

Putnam on the Magical Theory of Reference

Kanti Lal Das

Abstract: The concept of reference is the main contention of epistemology and metaphysics in general and semantics in particular. It is through the reference of the linguistic expressions that one may come to know about the world or reality or ontology. It has been acknowledged that language does refer and the referential function of language is helpful to know about reality. However, the paradigm of reference remains a bone of contention to the philosophers. Putnam uses the metaphor 'brains in a vat' to expose the magical theory of reference. According to Putnam, the magical theory of reference is a big default of reference. He, therefore, attempts to overcome the default of reference through the analysis of the metaphor 'brains in a vat'. The main strategy of this paper is to exemplify the philosophical insight and significance of the metaphor 'brains in a vat' while solving the problem of reference.

Keywords: epistemology, metaphysics, reference, brains in a vat, linguistic expression.

Hilary Putnam (1926-2016), the proponent of internal realism, uses the metaphor 'Brains in a vat' to solve the problem of reference *in the realm of metaphysics and epistemology*. Reference is the key issue in epistemology, metaphysics, and in the philosophy of language, particularly in semantics. In the case of reference, language plays a pivotal role. The simple truth is that *language does refer to*. There is no question about the referential function of language. But the serious philosophical and epistemological questions are: what does language refer to? What is the locus of language? Is language reference? Is the language reference is something different from language? Is language refers to something extra-linguistically? Are language and the object as refereed by language different? Is reference is independent of the mind? Is reference is ontologically neutral? These are some serious philosophical questions that need to be taken care of. Further, the concepts of truth and meaning are associated with the concept of reference. Linguistic philosophers, in general, have attempted to solve the problem of meaning and truth by reference. The world or reality as such is represented by reference. Thus the issue about the relationship between *language and reality*, *language and ontology* is taken care of by the whole host of linguistic philosophers. As the epistemological and metaphysical foundation is

dictated by reference, it seizes significant ground of philosophy. Considering the significance of the theory of reference there developed various theories on reference, such as the Classical theory of reference expounded by Frege, Russell, Searle, etc., Causal theory of reference expounded by Putnam, Kripke, etc., and New classical theory of reference as expounded by Fodor and others. The classical theory of reference may be thought of as a *thesis*, the Causal Theory of reference may be thought of as *anti-thesis* and the New-classical theory of reference may be thought of as *synthesis*. This suggests that there remained a criss-cross and overlapping relationship among the theories of reference. This is natural because the relationship between language and reality is dealt with by humans where the language is taken as a tool of the medium of communication and reality is something there in the world outside human mind or the mental construction of human mind. Here I try to develop this paper in the light of internal realism where the metaphor 'Brains in a vat' is used to explicate the metaphysical and epistemological foundation of the theory of reference.

Many faces of realism

As an internal realist, Putnam invokes *many faces of realism*. In his book, *The Many Faces of Realism* (1991) Putnam deals with the question of whether there we still say anything about reality and truth. Putnam here anticipates various faces of realism. In this regard, he remarks, "‘Realism’ can be claimed by or given to at least two very different philosophical attitudes (and, in fact, to many). The philosopher who claims the only scientific objects ‘really exist’ and that much, if not all, of the commonsense world, is mere ‘projection’ claims to be a ‘realist’, but so does the philosopher who insists that there *really are* chairs and ice cubes (and some of these ice cubes *really are* pink), and lead two attitudes, these two images of the world, can lead to and have led to many different programs for philosophy."¹ Putnam examines various faces of realism, such as commonsense realism, scientific realism, metaphysical realism, and internal realism. He then attributes metaphysical realism with capital "R" and internal realism with small "r". There are two important aspects of internal realism, first, *it is objective*, and second, *it works with a human face*. According to metaphysical realism, the structure of the world is *ontologically independent of the human mind* and hence it is radically non-epistemic.² Moreover, it talks about *adequate conceptual schemes* of any given domain where they can type-reduce one another. Putnam while developing his internal realism denies the foundation of metaphysical realism. Contrary to metaphysical realism, Putnam affirms that *the structure of the world is ontologically dependent on the human mind* and hence truth is not radically non-epistemic. Here he ponders causal epistemic justification. Regarding conceptual schemes, he equally critical about metaphysical realism as he denies the possibility of type-reduce in more than one adequate

conceptual scheme.³

The Magical Theory of Reference

Putnam in his book *Reason, Truth and History* (1981) outlines the significance of the metaphor 'Brains in a vat'. Let me explain it with his example. An ant is crawling on a patch of land and thus creates a sketch line in the land that fortuitously portrays the caricature of Winston Churchill. Does it then lead us to assume that the ant traced a picture of Winston Churchill which essentially depicts Churchill? The general answer would be negative because the ant has never seen Churchill and it has *no intention of depicting Churchill*. It simply traced a line that we can see as the picture of Churchill. Based on this we can claim that the line the ant traced is not itself a *representation of anything rather than anything else*. The issue at our hand is that if the similarity is not necessary or sufficient to make something represent something else, how can anything be necessary or sufficient for this purpose? What else remained as the necessary or sufficient condition of representing something? How on earth one thing can represent a different thing? Now let me further suppose after Putnam that the ant had seen Winston Churchill and the ant had intelligence and skill to draw a picture of Winston Churchill. Further, suppose that the ant produced the caricature *intentionally*. In such a case the line would have represented Churchill. On the other hand, suppose the line had the shape WINSTON CHURCHILL accidentally, then the printed shape would not have represented Churchill even though the printed shape does represent Churchill when it occurs in almost every book. Thus for Putnam, 'what is necessary for representation *is intention*'.⁴ In this regard, we can cite the notion of the intentionality of consciousness of Husserl. Husserl in this regard referred to Brentano's concept of intentionality. Our intentionality of consciousness is always directed to an object. To say that I am conscious is to say that I am conscious of something. In this way, our consciousness is directed towards an object. In this sense the relevance of reference matters. "Intentionality has been understood by Husserl as the unique characteristic of experience to be consciousness of something."⁵ That means to have the intention that anything should represent 'Winston Churchill'; one has to think about or conscious about Churchill. Without thinking about or without conscious of Churchill, he cannot be represented. Likewise, if lines in the sand and noises etc., cannot 'in themselves', represent anything, then how 'how is it that thought-forms can 'in themselves' represent anything?' How thought can reach out and grasp what is external?

A possible solution suggested by some philosophers saying that the mind is *essentially non-physical in nature*.⁶ It might be the case that the ant though lacks intentionality of consciousness or thought, but the ant's curve applies to any physical object. It is claimed that no physical object

can, in itself, refer to one thing rather than to another, but thoughts in the mind obviously do succeed in referring to one thing rather than another. Thus, the nature of thought essentially differs from physical objects. Unlike physical objects, thoughts have the characteristics of intentionality. Accordingly, thoughts can refer to something else, physical objects cannot so the same. One can assume some mysterious powers of mind that control the employment of physical object through intentionality of consciousness. According to Putnam, ‘Intentionality is derivative from some employment of that physical thing by a mind.’⁷ Again the question arises: how is intentionality, reference possible?

To respond to the question of how is intentionality, reference possible, Putnam brings the concept of *the magical theory of reference*. So far we come to know after Putnam that the ant’s picture has no necessary connection with Winston Churchill and we also resolve that ‘picture’ bears an accidental resemblance to Churchill does not make it into a real picture, nor does it make it a representation of Churchill. We also resolved that to acknowledge the ant’s picture as a representation of Churchill, the ant has to be an intelligent ant; it has to have intentionality of consciousness or thought. The ant as a biotic community lacks everything. Some philosophers, however, claimed that some representations, particularly names, have a necessary connection with their bearers. To Putnam, to know the ‘true name’ of someone or something gives one power over it. What is the nature of this power? What is the locus of power? Is this power mind-independent? Is it something that lies independent of the mind? We had a sense of it if we carefully read Frege. According to Frege, thought (power) is the third-realm, it is neither physical, nor mental, but something else. It is the meaning of the sentence and without it, the reference of the sentence cannot be determined. Frege then invokes different levels of thought, such as apprehension, assertion, and judgment levels. He has differentiated thought from the idea. Putnam uses the term power. In this regard, he remarks, “This power comes from the *magical connection* between the name and the bearer of the name.”⁸ However, the point that has been raised by Putnam here is that once we realize that a name only has a *contextual, contingent, conventional connection with its bearer*, then where the logic that the name should have any mystical significance lies. Putnam, of course, proposes a solution to overcome the magical theory of reference. He suggests that if we realize that what goes for a physical picture equally goes for a mental picture, or so to speak, mental representations in general then in such a case one can overcome the so-called magical theory of reference. In such a case, metal representation has a necessary connection with the physical representation. For example, Wittgenstein’s concept of mental image is a case in point.

Suppose, there is a planet somewhere on which human beings evolved. Suppose they have

never seen trees, never imagined trees. Further, suppose, one day a picture of a tree is accidentally dropped on their planet by a spaceship without making any contact with them. Imagine that they were puzzling over the picture and all sorts of speculations naturally occur to them even though they never come close to the truth. However, for us, the picture is a representation of a tree and for these humans, the picture only represents *a strange object*. Now again suppose that one of them has a mental image very similar to my mental image of a tree. As a result, he has seen the picture. But his mental image is not a representation of a tree, but only *a representation of the strange object*. Here his picture is mysterious and it mysteriously represents. Someone might argue that the mental image that we talk about is indeed a representation of a tree because the picture which caused the mental image was itself a representation of a tree, to begin with. The logic is that there is a *causal chain* from actual trees to the mental image even if it is supposed to be a strange one. One might say that this causal chain can be imagined absent. When? Suppose the ‘picture of the tree’ that the spaceship dropped *was not essentially a picture of the tree, but the accidental result of some drip paints*. In such a case, even if it looks exactly like a picture of a tree, but in truth, it was no more a picture of a tree than the ant’s ‘caricature’ of Churchill was a picture of Churchill. Even one might imagine that the picture drooped by the spaceship came from a planet which knew nothing of trees. But the humans’ still would have mental images qualitatively identical with my image of a tree. However, they never image which represented a tree any more than anything else. We can assert the same in the case of words. It may perhaps be the case that a discourse on paper might seem to be a perfect description of trees, but if it was produced by a monkey randomly hitting keys on a typewriter for millions of years, then surely after Putnam, the words do not refer to anything. Even if there were a person who memorized those words and said them in his mind without *understanding* them, they would not refer to anything when thought in the mind either.

Again imagine a person that has been hypnotized by words. Suppose the words are in Japanese and the person has been told that he understands Japanese. Suppose, he thinks those words he has a ‘feeling of understanding’. This does not ensure that he could use the words rightly. It may perhaps be the case that he could not use the words in the right contexts, answer questions about what he thought. In such a case, he did not understand the word what he thought to be. It may even be the case that someone thinks words that are in fact a description of trees in some language and simultaneously has an appropriate mental image. However, from this, it does not entail that he essentially understands the words or even knows what a tree is. We cannot rule out the possibility that a mental image might be caused by paint-spills and the language it is associated with is ‘non-sense-sentence’.⁹ Putnam remarks, “In short, everything

passing before the person's mind might be qualitatively identical with what was passing through the mind of a Japanese speaker who was *really* thinking about trees- but none of it would refer to trees."¹⁰ Putnam intends to claim here that all these are really impossible just like it is impossible that monkeys should by chance type out a copy of *Hamlet*. Even though it is really impossible does not mean it is logically or physically impossible. It has been accepted in logic that what is empirically impossible would be logically possible. The concept of logical space as expounded by Wittgenstein in his *Tractatus* is a case in point.¹¹ Who knows that there are a lot of intelligent beings on other planets. If it did happen, then it would be a striking demonstration of an important conceptual truth. According to Putnam, thought words and mental pictures do not intrinsically represent what they are about.

Brains in a vat

Let me explain after Putnam how the metaphor 'Brains in a vat' signifies the magical theory of reference. While analyzing the insight of the metaphor 'Brains in a vat', Putnam draws our attention to the *fiction of science*. Let me imagine after Putnam that a human being has been subjected to an operation by an evil scientist. One can imagine this in his or her own case. The person's brain (your brain) has been detached from the body (your body) and placed in a vat of nutrients which keep the brain alive. Now the nerves of the brains ending have been connected to a *super-scientific computer*. This super-scientific computer causes the person whose brain is in the illusion that everything is perfectly normal. The brains function under the illusion in such a way that there seem to be people, objects, the sky, etc. However, in reality, all the person including you is experiencing the result of electronic impulses traveling from the computer to the nerve endings. Here the function of the brain is fully controlled by the super-scientific computer. The computer is so clear that if the person tries to raise his hand, he will get the same feedback from the computer. That is, the computer will cause him to see and feel the hand being raised. Likewise, the evil scientist, if he wishes, can cause the victim to experience or to hallucinated any situation whatsoever. Moreover, he can also wipe out the memory of the brain operation. The victim is no way knows that there is an evil scientist who removes people's brains from their bodies and places them in a vat of nutrients which keep the brain alive. The insight of the metaphor 'Brains in a vat' is also linked with the classical problem of scepticism where the question: How do you come to know that you are not in this predicament about the mind-world relationship?

Putnam then goes on to say that the metaphor 'Brains in a vat' can be applied to everyone. We can imagine that all human beings are brains in a vat and may even be extended to all

sentient beings. It may perhaps be the case that automatic machinery is programmed to give us all *collective hallucinations* instead of separate unrelated hallucinations. And successful illusionary communication may be going on among the persons. This does mean to say that my words actually reach your ears, rather when I produce my words different impulses travel from my brain to the computer. It essentially causes me to hear my voice and in turn, causes you to 'hear' my words. Thus we are in a sense actually in communication. Here I am not mistaken about your real existence apart from brains. It does not even matter that the entire world is a collective hallucination. We are essentially brains in a vat without violating any physical laws and this is perfectly consistent with everything we have experienced. However, what we have experienced under the brains in a vat cannot possibly be true, but it is self-refuting in a certain way. To Putnam, "A 'self-refuting supposition' is one whose truth implies its own falsity."¹² A self-refuting statement appears as a well-known paradox. For example, let us consider the statement 'all general statements are false'. Now if this general statement is true then it must be false. Hence, it is false. Thus it is self-refuting. Sometimes, a thesis may be designated as self-refuting if it implies its falsity. The statement, for example, 'I do not exist' is self-refuting if it is thought by me. Here it implies its falsity. Following Descartes it can be said that 'I think, therefore I am'. Now to say that 'I think, therefore I do not exist', goes against Descartes 'Cogito Ergo Sum'. However, when are brains in a vat, we will live up with the self-refuting statement. We are brains in a vat has just this property. According to Putnam, we may subscribe compatibly along with physical law that there should be a world in which all sentient beings are brains in a vat. If it would be the case then it is quite natural to assume that all humans of that possible world have exactly the same experience that we do and they have the same thoughts or images as we have. In such a case what we think about reference? Can the people of a possible world under the brains in a vat refer to what we can refer to? Even though the people in a possible world under the brains in a vat can think and say just like we can think and say, but they cannot refer to what we can refer to. To Putnam, 'they cannot think or say that they are brains in a vat (even by thinking 'we are brains in a vat')'.¹³

Turning's Test

Putnam then brings the concept of *turning test* to explicate further the significance of the metaphor 'brains in a vat'. His idea of turning test is nothing but an intelligent conversation through the computer. British logician Alan Turing developed this test. Suppose someone carries on a conversation with the computer and a conversation with a person whom he does not know. Now if he cannot tell which is the computer and which is the human being, the computer is conscious. To Putnam, a computer machine is conscious if it can pass the 'Turning Test'. It

should be kept in mind that the conversations are not carried on face to face and the interlocutor does not know anything about the visual appearance of his two conversational partners (one is a computer and the other is a person). Further, no voice is used considering the fact that the voice of a machine is different from the voice of a human. Imagine that instead of voice, the conversations are carried on vis electric type writer where the interlocutor types in his statements, questions and in turn he gets the response from the partners vis the electric keyboard. Further, suppose that the machine can lie. If it is asked 'are you a machine', the machine might reply 'No, I am not a machine, but an assistant of the lab here'. Now Putnam brings the idea of 'Turning's Test' not for seeking philosophical inquiry, such as whether this test is a definite test of consciousness, but simply for exploring the notion of reference. Accordingly, the task at our hand is not to exemplify whether the partner is really a person or a machine, but to exemplify whether the partners are capable to use the words to refer as we are used to do. There it seems to me that Putnam actually takes not the 'Turning Test' as the definite Test for reference or shared reference.

Putnam immediately denies the possibility of preconceiving the Turning's Test as the definitive test of reference or shared reference. He ponders that the Turning's Test is very similar to the magical theory of reference. According to Putnam, the Turning Test for reference is not definitive, though it is supposed to be an excellent test in artificial intelligence in practice. It may perhaps be highly probable, and hence we cannot rule out its logical possibility. It might be the case that someone could pass the Turning Test for reference without referring to anything. It has an obvious limitation. Here we can extend our observation that words do not have a necessary connection to their referents. This position again is self-refuting as someone could pass the Turning Test of Reference without referring to anything. It is indeed an illusion. In the language of Advaita Vedānta, it is just like Maya where appearance seems not to be real. There is no determinate reference in the Turning Test for reference. Putnam thus asserts 'that the Brain-in-a-Vat Worlders cannot refer to anything external at all (and hence cannot say that they are Brain-in-a-Vat Worlders).'¹⁴ Putnam cites another example. Suppose he is playing the 'Imitation Game' in Turning's terminology where his partner actually is a machine. Let us suppose, this machine enables us to win the game, capable to produce beautiful responses in English to statements, questions, remarks, etc., English but lacks sense organs. Putnam thinks that Turning does not assume that the possession of either sense organs is necessary for consciousness or intelligence. What we should say about such a machine? Here Putnam contends that we cannot and should not attribute reference to such a device. Of course, it is indeed true to say that the machine can discourse beautifully, but it surely 'could not recognize an apple tree or an apple, a

mountain or a cow, a field or a steeple, if it were in front of one.’¹⁵

It thus seems that the electronic device can produce sentences in response to sentences, but the fact is that none of the sentences can be connected *to the real world*. The machine as we observe can play imitation game with humane through the type-writer, and even they go on fooling each other forever, but the real crux of the fact is that machines cannot talk of apples as referring to real-world apples than ‘there is to regard the ant’s ‘drawing’ as referring to Winston Churchill’.¹⁶ Thus, there prevails an overall illusion of reference, meaning, intelligence under which the machine’s discourse refers to apples, chairs, etc. Likewise, there is the illusion that the ant has caricatured Churchill for the same reason. Humans differ from the machine in the sense that unlike a machine, humans able to perceive, handle, deal with apples and fields. Our talk of apples and fields are intimately connected with our non-verbal transactions of the same. We do because language entry rules take us from the experience of apples to such utterances as ‘I see an apple’. Besides, the rules of language take us from decisions expressed in linguistic form, such as, ‘I am going to buy some apples’. Machines are lacking either language entry rules or language exist rules. Thus the way humans converse with each other with language is unlikely the way a machine converses with a man. It could be argued that the ant would have drawn the same curve even if Winston Churchill had never existed. In a slightly different way we can say that if apples, trees, etc., had not been existed, the programmers would not have produced the same program. There we cannot imagine any causal ground in ant’s caricature, whereas there is some causal connection between the machine and the real world apples. But both are suffering from the perspective of true reference. Putnam contends that the machine is utterly insensitive to the contingent existence of apples and if all empirically real objects things ceased to exist, the machine could still have existed. Accordingly, the machine cannot be regarded as a reference at all. Thus, for Putnam, the Turning Test though capable to play Imitation Game, but Imitation Game is certainly not referring to any more than a record player is.

The ‘brains in a vat’ do not have sense organs but they do have provision for sense organ. More importantly, the ‘brains in a vat’ are brains, they are functioning brains, and they function by the same rules as brains do in the actual or real world. Accordingly, it would absurd to deny consciousness or intelligence to them. But they are conscious and intelligent does not mean that their words refer to exactly the same what our words refer. Do their verbalizations, say, for example, trees, actually refer to trees? Can the ‘brains in a vat’ refer to external objects at all? The answer is negative as there is no connection between the word ‘tree’ as used by these brains and the real trees. Interestingly, these brains would still use the word ‘tree’ just as they think, thought, just the images they have irrespective of the fact that there were no actual trees.

Putnam conceives that the brains in a vat are not thinking about real trees when they think ‘there is a tree in front of me’, ‘because there is nothing by virtue of which their thought ‘tree’ represents actual trees. Thus there is no point in saying that the ‘brains in a vat’ can refer to external things. Even if the brains think about, refer to, represent trees, it is because of the way the program connects the system of language to non-verbal input and outputs. However, there are no such non-verbal inputs and outputs in the Brain-in-a-Vat world.

Concluding Remarks

Based on the above observation, I can say that Putnam’s metaphor ‘brains in a vat’ actually portrays the mockery of the referential function of language. According to Putnam, brains in a vat cannot say that they are brains in a vat. The brain in a vat thinks ‘There is a tree in front of me’, but his thought does not refer to actual trees. Of course, brain in a vat might refer to trees in the image or through electronic impulses and there remains a causal connection between the use of the word ‘tree’ in vat-English and the presence of trees in the image. Accordingly, the brain is right in thinking ‘There is a tree in front of me’. Likewise, ‘vat’ refers to the vats in the image in vat-English, but certainly not to real vats, since the use of ‘vat’ in vat-English has no causal connection to real vats. It follows that if their possible world is really the actual one, and we are really the brains in a vat, it then entails that ‘we are brains in a vat’ is that we are brains in a vat in the image. But part of the hypothesis that we are brains in a vat is that we are not brains in a vat in the image. Thus for Putnam, if we are brains in a vat, then the sentence ‘We are brains in a vat’ says something false. Simplistically, if we are brains in a vat, then ‘We are brains in a vat’ is false. Therefore, we are brains in a vat is necessarily false. It is linked with a combination of two errors, first, taking physical possibility too seriously and second, unconsciously operating with a magical theory of reference. Here Putnam inclines to say that the existence of a physically possible world in which we are brains in a vat does not lead us to assume that we are really brains in a vat. It is not physics, but philosophy that would rule out this possibility.

Here two important philosophical resolutions need to be taken care of. First, the magical theories of reference are wrong. It is wrong for mental representations and not only for physical ones. Secondly, one cannot refer to certain kinds of things, namely, trees, without causal interaction with them. Putnam’s clearcut position is that if there are mental representations that necessarily and essentially refer to external things, they must be of the nature of concepts, but not the nature of the image. While acknowledging this position, Putnam actually denies Brentano’s theory of intentionality where the role of image is colossal. What then are concepts? To Putnam, concepts are not mental representations that intrinsically refer to external objects. Hence concepts are

not mental representations at all, rather they are signs used in a certain way. Concepts cannot be identical with mental objects of any kind. Mental objects do not intrinsically refer any more than the ant's picture does. Thus we end the paper with Putnam's remark, "The doctrine that there are mental presentations which necessarily refer to external things is not only bad natural science, it is also bad phenomenology and conceptual confusion."¹⁷

Notes and References

1. Putnam, Hilary, *The Many Faces of Realism*, Open Court, Publications, Chicago.
2. Forrai, Gabor : *Reference, Truth and Conceptual Scheme: A Defense of Internal Realism*, Eotvos Lorand University, 2001, p. 20.
3. Ibid., p.23.
4. Hilary, Putnam : *Reason, Truth and History*, Cambridge University Press, Cambridge, 1981, p.2.
5. Bhadra, M. K., *A Critical Survey of Phenomenology and Existentialism*, Allied Publishers Ltd., Kolkata, 1990, p.65.
6. Hilary, Putnam : *Reason, Truth and History*, op. cit., p.2.
7. Ibid., p.2.
8. Ibid., p.3.
9. Ibid., p.5.
10. Ibid., p.5.
11. Wittgenstein, Ludwig : *Tractatus Logico-Philosophicus*, D. F. Pears and B. F. McGuinness (Trans.), London and New York, 1922.
12. Putnam, Hilary : *Reason, Truth and History*, op. cit., P.7.
13. Ibid., p.8.
14. Ibid., p.10.
15. Ibid., p.10.
16. Ibid., p.11.
17. Ibid., p.21.