*Naturalism, Realism, and Normativity*
By Hilary Putnam, edited by Mario de Caro
Harvard University Press, 2016, 248pp., £39.95 (hbk)
ISBN: 9780674659698

This review is due to be published in *Philosophy*. This is a pre-print and may be subject to minor alterations. The authoritative version should be sought from *Philosophy*.

Hilary Putnam's *Realism with a Human Face*[1] began with a quotation from Rilke, exhorting us to 'try to love the *questions themselves* like locked rooms and like books that are written in a very foreign tongue'. Putnam followed this advice throughout his life. His love for the questions permanently changed how we understand them.

In *Naturalism, Realism, and Normativity* – published only a few weeks after his death – Putnam continued to explore central questions concerning realism and perception, from the perspective of 'liberal naturalism'. The volume's thirteen papers were written over the past fifteen years (only one paper is new), and they show a man who fully inhabited the questions he loved.

But this presents a certain difficulty for the volume's would-be readers. Since Putnam had engaged with these questions so many times before, much of his reasoning take place 'off-stage', and only appears in the papers in thumbnail form. Compounding this problem, chapters 2, 4, 5 and 8 are replies by Putnam to papers written about him (by Williams, Sosa, Boyd and Wright), but Putnam's replies are printed here *without* the original papers. Conversely, chapters 7 and 11 are papers that Putnam wrote about specific authors (Dummett and Block), but they are printed here without *those* authors' replies. Unless, then, you are really *into* Putnam, I expect that some of this volume will fly past too fast. But, if you are really into Putnam – and I am – then there are some golden moments here.

Indeed, the main significance of this book is that it shows – implicitly, but very clearly – quite *how much* of Putnam's contribution to his philosophy is

[1] Putnam, H. *Realism with a Human Face*, (ed.) James Conant (London: HUP, 1990)

continuous with his 'The Meaning of "Meaning"' (1975).[2] That's what I'll try to highlight here.

# 1. Semantic externalism

Putnam tells us that, in the mid-1960s, he began to question the idea 'that the meaning of a word could be given by "semantical rules"' (204).[3] This drove him towards semantic externalism: the claim that reference is generally fixed by '*other people and the world*' (206). In the immortal slogan, '"meanings" just ain't in the *head*!'[4]

What everyone remembers about Putnam's semantic externalism, of course, is *Twin Earth*. We are introduced to Oscar and Toscar, respectively sipping $H_2O$ and XYZ, both using the word 'water', and we asked about the referent of that word in their mouths. Putnam reports that the 'majority of people' share his intuition that Oscar's word 'water' refers *only* to $H_2O$ (209, 217).

Confession time: I *don't* share that intuition.[5] But that doesn't matter much. Although Putnam describes it in this volume as his '"Twin Earth" argument', *Twin Earth* is less an argument than an *illustration* of Putnam's general approach to meaning. And, whilst I differ in my verdict on the illustration, I agree with the general approach.

Indeed, Putnam himself says that *Twin Earth* first occurred to him in December 1972, well *after* he had arrived at his semantic externalism (208). The thoughts that had actually *led* him to semantic externalism were:

1. *A via negativa*: if kind terms are mere descriptions, or governed by semantical rules, then '*what are those rules?*' (205; cf. Putnam's comments about Kripke, on pages 209–10).

---

[2] Putnam, H. 'The Meaning of "Meaning"', in his *Mind Language and Reality* (Cambridge: CUP, 1975), 215–71. It is also worth mentioning its predecessor: 'Meaning and Reference', *The Journal of Philosophy* **70**:19 (1973), 699–711.
[3] See also his 'Replies', *Philosophical Topics*, 20:1 (1992), 349, 358
[4] Putnam, H., 'Meaning and Reference' (1973), 704; 'The Meaning of "Meaning"' (1975), 227.
[5] See Button, T., *The Limits of Realism* (Oxford: OUP, 2013), 183ff.

2. *Considerations of actual practice*: such as the division of labour, the reference-fixing role played by paradigm samples, and the way in scientists have actually looked for '*hidden structure*' (206–8).[6]

Elsewhere in the volume, Putnam also hints at a third route which led him to externalism, namely:

3. *A transcendental thought*: that 'solipsism [is] unavoidable in any picture that limits mental life to what happens inside our heads' (166, 226).[7]

These are three big ideas that *really matter*. They – and not our somewhat malleable intuitions about Oscar and Toscar – are what should push us towards some form of semantic externalism.

Having once embraced semantic externalism, Putnam never let it go. He clung to semantic externalism steadfastly throughout his internal realism.[8] Indeed, semantic externalism provided one of the two premises in his (in)famous brain-in-vat argument (219–221).[9] By 1981, then, Putnam explicitly believed that Oscar's brain, Toscar's brain, and a brain-in-a-vat can be (type-) identical, *without* sharing the same thoughts. In a slogan: since *meanings* ain't in the head, *thoughts* ain't either.

Since thoughts are meant to be in the mind, McDowell concluded from this that *minds* also ain't in the head.[10] Putnam essentially agreed (181, 224).[11] And this leads to a surprising upshot: *Minds ain't brains; they ain't even* realized *by brains.*

These deep ideas all flow naturally from 'The Meaning of "Meaning"'. Moreover, they underpin much of this volume. In the rest of this review, I'll explain how.

---

[6] Cf. Putnam, 'The Meaning of "Meaning"' (1975), 235 and my *The Limits of Realism*, (2013) 182–3.
[7] See also 'The Meaning of "Meaning"' (1975), 220
[8] Putnam, H. *Reason, Truth, and History* (Cambridge: CUP, 1981), 18–19, 22–5.
[9] Putnam, H. *Reason, Truth, and History* (1981), ch.1; Putnam, H. 'Replies', (1992), 369. See also my *Limits of Realism* (2013), 118.
[10] McDowell, J. 'Putnam on Mind and Meaning', *Philosophical Topics* **20**:1 (1992): 347–408.
[11] Putnam, H. 'Replies' (1992), 356–7, 368–61;

## 2. Mind and perception

Putnam pioneered functionalism in the late 1960s. At that time, he thought of functional states (in humans) as realized entirely by *brain* states.[12] But we just saw that Putnam's semantic externalism entails that brains (alone) don't realize minds. And Putnam came to *reject* functionalism.[13]

That, however, was an over-reaction. The problem we just identified does not concern functionalism's core claim, that minds are multiply realizable. The problem is only with the internalist assumption, that the realizers are in the head. Recognizing this, Putnam *returned* to functionalism, but of a more liberal sort, which conceives of functional states more widely (167–8, 196).[14]

On the one hand, then, we have mere *brain states* (or whatever they realize), which are *caused* by objects in our environments. On the other hand, we have *beliefs*, which are *about* objects in our environments and which can be *justified* or not. We therefore need to say something to link the *narrow* with the *broad*; to link what is merely *caused* with what can be *justified*.

McDowell's *Mind & World*[15] influenced Putnam's thoughts about this link for a very long time,[16] particularly concerning the justification of perceptual beliefs. However, as Putnam explains in this volume, he gradually drifted away from McDowell and towards Block. In this volume, Putnam agrees with McDowell that justification requires concepts; but Putnam agrees with Block that there are qualia. Let me explain this in more detail.

Putnam describes qualia as 'unconceptualized' (147, 162–3). This is misleading, given how much Putnam thinks we can *say* about them. However, what he means here is fairly simple. As in §1, concepts have some *external* component (their extension). But even creatures that lack *conceptual* capacities (in this externalist sense) can have qualia. So, qualia are *purely* in the head. Indeed, Putnam suggests that they might ultimately be *identified* with (features of) brain states (160, 175, 183). So, on Putnam's final picture, nothing prevents us from saying that Oscar, Toscar and a brain-in-a-vat all

---

[12] Or maybe brains + central nervous systems; but it's easier just to speak of brains, and won't change any of the points I want to make.

[13] Putnam, H. *Representation and Reality* (Cambridge MA: MIT Press, 1988) 73–5.

[14] But this retelling of Putnam's intellectual history smooths over one interesting wrinkle: Putnam had *anticipated* aspects of this liberal functionalism in *Representation and Reality* (1988, 75ff), but *nevertheless* rejected functionalism (*tout court*) for further reasons.

[15] McDowell, J. *Mind and World* (Cambridge, MA: HUP, 1994)

[16] See e.g. Putnam, 'Replies' (1992), 361

have *identical* qualia. (This represents a clear departure from his earlier view, that qualia-talk should be *rejected* on the grounds that we can make no sense of interpersonal comparisons between qualia; see pages 170–2.)

Now, a full answer to the *causal* question of why I have a particular perceptual belief will probably mention my qualia (195–6). However, Putnam holds that qualia cannot provide us with *justifications* or (good) *reasons* for holding our beliefs. Although Putnam doesn't quite put it this way, this is because qualia are completely in the head, whereas beliefs involve concepts that reach *outside* the head. To justify a perceptual belief, then, we seem to need something which is 'conceptualized' in the way that qualia aren't: we need something that exists *only* in creatures which possess concepts (in the externalist's sense); something which calls concepts into play (147, 151, 184); something, briefly, which *ain't in the head* (149, 225–6). Putnam calls these things *apperceptions*. They are something *like* conceptually-saturated perceptions.

At this stage, it is tempting to ask: *How do we 'get' from qualia to apperceptions?* But this question is dangerous. On the one hand, if the 'get' is just meant *temporally*, the question is just (for example) how a baby becomes (mentally) an adult. That's an interesting question, but *a priori* reasoning only has so much to say about it. On the other hand, if the 'get' is meant to involve a *rational reconstruction*, which has us treating qualia (alone) as *epistemologically* fundamental building blocks from which apperceptions are constructed, then Putnam explicitly *denies* that we 'get' from qualia to apperception (193–6).[17] And this denial is inevitable: since apperceptions ain't in the head, but qualia are, there is *no* function from qualia to apperceptions. (So, in particular, we do *not* 'logically construct' apperceptions from qualia.)

To be sure, there are many questions one might well raise about this view (cf. 210–2). For example:

- Are qualia 'immediately given' in experience, or do we 'infer their existence? (cf. Putnam's discussion of Reichenbach, pages 104 –5.)
- Are qualia really (features of) brain states? And either way, why not *simply* discuss brain states and ditch qualia-talk?
- What are the detailed relationships between qualia, apperceptions, and perceptual beliefs? And must the boundaries between them be completely sharp?

[17] Cf. Button, *Limits of Realism* (2013), 84–6

Had Putnam lived longer, he would doubtless have continued to explore these and other questions.[18] But the core thought, which I think that Putnam would have wanted to preserve no matter what happened with these questions – indeed, even if he ultimately rejected qualia-talk once again – is this. *Brain states are often caused by things outside the head; but the content of a belief or (ap)perception constitutively depends upon what is outside the head*. Indeed, this core thought is just a sympathetic extension of Putnam's own argument, that externalism is the key to any *real* challenge to Russell's theory of perception (180ff.)

# 3. The importance of reference

In the last section, I explained some consequences of Putnam's semantic externalism for his philosophy of mind. I now want to comment on its consequences for his approach to *truth*.

In this volume, Putnam emphasizes that the notion of *truth* depends upon the notion of reference (30–8). He takes this as the lesson from Field's 1972 paper on Tarski,[19] but here is a quick way to make the point. In the standard definition of a truth-predicate for a first-order language, we offer recursion clauses like this:

- $\ulcorner \neg p \urcorner$ is true iff $p$ is not true

However, our *base* clauses look something like this:

- '*F*' refers to all and only the white things.

and *reference* (or other cognate notions, like *extension* or *true-of*) is never *eliminated* from our definition.

Now, from 'The Meaning of "Meaning"' onwards, Putnam has held that reference is crucial to the meaning of a term, and hence to translation.[20] But here, Putnam uses this point to offer a deep challenge to deflationists. Deflationists need to accommodate the fact 'that *translating sentences presupposes knowing what their descriptive constituents refer to*' (38).

[18] Indeed, he was working on a book with Hilla Jacobson (and mentions this on page 158); Jacobson intends to finish and publish the book.
[19] Field, H. 'Tarski's Theory of Truth', *The Journal of Philosophy* **69**:13 (1972), 347–75.
[20] See also *Realism with a Human Face* (1990), 291–5

This is one of those golden moments I mentioned in the introduction. Many people have criticized deflationists, on the grounds that a theory of meaning *requires* a theory of truth. Putnam's point cuts deeper. Once you accept semantic externalism, the very notion of *meaning* depends upon the notion of *reference*. But this leaves the would-be deflationist having to take the plainly semantic concept of *reference* as a primitive, contrary to their aims to deflate (all) semantic notions.

To avoid taking reference as a primitive, the deflationist might try to argue that reference can be *reduced* to non-semantic notions. However, that would not help them much, for the *reduction* of reference will then provide us with a *reductive* (rather than a *deflationary*) theory of truth. And, in fact, Putnam holds that there cannot be a reduction of reference (or truth) to *non-intensional* notions (23–4, 71–2).

I believe that this attack on deflationism is new to this volume. But it draws on two ideas which Putnam has held for a long time, namely that *reference is central to truth* (a point clear since 'The Meaning of "Meaning"') and that *reference is non-reducible* (a point which Putnam had suggested in several places).[21]

*But wait* – you might say – *didn't Putnam at one time think that there were* no *determinate reference facts? Wasn't that the whole* point *of his model-theoretic arguments?* No! The model-theoretic arguments had the structure of a *reductio*: they sought to show that his *opponent* was forced to embrace the *absurd* claim that reference was radically, hopelessly, indeterminate. By contrast, Putnam's *own* position – his internal realism – was deliberately designed to *provide* the *determinacy* of reference.[22]

# 4. Internal realism

Let's turn, at last, to Putnam's rejection of 'internal realism'. Here are four points which are characteristic of his 'internal realist' writings:

---

[21] Putnam, *Reason, Truth, and History* (1981), 46–8; *Realism with a Human Face* (1990), 37–8, 80–95; *Renewing Philosophy* (Cambridge, MA: HUP, 1992), 35–59.

[22] Putnam, 'Realism and Reason', *Proceedings and Addresses of the American Philosophical Association*, **50**:6 (1977), 483–98, especially pages 488, 495; *Reason, Truth, and History* (1981), 53–4

(i) the model-theoretic argument(s) show that something is wrong with a particular philosophical position (held by his opponents).
(ii) the brain-in-vat argument shows something similar.
(iii) there is not just 'one single metaphysically privileged use of "thing" (or "object," or "entity")' (76), and this observation yields an effective attack on certain kinds of metaphysical inquiry.
(iv) truth is an idealization of warranted assertibility.

Putnam explicitly endorses (iii) in this volume. He also discusses (ii) in broadly approving terms (218–21). He scarcely mentions (i) here, and I would his relationship with the model-theoretic argument(s) was somewhat complicated. Nonetheless, in a fairly late paper, he suggested that both (i) and (ii) were effective criticisms of 'Cartesianism-cum-materialism' and, so far as I know, he never retracted this.[23]

Indeed, the philosopher after whom the 'hilary'[24] is named only *clearly* changed his mind on point (iv). He did so around 1990 and, in this volume, he offers two reasons why. One of them seems rather unfair on his past self; the other is excellent.

I shall start with the unfair argument. It runs as follows:

> … let us suppose, as seems reasonable, that whatever makes it rational to believe that *S* [also] makes it rational to believe that *S* would be justified were conditions good enough. If my understanding of the counterfactual '*S* would be justified if conditions were good enough' is exhausted by my capacity to tell to what degree it is justified to assert it, and that is always the same as the degree to which it is justified to assert *S* itself, why did I bother to mention the counterfactual at all?… On the other hand, if I repudiate the justificationist account of our understanding of counterfactuals, the Charybdis of the metaphysical realism I was trying to avoid sweeps me into its whirlpool. (124)

On encountering this argument, we – or an earlier timeslice of Putnam – should immediately ask: *why think that* my *understanding of the counterfactual is especially relevant?* Putnam always insisted that language was social. So surely

---

[23] Putnam, H. 'Das Modelltheoretische Argument und die Suche nach dem Realismus des Common sense', in Willaschek (ed.), *Realismus* (Paderbon: Ferdinand Schöningh Verlag, 2000), 125–42.
[24] 'hilary, n. (from *hilary term*) A very brief but significant period in the intellectual career of a distinguished philosopher. "Oh that's what I thought three or four hilaries ago."' From *The Philosophical Lexicon*, http://www.philosophicallexicon.com/

what matters for the 'justificationist account' is not *my* understanding, but (at least) our *society's* understanding, of the counterfactual.

In effect, Putnam replies to this question a few pages later, when he claims that 'while language is indeed social, *competence* is individual' (126). But where the *competence* in question concerns linguistic understanding, I think this is wrong; and not *just* wrong, but indeed wrong by Putnam's own lights. According to Putnam: Oscar, Toscar and an utterly isolated brain-in-a-vat differ in their linguistic competence, all having different thoughts (if any) when they entertain 'water is wet'. This illustrates the fact that linguistic competence can be *external*; and to make linguistic competence more explicitly *social*, we can switch from Putnam-style to Burge-style cases. So: Jane's (in)competence with the word 'arthritis', when she worries 'I have arthritis in my thigh', depends in part upon how the word 'arthritis' is used in her community.

Putnam's error here is not merely one to which he should have been alive, given the line he advanced in the 'Meaning of "Meaning"'. In a double-irony, it *repeats* an error he made in that very paper, when he attacked Frege by saying:

> … even if meanings are 'Platonic' entities rather than 'mental' entities on the Frege–Carnap view, 'grasping' those entities is presumably a psychological state (in the narrow sense).[25]

Putnam's claim here is roughly as follows: even if we grant that Fregean senses are entities *outside* the head, Fregeans must agree that our 'grasp' of them is compressed, as it were, through what *is* (merely) in the head. But, setting aside whether Fregean senses should really be thought of as *entities*, this is exactly what Fregeans should *reject*. To 'grasp' a Fregean sense should depend, not just upon what an individual does, but also upon the community and environment within which she does it. That would surely be the right moral for Fregeans who read 'The Meaning of "Meaning"'.

But let us return to Putnam's reasons for rejecting point (iv) of his internal realism. His second (better) reason is as follows. Suppose that our 'idealized best theory', whatever it is,

> … tells us that the cosmos, or at least the part of space-time that might contain intelligent life, is finite, and that there is no one place

[25] 'The Meaning of "Meaning"', (1975), 222

> within that finite region from which one can survey (or even receive causal signals from) the entire region… (132)

The same theory might also tell us that it is *extremely probable* that there *is* intelligent life elsewhere in the universe. However,

> … it is a conceptual truth… that *what is highly probable does not always happen*. Thus it is *internal to physical theory* itself…. That ['there are no intelligent extraterrestrials'] may be true. (133)

So: there may be a truth (e.g that there are no intelligent extraterrestrials) which we could *never* be warranted in asserting, and would in fact have excellent warrant to *reject*. That is why Putnam rejects internal realism, in the sense of rejecting (iv).

It is fitting that Putnam's (good) reason for abandoning internal realism is *internal* to our own theorizing. The whole point of *internal* realism was to allow Putnam to say, '*within* a theory', essentially what his opponents wanted to present in a metaphysically loaded way.[26] In this sense, the (good) reason for abandoning (iv) does not tell against the original *ambition* of internal realism. Quite the contrary: reflection on intelligent extraterrestrials should lead those *with* that ambition kind of ambition to abandon (iv). (Indeed, in the same way that considerations about actual or reconstructed scientific practice should lead us towards semantic externalism – see point (3) of §1 – so considerations about actual scientific practice should lead us to reject (iv).)

# 5. A liveable philosophy

But Putnam goes further. He tells us that it is not enough to try to say realist-sounding things '*within* a theory'. Rather:

> *… to preserve our commonsense realist convictions it is not enough to preserve some set of "realist" sentences; the interpretation you give those sentences, or, more broadly, your account of what understanding them consists in, is also important!* (112, his emphasis)

He explains this point as follows:

[26] 'Realism and Reason' (1977), 484.

> If[, with the positivist,] I view my whole language as just a device for predicting what experiences *I myself* will have – if even statements about my family, and about what will happen to them after I die, are no more than gears in a prediction machine, a machine whose whole purpose is to predict what I will experience here and now – then that view will violate the deepest intuitions we have about *what we are doing* when we utter sentences about others and about events after (and before) our own lives. (112; see also page 120)

This is another one of those golden moments. And, again, it is one with very deep roots. Twenty-five years earlier, Putnam had suggested that one might simply *shrug* at the thought that *tables* are logical constructions out of sense data, but insisted that we *must* deny that *other people* are such constructions. This is because the picture of others – real *others* – about whom we talk, and with whom we sympathise, '*is essential to our lives*'. In short – and, again, in accord with his semantic externalism – 'that truth genuinely depends upon what is *distant*, is part of a picture with enormous *human weight*'.[27],[28]

Tim Button

button@cantab.net

[27] Putnam, H. *Metaphilosophy* **22**:1 (1991), 1–13

[28] **Acknowledgements.** I wrote this paper during a period of research leave which was funded by a Philip Leverhulme Prize (awarded by the Leverhulme Trust, PLP–2014–140). Thanks to Maria Baghramian, Ned Block, Hilla Jacobson, Rob Trueman, and (most especially) Ori Beck.