True)
3. (BIV1, BIV2) implies we are not brains in vats.

Consequently, the notion that " I am a brain-in-a-vat" appears to be a self-refuting according to Putnam.

After Putnam believes he has established this self-refutation, he must form a generic (i. e. universally applicable) argument (U). Hence, the following instance:

1. Iteration of (U1) " I am a brain-in-a-vat" (false, necessarily);

2. (U2) “ I am not a brain-in-a-vat” (from U1)(true, necessarily)
3. (UC)If I am not a brain-in-a-vat, then we are not brains in vats.
4. We are not brains in vats. (UC)

The uttering of “ I am a brain-in-a-vat” must be false since the instance BIV concludes that we are not brains in vats. As a result, we mustn’t be brains in vats according to this logic. While on the surface this may seem sound, I intend to show how Putnam may have missed the mark.

At first glance, the logic behind the Tangible world instance, the BIV instance, and the Universal instance may seem identical insofar as they each deduce we are not brains in vats; however, each stipulation incorporates a different meaning of the word “ vat.” The “ vat” used in the Universal instance represents an obscure term between the first two mentioned instances (i. e. Tangible world and BIV instance); the Tangible “ vat” represents vats from the tangible world just as we would perceive it today; and the BIV “ vat” stands for the virtual vat that the super phenomenal computer has created for us with its electric signals. Putnam’s error occurs when he doesn’t universalize the vat definition by using the latter sense of the artificial vat throughout instance BIV. While a bit confusing, it seems Putnam considers the instance (BIV1) since the only time it is true is in the latter sense of “ vat.” Putnam also wants to tie this definition to the Tangible world. After all, we all live in the tangible world and would want to believe we are not brains in vats while in the tangible world. Unfortunately, using slightly different definitions during an attempt to prove this conclusion hampers the argument. In other words, his argument is either that ‘(BIV1)

implies (BIV2) implies (TC)' or that '(BIV1) implies (T2) implies (TC); however, these arguments fail to hold true.

It is unnecessary to consider both possibilities in depth independently, since they can both be repudiated on the same criteria. Whether going from (BIV2) to (TC), or from (BIV1) to (T2), Putnam makes an observation about BIV-vats, and then uses that to make a claim about Tangible-vats. The true statement, (BIV2) " I am a not a brain in a virtual-vat" fails to imply " Not being brains in tangible-vats." Likewise, (T2) " I am not a brain in a tangible-vat" being true, fails to follow from (BIV1) " I am a brain in an artificial-vat" being false. The lack of a constant definition of " vat" poses one significant stumbling block for Putnam; however, if you maintain a certain degree of what constitutes " vat," the argument still remains invalid and runs into other problems which I will not address here.

I have attempted to argue that one cannot get to (TC) from (BIV1); however, anyone who subscribes to Putnam's argument against brains-in-vats might argue the opposite. It is important to recognize the conditions surrounding the argument. More precisely, Putnam presupposes we live in the natural world. The skeptic's main argument is that we lack the knowledge to discern whether or not we exist as brains in vats. We would only be able to differentiate between the instances and the vat usage if we new definitively which instances were being referenced.

Semantic externalism lacks the ability to liberate us from a skeptical hypothesis. Putnam fails to show that we can not be Brains in Vats, but he does show that if we were a BIV and uttered the phrase " I am a BIV," we

would not know it. He fails to cinch a logical loop and all he really shows is that the brain in the vat couldn't know it was a brain-in-a-vat. If Putnam's version of semantic externalism is correct, but we are brains in vats, then we lack the ability to discuss the tangible-world. For example, if we assume that we are brains in vats, when I said the worlds " tangible-world" in the previous sentence, I was actually referring to the artificial world because that is what I have been interacting with. Therefore, provided that the skeptic is correct that we would not know if we were brains in a vat, and also assuming that Putnam's semantic externalism is similarly correct, then we simply do lack the knowledge of the meaning of the words in our vocabulary. More precisely, assuming that interaction remains critical to gaining understanding and meaning, then it becomes a necessity to know precisely with what we must interact. Hilary Putnam attempted to embrace semantic externalism in hopes of freeing us from skeptics; however, Putnam did illustrate that BIV's wouldn't be able to say or even think they were BIVs, thereby demonstrating a quandary for semantic externalism.

Works Cited

DeRose, Keith. " Responding to Skepticism." Skepticism A Contemporary Reader. New York: Oxford UP, USA, 1999. Print.

Putnam, Hilary. " Brains in a Vat." Putnam on Brains in a Vat. 2 Mar. 2005. Drexel University. 10 Oct. 2009 .

Unknown. " Brains in a Vat." Stanford Encyclopedia of Philosophy. 29 Oct. 2004. SEP. 10 Oct. 2009 .