



## Thought and Experience

*Naturalistic theories of mental content*

### Alex Barber

Tim Crane suggested that it may not be necessary for an advocate of the computational theory of mind to give a scientific or naturalist account of representation. They could instead take it as a primitive notion, just as biologists take the notion of fitness to be primitive. David Papineau, Professor of Philosophy at Kings College London, does not take this easy route. In my interview with him we talked about his preferred naturalistic theory of mental content, what he calls “success semantics”, which is a version of tadao semantics. I began though by asking him why he is attracted to naturalism about the mind and to say what this involves.

### Professor David Papineau

I find it rather difficult to define the notion of naturalism. It's a term that's used a lot by philosophers nowadays but I don't think it's got any precise meaning. It refers to a kind of family of positions rather than anything very definite. I mean one idea is that we want philosophy to be continuous with or perhaps even part of science. I am not sure that is a very good understanding of naturalism. Or if you do understand naturalism in that way I am not sure it's a very good idea to be a naturalist because there is plenty of things that are worth thinking that are not scientific things and perhaps even within philosophy.

Another idea is much more specific. That you are a naturalist if you are not a dualist, if you think the mind is just a brain, you don't believe in any extra mind stuff but I am not sure that is a very good notion of naturalism either. I mean there are some dualists, David Chalmers, perhaps who seem to me very much naturalists and perhaps one might not be a dualist and yet fail to be a naturalist because of ones views in other areas about morality or mathematics say. I think the simplest way of understanding what naturalism is is it's how philosophy goes if you try and do it without God. I think a lot of things in our philosophical tradition seemed for a long time unproblematic. I think that when you take away God as has effectively happened within Western philosophy in the last couple of hundred years, then certain things start to be problematic and what is distinctive about naturalist philosophy is it worries about these things. For instance the existence of self-standing, fundamental, unexplained norms is something which – well this isn't straight forward – but arguably is not problematic if you have fundamental deity but is problematic without. And similarly it seems to me notions of representation would have seemed unproblematic if you thought of our minds as fragments of the Divine mind but once you take away the Divine mind and its power to imbue physical bodies with intentionality then intentionality representation becomes very problematic indeed.

### Alex Barber

What I am not quite seeing is why taking away God makes the thought that physical organisms like us have representational capacities problematic?

### Professor David Papineau

Well, if you believe in traditional God then surely one of the basic facts about the universe would be that God can represent things. God isn't urgent, God thinks, God has thoughts about you and me and the rest of the world and if one believes in the basic fundamental God, one is not going to be explaining that power in terms of something else. That is just basic to the nature of God. He is capable of intelligent thought. And maybe when you have God his powers of intelligent thought could somehow distil down into our minds. Our minds are fragments of the mind of God and thereby get the power of representation from God's power. But if you take away God and then you have left at first pass the world as described by modern science, it's not clear that there is any representation in there. It just doesn't look like that's one of the basic properties of the natural world. I mean to see the problem just think about words on paper. I mean every philosophically inclined child might well think “how come

that word Lima stands for a City on the other side of the earth?" I mean that is just marks on paper or if you are saying it it's just sounds in the air. How can sounds in the air have the power to represent something else? And well words stand for what they do because of the way we understand the words but now we come down to what makes our understandings, our mental states stand for anything else, if you think of our mental states as the brain states, you have just the same problem. How can a brain state stand for something else? And it looks like something that just cries out for explanation.

**Alex Barber**

So you are sympathetic to the computational theory of mind and that uses the notion of representation and you're keen on giving a naturalistic account of that notion. Can I ask you then for your views on some of the more popular attempts to give a naturalistic theory of the notion of representation? How sympathetic are you for example to functional role or as it is sometimes called "conceptual role theories of representation"?

**Professor David Papineau**

I am not very sympathetic at all and it is something of a change of mind for me because if you were so inclined to go back to my first book which was written now some years ago, you will see that I was a mad conceptual role semanticist then to the extent that I thought that no two people could possibly share a concept because the roles in their heads would always be different. But I have now given up on this approach almost entirely. One obvious difficulty is that if we are trying to figure out the content of a belief or concept just by seeing how it relates to other beliefs or concepts inside the individual's cognitive economy cognitive structure in terms of what inferences they are inclined to make: what they infer from what else and so on. There is an obvious difficulty which is that how does one belief or concept get a content by being related to other beliefs or concepts if they don't already have some content and if their content is going to be given in the same way. It's rather like everything is unsupported. Each belief or concept is supposed to get its content from its relation to other beliefs or concepts. And all you end up with is a kind of here's a standard picture, a net in which the nodes are concepts and the strings are dispositions to make inferences and this net is not tied up to the world at all. So you would just be looking at it as a structure of inter-related concepts with no relation to the outside world and from that point of view it could represent anything or nothing.

**Alex Barber**

Could a conceptual role semanticist perhaps reply to that by saying that the whole net, if you like to use that metaphor, isn't tied down at the periphery and that concepts like red for example are very closely related to the external world whereas concepts like a light bulb for example or letter box are very distantly related. So conceptual role semantics might work by treating the meaning of these higher order concepts as holistically given but the whole is tied down by their relation to perception and action.

**Professor David Papineau**

Once you see the need to tie the network of beliefs down to the external world then I think the motivation for bringing in the network in the first place falls away. I mean why stop with perception? Why not think of my concept of cats say and my beliefs involving that concept, as being related to cats rather than my perception of a cat and indeed why not just think of other various concepts as having their content because they bear a certain relation to things in the external world. What I am suggesting here is pretty close to a notion of meaning which is just meaning as reference. Now you might say but don't you want some other notion of meaning – my concept of cat isn't just something that represents the things in the external world. It's something that I reason with in a certain way. It works a certain way in my thinking and doesn't its conceptual role capture that? And I think a very good answer to that challenge is to say well it does work a certain way in my thinking and we do want to capture that but you don't have to build the way it so works into its meaning to capture that because after all you have got all my beliefs about cats already in the network which is considered as beliefs. So if you think of my term cat as having its meaning, its content just as its reference you can still explain all the reasoning I do as a result of my cat beliefs in terms of which beliefs I have. You don't need to build my structured beliefs into the content of my concepts because the structure of beliefs would be there even if you don't. So I see no virtue in conceptual role semantics at all. Just think of semantics in terms of reference.

**Alex Barber**

A lot of people who would agree with that response to conceptual role theories would try to spell out what the reference relation is in terms of some kind of indication relation. Do you think that that's a good way to begin to start with the simple notion of indication or causal co-variants as Fodor for example has called it and to try and deal with the problems that emerge such as the problem of misrepresentation? Or do you think there is completely some other way of doing it?

**Professor David Papineau**

Well I do think that the indicator approach is an improvement on the conceptual role approach in that it puts the relation of our concepts to their reference kind of at the centre of analysis of meaning. But even so, it seems to me there is something very unhappy about trying to analyse that notion of reference by starting with indication.

The basic idea there is that a term stands for the circumstances that leads you to apply the term and it's a natural way of thinking given our philosophical tradition but I don't think from another perspective it's particularly natural at all. What I have in mind here is if you think in that way it becomes difficult to see how anybody can have false beliefs. If the content of the beliefs is a circumstance that leads you to form it well then you will form it when its content is there and it will be true. But surely it's all too easy for humans and any other thinkers that there may be to have false beliefs and so you may well think that the theory that makes false beliefs problematic is starting in the wrong place. Let me try and make this more graphic by giving you a simple example. Supposing – I mean this is in fact reasonably ethologically accurate – there are monkeys, Vervet monkeys that will go and hide under bushes whenever they get into a certain brain state. Hiding under bushes is a good way to avoid being eaten by eagles. I say that those facts alone are enough to tell us that that brain state represents there is an eagle near by. But note that the story I have just told is consistent with the monkeys being terrible at telling at whether eagles are nearby. Maybe only one in a hundred times when they get into this brain state is there actually an eagle nearby. So the brain state's a lousy indicator of an eagle but nevertheless I want to say it represents an eagle and that is shown by the fact that when the monkeys are in this brain state they will act in a way that is advantageous, that is appropriate to the presence of an eagle and that's what shows their brain state represents the presence of an eagle. Now I haven't spelt out explicitly what kind of theory of representation I am assuming here but I hope that this example makes it clear that indicator semantics is a misguided place to start.

**Alex Barber**

Perhaps I could follow on by just asking you to spell out an alternative account of the reference or representation relation that copes with mis-representation, at least in this case.

**Professor David Papineau**

What I had in mind there when I was talking about the monkey example was what's often called success semantics. If you have a belief like representation, according to success semantics the way to figure out its content is to look at what actions it causes. Look at what actions that representation prompts. And then ask in what circumstances would those actions be a sensible thing to do? What circumstances would if they are obtained, lead that action to succeed and you can think of the content of the belief then as success conditions? The conditions under which the actions of belief prompts will succeed. I mean it's a terribly intuitive idea. I want a beer. I have some brain state that makes me go to the fridge. You can infer from that that the brain state must represent to me that there is beer in the fridge because it's in those circumstances that my action, given my desire, it's a sensible thing to do.

**Alex Barber**

OK. That sounds very intuitive but I am not sure whether it's intuitive because it's vacuous or empty or uninformative. I mean your theory in effect is this. That a state, a brain state, represents those circumstances in which the actions that it causes would be sensible. So you are relying on the notion of an action being sensible and that sounds at least as in need of reduction or explanation as the notion of representation that we are trying to explicate or

explain so can you say something about why it is OK to appeal to the notion of an action being sensible?

**Professor David Papineau**

Good. When it comes to sophisticated thinkers like human beings, I think the appropriate notion of an action being sensible, an action given the circumstances being apt to succeed, is a notion of an action which will lead to the satisfaction of the agents desires. So in that kind of context where you have got a structure of beliefs and desires the thing to say about beliefs is that the content for any belief is that circumstance that will guarantee that the actions it prompts will lead to the satisfaction of the desires it's operating in concert with. And that's a bit of a mouthful but if you think of the fridge example I hope it will become clear. But you're right. I am assuming here crucially the notion of a desire being satisfied and that's itself a representational notion. I mean desires are for certain states of the world. I desire that I have a beer in my hand. I desire that there be peace on earth. I mean desire represents, in a different way from beliefs, but still represents that the world should be a certain way. And if one finds representation puzzling then one ought to find the notion of desire and desire satisfaction puzzling. Now this is the place where I and others, most prominently Ruth Millikan, appeal to biological considerations. We say that to understand the satisfaction conditions of a desire, to understand what a desire is aimed at, its target, you need to think in biological evolutionary terms. You need to think what affect was that desire selected to produce? This is to apply the notion of biological function to desires. In general in biology we talk about items having a function when they produce some affect and what's more the fact that they produce that affect is the reason they are here now. The polar bear's white fur has the function of camouflaging it and the standard way of understanding that in biology circles is to say that means the polar bear's white fur does camouflage it and what's more that explains via a story of natural selection why the polar bears have white fur now. In the past ancestors of polar bears with white fur were camouflaged and as a result they had more offspring and that's why all polar bears have white fur now. Similarly one wants to say here my desire for water when I am thirsty is a state which has the function of getting me water and that's a matter of its being here now and me because ancestors who had the state and were thereby lead to water survived and reproduced better. So what I and others want to do is to take the basic idea of success semantics of beliefs and desires having co-ordinated contents in the way the fridge example showed and then place all that in a biological functional context to think of the mind or the belief desire part of the mind, as a system which has been designed by natural selection to allow agents to produce certain results as guided by their beliefs. And so I think that you are right just appealing to an action being sensible or appropriate or satisfying desires without saying anything more would be no good. But once we think of desire satisfaction as a biological matter, as a special case of biological function, then I think that an appropriate explanation has been given.

**Alex Barber**

You have been suggesting that a theory of mental content might draw on a notion of function that's grounded in human evolutionary history. This may work for some mental states but I am worried about mental states like hoping that your computer doesn't crash for example. Our evolutionary ancestors didn't have computers so that particular hope, that particular desire can't be accounted for in the terms that you suggest, can it?

**Professor David Papineau**

OK. That's a good question and I obviously need to have some answer to that. I started with the example of the desire for water because that's a kind of desire one might plausibly think is the result of biological evolution, that you desire water when you become de-hydrated. But obviously as you have just pointed out, we have many desires for things that weren't around in the period when our ancestors were evolving and the content of those desires obviously can't be explained in the terms of selection genes over generations. It doesn't follow from that that we shouldn't think of even desires like that as having functions, indeed biological functions. And there are a couple of options here. One line of thought is to say that well, those concepts are learned but learning is a selectional process just like the inter-generational selection of genes. I mean just to take an example. Take standard instrumental conditioning: pigeons being taught to peck a bar in order to get food. You can think of their learning this as a matter of selection. First they make random movements. Some movements are followed

by food and other ones not. They embody some mechanism that encourages the neural pathways that give rise to food and discourages other ones and as a result they end up being programmed to peck the bar in the circumstances. And in line with that it's perfectly natural to say that the pigeons are pecking the bar in order to get the food. The function of their pecking the bar is to get the food. And there is a case where you get the same talk of function as a result of seeing the thing as an item that has been selected. But the selection here isn't inter-generational selection of genes - its selection in learning. But there is a different option which is the one that is embraced by Ruth Millikan. She says no, the processes that give rise to beliefs and desires in learning do instil functions in those beliefs and desires. Not because learning is a selectional process but because the mechanisms that give rise to beliefs and desires themselves have the function of producing certain items that will serve certain purposes. She uses an analogy. Take the chameleon's mechanism that turns it the colour of whatever background it is - OK? Once it's turned that colour, that colour in its skin has the function of disguising it against that background. You could say the redness of its skin if it's on a red sofa as you are now, is designed to camouflage it against the sofa. But it's never been on that sofa before. And its ancestors may never have been on an item of that colour. So the function of the red skin, Millikan says, is a derived function and it derives from the prime mechanism, the colour producing mechanism, whose purpose is to give the chameleon skin the colour to disguise against whatever background it is. So Millikan similarly thinks that we have in us mechanisms that are designed to give us concepts that will relate us to items in our environment whatever they might be even if they're items that were not around in evolution. And so she thinks that there are specific and non-selectional mechanisms that give us concepts and thereby beliefs and desires for items in our own environment even though those items were not themselves present in our evolutionary history. Rather the mechanism is a kind of general mechanism, like the chameleon's mechanism is a general mechanism for making it the same colour as the background whatever that might be; we have in us the concept producing mechanism, general mechanisms for relating us to features of our environment whatever they might be.

### **Alex Barber**

Let me ask a different question about the appeal you make to evolution and it's one that you will be familiar with. Suppose that I am standing next to a tree in a swamp and the tree and me are both struck by lightning and I dissolve into nothing. At the same time by astonishing coincidence the tree is converted into a physical replica of me out of different molecules. This replica, let's call it "swamp Alex", moves exactly as I did. It leaves the swamp. It encounters and then seems to recognise and to speak with my friends in English. It moves into my house and starts writing Open University Course materials. And no one can tell the difference. Perhaps I am this Swamp Alex now. If I were, would I have beliefs and desires? Now it seems to me that you are committed to saying that I don't; that I wouldn't. And that seems very counter-intuitive.

### **Professor David Papineau**

Well I used just to say you don't. I don't care if you look like Alex Barber. I mean you don't have any beliefs and desires if you are a swamp man. But I have rather changed my mind and I am inclined to tell a more complicated story now and perhaps I can tell you why. My original line was well it might be counter-intuitive to say that a swamp man has no beliefs and desires. But we have got a good theory here; a semantic theory and good theories often make us revise our intuitions a little bit. Once people had got a theory of what gold was they realised that certain substances that looked like gold were not really gold and now we have got a good theory of what representation is we realise that certain people, beings, who look like they are representing aren't really representing. And I just used to say well lets bite the bullet and override any intuitions we have about swamp man being a believer and a desirer. That was my old view.

But I have now been persuaded out of it. And I can remember a student of mine pressing me and saying look, your view is horrible. Suppose you came across a swamp person. Suppose you came across a swamp child. In our view as far as I can see it would be all right to kill the swamp child and eat it as meat. And I said it would be wrong to harm the swamp child, swamp person, because I have never denied that they had feelings. I never denied that they would be conscious. It seems to me that they would be conscious and have feelings. It

would be something that was like for them. I am just saying that they don't have any representational state. So I thought I could get out of this argument by saying well since they would have feelings it would be wrong to kill them and eat them. But my student didn't let me off. He said, "No, no, after all you kill cows don't you?" And I said, "Well I guess so." And he said "well look - they have feelings but you think it's alright to kill them provided you kill them quickly and painlessly and no doubt you think that's OK but it's not OK with higher animals, humans, perhaps chimpanzees because these higher animals have hopes and fears and desires for their children and so on." And I agreed with the student and I agreed that it was precisely the ability to form plans, have hopes and projects that made it wrong to kill beings with high cognitive powers. And my student pointed out that is just what I said that swamp people didn't have. They didn't have any representational state so they didn't have any power to form plans and anything like that so I should happily kill them. And it seemed clear to me that that was wrong. That it would be wrong to kill a swamp man even more so to kill a swamp child. So I thought about this long and hard and I think the right answer is now rather different. And it's to do with the fact that swamp people are made up. There aren't actually any swamp people. I think if there were lots of swamp people, well that would just show that the Tadao semantics theory was wrong but since there aren't, the Tadao semantics theory could be right. I mean if there were lots of swamp people not only would the Tadao semantics theory be wrong but it would be wrong to kill them as well for that reason.

Perhaps I could make the point clearer like this. Think of it rather like gold or some other scientific concept, water might be a better one. We start off with an everyday notion of water - this odourless, colourless liquid and then we get a scientific theory of its underlying essence. It's made of H<sub>2</sub>O and that's an illuminating and helpful thing to know.

I wanted to think of Tadao semantics in the same way. We start off, and this is what we talked about in the beginning, we start off with an everyday notion, a common sense notion, of beliefs and desires and their representing certain things and then I think we should regard Tadao semantics, as telling us what in the actual world is the scientific essence of those states. It's that they have a biological, selectional history. And that's a useful and illuminating thing to know. To think of these states as having a biological and selectional history, gets us to some deeper and more important feature of their nature. So I regard the swamp man argument as no more effective than this. Somebody comes along and says, look water can't be H<sub>2</sub>O because imagine that all our rivers are full of X Y Z. That's not a good argument against a claim that in the actual world water is H<sub>2</sub>O. No more do I think it's a good argument against the claim that representation involves selectional history to say imagine that lots of the people walking around with beliefs and desires don't have any selectional history. The fact that this possible world with swamp men in doesn't cut any ice against the theory at all.

**Alex Barber**

David Papineau, thank you very much indeed.