



Thought and Experience

Thought experiments in philosophy

Alex Barber

I asked Tim Crane, Professor of Philosophy at University College London, about the implications of Hilary Putnam's famous twin earth thought experiment. We started by talking about its implications for language and moved on to talk about its implications for thought. In between, along with Barry Smith, we discussed the legitimacy of using thought experiments in philosophy. Tim crane began though by outlining the view that Putnam was criticising with his twin earth example.

Professor Tim Crane

Putnam identified something he called the traditional picture of meaning, which involves two very important assumptions. The first assumption is that knowing the meaning of a word was a matter of being in a psychological state; a state of mind. It's when you know the meaning of the word that is a matter of having certain beliefs about how things are. The second assumption was that the meaning of a word determines its reference. What that means is that what the word means determines what it is about in the world. There has to be some connection between the meaning of the word cat and the fact that it refers to cats. That it is real cats in the world that the word refers to. Putnam expressed this by saying that the meaning of a word determines its extension or its reference. The words extension and reference can mean the same thing for the time being. Now he took these two views that meaning is a matter of the psychological state that you are in and that meaning determines reference, to be part of the traditional picture of meaning and I think he is certainly right about that. Many of the great philosophers of language of the Twentieth Century held these views. Now Putnam wanted to argue that these views were actually inconsistent when we take into account certain obvious possibilities. And he provided a very powerful argument against the view that meaning, or knowing the meaning, was a matter of being in a certain psychological state.

Putnam wanted to argue that two people could be psychologically exactly the same. That is they could be in the same mental states even though they meant different things by their words. The way he argued this was by using a thought experiment, an imaginary situation which brings out the consequences of some of the things that we think. His thought experiment was this. Suppose there is, in this universe, a planet, which is identical in every respect to this planet, to earth. And we can call this planet Twin Earth. It's identical in every respect except one. The thing that they call water on Twin Earth is not the same substance as the thing that we call water on earth. It's got some different chemical constituent. So when someone on Twin Earth says "water, water, everywhere, nor any a drop to drink" they are referring to something different than what someone on earth is referring to when they say those words. Now Putnam says let's suppose within the scope of this fantasy, this thought experiment, that each of us has a duplicate on Twin Earth, someone who is exactly the same in every respect as us. Now, call these people the twins of people on earth. So each of us has a twin. Now our twins are by hypothesis psychologically identical. They are in all the same psychological states. But when my twin uses the word water he is referring to something different from the thing I am referring to when I say water. Well you might say "so what? Big deal" once you add the assumption that meaning determines reference then you can see that the meanings of our words must be different too according to Putnam. It's very important to realise that determines here means the following thing. It means that if meaning determines reference then if two words differ in their reference they differ in their meaning but not vice versa. So if the meaning of cat determines the reference namely the property of being a cat, if the word were to have a different reference it would have a different meaning. So apply that to the Twin Earth case my utterance of water refers to something different from my twin's. Therefore, by that principle, meaning determines reference. Our words mean something different. And that's the situation Putnam wanted to generate which is this that my

twin and I are psychologically identical. We have all the same mental states but our words mean something different.

Alex Barber

OK. So that's the argument. But can I ask about the methodology of that argument? In fact can I ask both of you Barry too, about the methodology of thought experiments because Grice often uses very fanciful scenarios and then asks us to accept certain conclusions based on them, and the Twin Earth scenario was the most fanciful of all. A physicist or a geographer would not be impressed by somebody who came along and said well your theory is wrong because I can imagine a situation where it doesn't hold. So why is it OK to refute a philosophical theory using a thought experiment?

Professor Tim Crane

Well I would like to take issue with an assumption that you made there which is that thought experiments are not used in science. I don't think that is quite right. If we think of Gallileo's thought experiment about the notion of bodies, which he used to refute Aristotle's theory of motion. Gallileo said that according to Aristotle, heavier bodies fall faster than lighter bodies. Aristotle also held that if a slower moving body were attached to a faster moving body it would slow it down. If you put these two ideas together and you simply imagine an object that weighs two pounds added to an object that weighs three pounds then Aristotle's theory implies that the object that weighs two pounds would slow down the other object. However, an object that weighs two pounds added to an object that weighs three pounds is an object that weighs five pounds so it's a heavier object. So it should move faster. So we have contradicted ourselves and we have done that without leaving our armchair and we have shown that Aristotle's theory implies a contradiction and we did that by a thought experiment. So I don't think it's quite fair to say that thought experiments are just used within philosophy.

Alex Barber

Barry, do you think they are legitimate?

Dr Barry Smith

Another common objection to thought experiments is that they deal with things which are too fanciful. So in the case that Tim has been describing we talk about a Twin Earth just like this one with creatures who are all duplicates and replicas of us and we try to use this experiment to draw the consequence that two people who are alike in all their internal psychological states and using the same word, are in fact referring to different things. Now the reason we use such seemingly outlandish examples is to show that often the notions we are using when we look at them in the domestic every day case we can't really see all their features. And if we take them and test them in the most extreme environments we test them to destruction as it were we can find out that notions we thought were the same come apart when we take them to those extremes. And by looking at how ordinary notions behave in extraordinary circumstances we find out more about what they are really like.

Alex Barber

People sometimes make a distinction between externalism about words and externalism about psychological states or mental content. Can you explain what that difference amounts to and then say something about how they relate to one another?

Professor Tim Crane

Putnam's argument can be applied not just to the case of the meanings of words but as people soon discovered you could apply the argument to thoughts as well.

Alex Barber

So thinking ain't in the head.

Professor Tim Crane

That would be the conclusion which sounds very paradoxical but the argument again would go in the same kind of way that my twin and I could be both thinking about water, one of us is thinking about the real water on earth the other one is thinking about this other water on Twin Earth, twin water. If our thoughts are supposed to determine reference then we can't be

thinking the same thoughts: that is to say thoughts with the same content. So we are thinking different thoughts even though as a hypothesis of the thought experiment we are actually physically identical. So this has the conclusion that as you say thoughts aren't in the head or in other words that what you are thinking the content of your thought is not determined purely by internal properties of you. So this thesis about thought is a very radical thesis. It says that what the nature of someone's thought is it cannot be fixed simply by considering that person just on their own. That thought rather stretches out into the environment.

Alex Barber

Could I just ask you to mention some of the objections to mental externalism because not everybody is won over by the argument that you just ran through.

Professor Tim Crane

Yes. There are two main kinds of objection. One appeals to the idea of the subject's point of view on the world and the style of objection says "from the point of view of me on earth and my twin on Twin Earth things seem exactly the same. Not only does the world seem exactly the same but my thoughts seem exactly the same. Yet Putnam's argument shows that we have different thoughts. Now how do we square this with the idea that our thoughts seem exactly the same? Well the worry here is that it is very plausible to think that our knowledge of our thoughts is in some way immediate or privileged. Sometimes it is said that we have privileged access to what we are thinking. I know what I am thinking in a way, which has a kind of authority, which you don't have over what I am thinking. You can't tell what I am thinking in the same way that I can tell what I am thinking. Putnam's argument seems to threaten this because it seems to say that I couldn't tell just by reflection on my situation now whether I am on earth or Twin Earth and therefore I couldn't tell whether I am thinking about water or twin water. So Putnam's theory threatens the authority that we have over our own thoughts, the knowledge of our own thoughts. That is one kind of objection and people say that because this authoritative knowledge is a very well established fact that we should therefore be sceptical about Putnam's argument.

The other kind of objection comes from the role that psychological states play in causing and explaining what we do. It's a fundamental aspect of our common sense picture of ourselves that what we do depends upon what we think and what we want. So we are caused to do the things that we do because we want certain things and we believe that we can get them by doing other things. Consider the Twin Earth case. If my twin wants to do something, to drink a glass of water, then he will behave in exactly the same way that I do if I want to drink a glass of water, assuming that we have the same other states. There seems to me no difference in the behaviour that's produced. When I aim to drink a glass of water and when my twin aims to drink a glass of water. We are both trying to do the same thing. So why do we need to introduce a difference in thought when trying to understand that difference in behaviour? Yet Putnam's argument does introduce a difference in thought. It says that our thoughts are completely different. One of our thoughts is about water and the other thought is about twin water.

Alex Barber

So the point is something like this. Thought that depends on the environment isn't the kind of thought that explains behaviour and we should in fact stick when we are talking about mental content, just stick to the latter kind of thought.

Professor Tim Crane

Yes. I think so. The idea would be that the difference between earth and Twin Earth is invisible to the subject by hypothesis so it's not something that can show up in the thoughts that they have which explain why they do what they do.

Alex Barber

Up to this point we've been focusing on meaning and language but I would like to turn now to mental representation and in particular attempts to understand it scientifically, part of the project of naturalising the mind as it is sometimes called. Rightly or wrongly what gave impetus to these attempts was the rise of the computational theory of mind. So Tim, I wonder if you could begin by spelling out the central tenets of this theory?

Professor Tim Crane

The theory is a theory about those states of mind which represent the world. So states of mind like belief and thought and desire and judgement and so on. The theory says that when someone is in one of these mental states there are certain symbols in their head which represent things in the world. These symbols are like the symbols of a language. It is often said that they have semantic properties. That is to say that they represent things and they have syntactic properties, which are properties like the property of being a word for example. The idea is that you have a mental representation which means dog in your head whenever you represent dogs and it's the same symbol whenever it occurs. It's this combination of semantic and syntactic properties, which the representations have which lead its main proponent Jerry Fodor to call the theory "The Language of Thought". So the theory is that we think in an inner language which is these mental representations or symbols in our head.

Alex Barber

Why would anybody hold such a strange view?

Professor Tim Crane

I think there are three main reasons. The first is to show how thinking can be embodied in the material world. Thinking is not an ethereal process but it is something that must be actually realised in the physical matter of our brains. Fodor has used the analogy with the way a computer program is embodied in the hardware of the computer. The computer program isn't an ethereal thing and it involves symbols, which are physically embodied in the actual computer. So that's the first reason. The second reason is that mental processes seem to have a certain character. Thinking has a certain kind of order, which Fodor thinks is best explained by treating symbols as having this structure. For example if you infer a conclusion from a piece of reasoning your conclusion, like for example that you want to go to the pub and have a beer, you might infer that conclusion on the basis of certain arguments. These arguments actually cause you to go and do something that is to say go to the pub and have the beer. So somehow you are being caused to do something by the fact that you're reasoning in a certain way. So your reasoning must have a physical embodiment. This mental process must have a physical embodiment, which reflects the structure of the reasoning itself. And that's the second reason. The third reason is more specific which is that thought seems to have this systematic character. If you can think that John loves Mary then you are capable of thinking that Mary loves John. Fodor has argued that the best explanation for that systematic character is that thought itself has a linguistic structure; that there are separate elements that put symbols or mental representations for Mary, loves, and John. And you can re-arrange them in this way just as you can in a language.

Alex Barber

Do those who posit this language of thought ever say in physical or scientific terms what it is for a symbol of this language to represent or to have semantic properties as you put it or do they just take the notion of representation as somehow given?

Professor Tim Crane

Well it's an interesting question because on the one hand what the Theory of Mind was trying to do was trying to explain how thought can be embodied in a physical system. And it did this by saying that there are representations in your head. But as you say it doesn't explain what makes something a representation. It just assumes that there are these representations and it tells you how they interact with each other. You might think this is a problem if you were a naturalist and you thought that all the facts about the world ought to be stateable in a physical vocabulary, a vocabulary that doesn't use notions like meaning and representation. People who think that would want to add something else to a representational theory of mind. The dominant approach to that question is to say that what makes a symbol represent the things that it does is its causal connection to the environment. So simplifying a lot, the basic idea is that your mental symbol for a cat represents cats because it was caused by an interaction with a cat.

Alex Barber

But you seem to be suggesting that there is a kind of alternative position, which is just the computational theory of mind without the attempt to naturalise the notion of representation.

Well I am not sure that could be a comfortable position. And my reason is just that one of the attractions of the computational theory of mind was its materialism. This was the first of the three attractions that you listed earlier on. Surely anyone who is attracted to the theory, the computational theory of mind, for that reason, will feel obliged to give a naturalistic definition of the notion of representation as well.

Professor Tim Crane

Not necessarily. I think these are two separate issues. One is whether there are representations in your head, whether there is a language of thought. The other is whether you can give a naturalistic or a physicalist reduction of the concept of representation. Whether you can explain what it is for something to represent something in non-mental or non-representational terms. And it seems to me that you could be a materialist, or a physicalist or naturalist, if these things, or these words all roughly mean the same thing, and nonetheless believe that within the theory of mind the concept of representation is a primitive concept. It's one of the basic concepts of the theory. What we are trying to do is to explain how the mechanisms of the mind work and as long as we have got something which is consistent with what we find in the other sciences, then we shouldn't worry that we have introduced new concepts which can't be explained in physical terms. The biological concept of fitness for example cannot be explained in physical terms. Nonetheless that doesn't make biology inconsistent with physics. All that is really needed for a naturalist picture of the world is that everything should be determined by the fundamental physical facts. I am not saying this picture of the world is true. This is all a naturalist needs to say. I don't think they need to say in addition that all concepts including the concept of representation can be defined in physical terms.

Alex Barber

Tim Crane, Barry Smith, thank you very much indeed.