

External Reference and Residual Magic

Paul Schweizer

University of Edinburgh

Abstract: Putnam famously criticizes the traditional theory of meaning and reference, and argues that no internal configuration of a cognitive system is able to capture the intended objects of linguistic reference. Indeed, a ‘magical connection’ between language and world would be required for the traditional theory to work. Putnam attempts to replace the magical dependencies assumed by the traditional theory with an externalist account based on causal chains and direct access to the world. However, I argue that he does not carry the exorcism far enough, and that his own theory assumes a literal and substantive interpretation of reference which itself must be relinquished if his critique of the traditional view is carried to its natural conclusion.

1. Old versus New

In one of the more seminal papers of late 20th century philosophy, Hilary Putnam (1975) argues against the traditional theory of meaning and reference, and advocates the now widely held view that ‘meanings ain’t in the head’. According to this view, no mere internal configuration of a cognitive system, be it computational, neurophysiological or conscious/phenomenal, is able to capture the intended objects of linguistic reference. Hence the representational capacities of internal states are, in the general case, too weak to carry the referential burdens of natural language.

When criticizing the traditional theory, Putnam makes a number of amusing allusions to its underlying ‘magical’ properties. For example, in (1981) he compares it to the belief held by various “primitive people” that “to know the ‘true name’ of someone or something gives one power over it. This power comes from the *magical connection* between the name and the bearer of the name” (p. 3, his italics). And again in (1989) he observes that the traditional theory of reference seemingly requires a ‘noetic ray’ that emanates from the mind and pinpoints the object of reference.

To replace the ‘superstitious’ belief that meanings are some type of psychological entity or rely on occult mental powers, Putnam supplies an alternative ‘externalist’ account based on direct appeal to the environment, and on the communal division of linguistic labor. The reference of natural kind terms such as ‘water’ is not determined by internal states and images, nor by qualitative identifying properties that the individual speaker may associate with the term. Instead, the extension is based on the actual microstructure of the environmental liquid, the stuff in the world jointly accessed by members of one’s sociolinguistic clan.

However, I argue that Putnam does not carry the exorcism far enough, and that his own externalist theory still places crucial reliance on covert ‘magical’ powers. Putnam assumes a very robust and substantive interpretation of the relation of linguistic reference, and I argue that such an interpretation tacitly depends on strong internalist assumptions which are at odds with his own critique of the traditional theory. Hence if Putnam’s critique is followed to its natural conclusion, then the robust and traditional view of reference must itself be relinquished in favour of a much more modest prescriptive account.

2. Brains in Skulls

In (1981) Putnam argues that a community of envatted brains might undergo internal states qualitatively identical to members of some normal English speaking community, but they would still fail to refer to real physical objects such as brains and vats with their words 'brain' and 'vat'. Brains in a vat are deprived of the right kind of causal links to their actual surroundings. Their phenomenal and linguistic episodes are not suitably related to the normal environment of spatially located macroscopic objects and direct physical interactions, and this cuts off their referential access to items that *we* can successfully talk about.

So a key aspect of Putnam's critique of the traditional theory relies on drawing a sharp distinction between our case and that of a community of envatted brains. In contrast to the hapless brains in a vat, *our* phenomenal and linguistic episodes are underwritten by direct and robust patterns of interaction with other human organisms and the surrounding physical world. For example, unlike disembodied brains, we have eyes, and H₂O is the salient source of reflected light that actually stimulates our retinas when we have experiences of 'seeing water'. In English speaking communities, these causal interactions are accompanied by assorted linguistic behaviours involving tokens of the term 'water', and these events take place in a shared spatial context. All of this is central to the overarching circumstances in which our natural language practices are embedded.

Thus when it comes to the nature of our causal interactions with the environment and our fellow language users, it's obvious that we differ markedly from the unfortunate inhabitants of Putnam's vat. And Putnam takes this difference in causal circumstances to have profound *semantical* effects. Unlike the referentially disabled vat dwellers, our words

such as 'brain' and 'vat' *really do refer* to actual brains and vats in the external world. Putnam holds that there is a substantive fact of the matter regarding the intended interpretation of English expressions, and he uses a variety of traditional locutions to convey this realist view of the relation of reference. So, in the case of a community of normal human agents, properly situated in their physical context, it is now variously said that reference is 'brought about', 'occurs', 'takes place' and is 'successful'.

On such a view, reference is successful in our case precisely because the theoretical stipulations of externalism are satisfied. 'Water' refers to H₂O in Earthian English because the appropriate causal ties to the environment, history of word use in the English speaking community, dispositions of native speakers, etc., actually obtain. H₂O has exactly the right kind of spatial proximity and interactive ties required. Reference to real water 'occurs' in our case, but not in the case of the transplanted grey matter, because of a fortuitous pattern in which internal states and external factors satisfy the stipulated conditions. We enjoy the appropriate kind of alignment between mind, language and world.

But contrary to Putnam's story, I would maintain that there's nothing about a *brain in a skull* that could legitimately exploit the external factors which distinguish our environmental context from that of an envatted brain. Putnam's critique of the traditional theory assumes the standard narrow interpretation of psychological states, and if this interpretation is maintained, then there's nothing special about an embodied brain that could give the expression 'water' the power to refer to the environmental liquid to which it bears all of the appropriate causal relations. How could a brain in a body have the power to reach out and utilize the external factors upon which the relation of reference is said to depend? How could it access any of these outside influences, or have the ability to

semantically hook-up to the 'right' set of environmental circumstances?

Contrary to mentalistic renditions of meaning, Putnam persuasively argues that nothing in the head and no mere intentional state is able to underwrite reference to natural kinds such as water - we need to appeal to the environment itself and our causal relations to the kinds in question. But the same line of reasoning equally implies that nothing in the head is able to underwrite cognitive access to the environmental factors that support our 'successful' reference to water, brains, vats, etc. Invoking causal connections and the intended external relations does not expunge 'occult' forces from the picture, but instead it merely postpones their deployment. Rather than solving the fundamental problem it simply pushes it one level further away.

3. Cause versus History

Putnam eschews the 'magical' presuppositions of traditional accounts, but in the new 'causal' theories of reference, it is *indexicality* and *ostention* that tacitly preserve the old magical ties between mind and world. For example, consider the 'initial baptisms' which supply the spatio-temporal origins of the use of proper names. These baptismal events are fundamental to the new theories of singular reference, and they rely on a direct ostensive tie between sound and thing. But it is crucial to note that there is no physical or properly causal link formed by such baptisms. There is merely a ceremony in which a sound is produced in some local context consisting of a myriad of different particles and aggregates, the intended object of reference being some fluctuating and loosely-defined collection of molecules off of which these sound waves presumably bounce. But the sound waves don't stick to the person as some kind of physical tag for future reference (perhaps *branding* would be a ceremony more conducive to the needs of a causal theory of naming).

There is no physical trace, imprint or strictly causal connection established by the baptismal act. From a naturalistic point of view, the effects of this ritual are more or less intangible. Indeed, it bears a rather uncanny resemblance to a magical rite. This elusive and momentary rite is then supposed to provide the naturalistic cornerstone of singular reference. It is said to allow contemporary speakers to reach back thousands of years into the past and ‘refer’ to various individuals in antiquity. But surely this requires an inescapably *intentional* correlation between sound and thing – initial baptisms and chains of use may form vital elements in the intended correlation, but purely mental factors still play an essential role.

As another example of covert magic, consider the role of *indexicality* in the semantics of natural kind terms such as ‘water’. Putnam relies on the idea that we have ‘direct access’ to this liquid in our surroundings, and thus ostensive appeal to the actual stuff in the world underpins our ability to refer. But what is the precise nature of the causal chain between human language users and H₂O? I would argue that ‘direct access’ via such chains is doubly ambiguous in this type of naturalistic scenario. First it’s ambiguous as to exactly *where* in the long and complex causal sequence the object of reference is located. Mere gesturing to perceived water is supposed to indicate a unique *liquid*, rather than, say, the retinal images produced by observing this liquid, or the ambient light that produces these images, or the shimmering surface that reflects the light.

And it’s ambiguous as to exactly *what* at this point in the causal chain is the intended object of ostension. The earthly realm contains a host of chemically impure liquids in which normal H₂O (along with its various isotopes) serves as the primary solvent, and in which any number of other chemicals abound in both solution and suspension. The term ‘water’ is supposed to pick out the equivalence class supported by

the same unique molecular category in all these complex and chemically impure liquids. Via mere ostension we are granted miraculous access to the *universal* molecular category H₂O, rather than just to some huge disjunction of environmental liquids in which H₂O is the primary solvent, or to only those particular samples of liquid that have been directly perceived and pointed at. In order for the story to work, there has to be something very special about brains in skulls, something that enables us to pick out the correct causal circumstances and the *imperceptible* microstructures underlying macroscopic regularities in the external world. And surely this purported access to generalized microstructural *types* cannot be accounted for as a natural effect of encountered instances of H₂O. Something spooky is still going on.

Perhaps an externalist would reply that it is the *intention to refer* on the part of the language user that hooks us up to the relevant causal connections and thereby gives us direct access to the environment. But what is this ‘intention to refer’, other than a special type of qualitative feeling, a purely subjective internal state which, according to Putnam’s own critique, should have no semantical power. By hypothesis, the envatted brains have internal states identical to our own in narrow terms. So brains in skulls should not have the psychic power to reach out and connect with the salient *causal chains*, any more than brains in a vat have the power to reach out and connect with real brains and real vats.

In the former case the gap separating internal states from external factors may seem smaller, but it is no less unbridgeable. And it doesn’t help to add some internal monologue attempting to express one’s intentions, e.g. ‘I intend to pick out the natural kind sharing the same_L relation to *this* stuff in the environment’. Such an incantation is just a pattern of sounds pronounced internally – it doesn’t have the power to reach outside the head and fix upon external microstructures. And indeed, this intention is far more rigorous and

sophisticated than those possessed by average English speakers who presumably refer without it. Rather than aiding reference, such articulations are part of a subtle and very specialized language game (Wittgenstein, 1953) played by professional philosophers.

4. Conclusion

In the original Twin Earth thought experiment, doppelgangers Oscar₁ on Earth and Oscar₂ on Twin Earth are in identical psychological states with regard to the environmental liquid they refer to with the term 'water', but it refers to H₂O on Earth and XYZ on Twin Earth. The scenario is strategically set in 1750, when knowledge of the molecular structure of the respective liquids could not be used to differentiate their psychological states. Internalism is taken to be refuted because all salient internal factors remain constant while extension varies. For this type of argument to go through, we must agree that natural kind terms such as 'water' are rigid. Thus we must accept Putnam's semantical principle

(i) (For every world W) (for every x in W) (x is water \leftrightarrow x bears same_L to the entity referred to as 'this' in the actual world W₁)

as opposed to the alternative principle

(ii) (For every world W) (for every x in W) (x is water \leftrightarrow x bears same_L to the entity referred to as 'this' in W)

Putnam takes (i) to properly characterize the reference relation in English, yet this principle does not supervene upon any *external* factors or circumstances. It is a prescriptive principle which characterizes what externalists hold to be the correct *intentional attitude* regarding the meaning of 'reference' in English. Hence even on Putnam's account, when fully purged of the covert magical forces assumed by the traditional view, the analysis of reference ultimately boils down to the prescriptive

characterization of an internal phenomenon. If the externalist nonetheless wants to maintain a traditionally robust and substantive interpretation of reference, in which it literally ‘takes place’, ‘occurs’ or is ‘brought about’ under the correct circumstances and ‘fails to occur’ in others, then residual magic is still at work.

Literature

Kripke, S. 1972. *Naming and Necessity*, Harvard University Press.

Putnam, H. 1975. ‘The Meaning of ‘Meaning’’, appearing in *Mind, Language and Reality*, Cambridge University Press.

Putnam, H. 1981. ‘Brains in a Vat’, appearing in *Reason, Truth and History*, Cambridge University Press.

Putnam, H. 1989. *Representation and Reality*, MIT Press.

Wittgenstein, L. 1953. *Philosophical Investigations*.