Metaphysics // Fall 2021

Handout 12

Internal realism: Putnam

Representation Putnam first explores the question of representation which is also the question of reference. In order for any material sign to refer to an object (or: to be a sign at all), physical resemblance alone is not enough. One problem, not mentioned by Putnam, is that anything resembles anything else (Goodman). Another problem is that representation requires intentionality. So, as Putnam notes, some philosophers (Brentano) concluded that the mind must be non-physical or better non-material, precisely because material objects lack intentionality.

Putnam thinks that this fails to establish immateriality of the mind, but the question remains: how is reference possible?

Some very clever thinkers, like Florensky, believed in the magical connection between names and objects. This we put aside as a queer religious doctrine. With Locke, we believe that names are arbitrary, and that there is no intrinsic connection between a particular name and a particular object. But more generally, nor is there a connection between a mental representation (image, idea) and the thing represented. As Wittgenstein asked, supposing I'm now thinking of some NN in New York, what makes my thought about *NN*?

To dramatise this problem, Putnam imagines a planet whose inhabitants never encountered trees. Suppose their spaceship drop an accidental replica very similar to our earthly tree. In this case, he says, we won't have a mental image of a *tree*, only of an image of something qualitatively similar to a tree.

It is hard, I think, to take these intuitions seriously, since in many of these unusual cases we should rather plead agnostic, instead saying something definitive. For instance, if the alien replica is an artefact, a prior question is whether it itself represents a tree. If it doesn't (and arguably it doesn't), then it's a moot question whether our images of it represent a tree. The issue depends, in the first place, on our view of artefacts.

Anyway, let's turn from images to words. Even if a person who doesn't speak Japanese could say Japanese words under hypnosis, and even if he had a feeling of understanding Japanese (?), we still wouldn't say that he understands Japanese.

Putnam's lesson is this. There is no intrinsic connection between mental representations, or signs, and what they represent (refer to). By ' intrinsic ' he means a connection specifiable independently of the causal relation and of the dispositions of the user (speaker). And the other way round, for a proper representation we demand that it came about in a ' right ' causal way, and that it is accompanied by the ' right ' dispositions of the user.

Brains in a vat. I won't spend time to describe this scenario, since it is exactly the scenario of the *Matrix* (minus Morpheus' holdouts—this proviso is vital).

Putnam asks: supposing that we are brains in a vat, could we think or say that we are brains in a vat? He wishes to argue that we couldn't. Here is the claim:

(12-1) For every x, if x is BIV, then: x's utterance or thought ' I am BIV ' is self-refuting.

As Putnam notes, this should be understood by analogy with the cogito proposition:

(12-2) For every x, x's utterance or thought ' I do not exist ' is self-refuting. [we ignore the selfhood complications, obviously]

A Turing test for reference. Putnam deploys Turing test to show that a BIV can't refer to objects. That is, in the course of a dialogue a human interlocutor will discover that the other party (a machine) does not speak the same language and does not refer to the usual objects.

And why, exactly, not? Because, says Putnam, although a machine could converse fluently about all sorts of subjects, it couldn't *recognise* objects if presented with them. Not just that, however: we are also able to ' deal and handle ' those usual objects. So there is not only a failure of a passive perceptual ' recognition ', whatever this is, but also of action and engagement with those objects. A machine doesn't have the practical ability to manipulate the objects.

Putnam puts this difference in competence in terms of ‘ language entry rules ’ and ‘ language exit rules ’. The former take us from mental images (can we call them representations?) to utterances. The latter takes from utterances to non-verbal actions. 11

Now Putnam admits that this argument is more convincing for Churchill-drawing ants. With ants, we can easily say that they don't recognise Churchill when they see him. Nor do the (speaking) ants practically interact with Churchill (well, this is already a bit stretched!). But in any event, what's the intrinsic limitation of the machines that we now imagine? There is no causal connection between cabbages and kings that the machine is discoursing about and the states of the machine. Well, Putnam immediately notes, there may be a causal connection via the programmers. But it is too ‘ weak ’. 11

This weakness argument is a bad one. For by the same logic, we can't now talk *about* dead people, people or places we have never encountered face-to-face, and so on. Putnam has another bad argument when he says that the machine discourse is insensitive to the continued existence of the objects. 11

I think his best shot is simply to imagine a world where cabbages and kings don't (physically) exist at all, but where the machines converse continuously and smoothly ‘ about ’ them the same way as we or other machines converse in the world that does contain cabbages and kings.

That's still not ideal, I think. The machines will have, by assumption, no language entry rules, because there are no cabbages to begin with. And they have no exit rules: by another assumption, they have no sensory organs. Let's put it this way: they have programming routines to recognise cabbages when presented with them (brought from another world!). But why can't the very same machines be equipped with just those routines and sophisticated sensors? Again, with ants, however intelligent, it would be possible to say that they have no sensory organs to recognise Churchill as Churchill. Or if they do, then they aren't ants any more. But to say, as Putnam seems to be saying, that the machine's limitation is just the absence of cameras and other sensors is not quite interesting, since we already have machines with very sophisticated sensors.

In any event, where is the *test* supposedly analogous to the Turing test? At what point do I conclude that the machine doesn't refer, ‘ clearly ’ so, as Putnam says? I think it's the point when I say, ‘ Fetch me the cabbage! ’, and the machine helplessly lingers. But that's like playing a cruel joke on an invalid! 12

The case of BIVs. Actually, the qualms about exit rules turn out to be of little consequence (I think). At once Putnam recognises that BIVs may have quasi-organs. I take it this means that they may well engage with cabbages if presented with them. 12

The problem is with entry rules, that is, with the causal flow from the world to the talk. Recall where the argument is going. We need to show that BIVs can't say that they are BIVs. So in order to say, legitimately, that they are BIVs they must be in contact with BIVs. But they are not in contact with *brains* or *vat*, but only with brain-images or vat-images, Matrix-like. 13–14

The causal contact is a pre-condition for successful reference. It is not fulfilled for BIVs. So they can't refer to BIVs. So they can't say of themselves that they are BIVs. 16

YSB