

Empiricist Semantics and Indeterminacies of Reference

Igor Douven*

Department of Philosophy, Utrecht University

email: Igor.Douven@phil.ruu.nl

December 15, 1997

1 Introduction

In concert with his overall empiricist outlook, Quine urges ‘to approach semantical matters in the empirical spirit of natural science’ ([21, p. 8]). Among many other things this means that theories of reference, or interpretations—i.e., joint ascriptions of referents to the words a certain speaker or group of speakers uses—should be accepted or rejected on the basis of exactly the same criteria that govern theory choice in natural science. To be more precise, theories of reference are, on this view, exclusively to be judged on the grounds of so called operational and theoretical constraints, i.e., on the grounds of whether they are in accordance with the observational data¹—the operational constraints—and exhibit such virtues as consistency, explanatory force, and coherence with accepted theories in semantics as well as in other domains—the theoretical constraints.

In his [11] Putnam launches a vigorous attack on this kind of semantics. He claims that it entails referential indeterminacy of a quite radical sort. If theoretical and operational constraints were all we had to go on in semantics, then it might be that

... half of us (the females perhaps) use ‘cat’ to mean ‘cat*’, ‘mat’ to mean ‘mat*’, ‘look’ to mean ‘look*’, ‘tells’ to mean ‘tells*’, and so on [whereas] the other half (the males) use ‘cat’ to denote cats, ‘mat’ to denote mats, ‘look’ to denote looking, and so on... ([11, p. 36]),

without this situation ever coming to light, although ‘cat*’ and ‘mat*’ are defined so as to actually refer to cherries and trees respectively. For, according to Putnam, a world in which women use ‘cat’ to mean ‘cat*’, etc., and men use ‘cat’ to denote cats, etc., may be indistinguishable in every empirical respect from a world in which men and women use the same words to refer to the same

*I am grateful to Lieven Decock and Mark van Atten for valuable comments on the penultimate version of this paper; also to Michael Devitt, Laurie Paul, Jaap van Brakel and Roger Vergauwen for comments on a very early draft, and to Bas van Fraassen for helpful discussions on empiricist semantics.

¹In the case of semantics the relevant data are, briefly put, ‘pattern[s] of associations of sentences with one another and with non-verbal stimulation’ ([19, p. 27]).

(kinds of) things. But the indeterminacy pictured in the example is plainly absurd, Putnam thinks, and hence so is empiricist semantics.

Quine had already famously argued that, on empiricist premises, reference is indeterminate ([19, ch. 2], [20], [22]). However, although from his writings it isn't crystal-clear exactly to what extent he thinks reference to be indeterminate, his examples definitely suggest that the indeterminacy he envisages is of a rather moderate type. Typically Quine makes his claim concrete by saying such things as that it's indeterminate as to whether 'rabbit' refers to rabbits or rather to temporal rabbit-stages, or to cosmic rabbit-complements, or to singletons of rabbits (i.e., sets containing a rabbit as their only member), or to undetached rabbit-parts (cf. for such examples, beside the works just cited, [23, p. 33], [24, ch. 7]). To be sure, temporal rabbit-stages, cosmic rabbit-complements, etc., are no rabbits, but they are 'rabbit-related' in an intuitively evident sense. Already this moderate kind of indeterminacy may seem far from plausible. Yet Quine does not take it to reduce empiricist semantics to absurdity; in his view it is something we must learn to live with.

In this paper I will assume that both Quine and Putnam are right that empiricist semantics leads to *some* kind of indeterminacy of reference.² The main question to be addressed then is who is right about the *nature* of the entailed indeterminacy. What I will seek to establish is that Putnam's argument for the claim that the indeterminacy is of the radical variety of his cats*/mats*-example, and not just of the modest Quinean sort, fails, at least against an opponent like Quine, who advocates, along with an empiricist semantics, a naturalized epistemology. From that perspective it can be argued that, even though a world in which men, say, use 'cat' to denote cat-complements and women use 'cat' to denote cats may be indistinguishable from a world in which 'cat' is uniformly interpreted by all its inhabitants, there are good reasons to believe worlds in which interpretations diverge as radical as in Putnam's example can be distinguished from each other on purely empirical grounds. However, it will also be seen that the indeterminacy the empiricist is committed to is not quite as moderate as Quine's examples would make one believe, but is located somewhere in between the indeterminacy suggested by these Quinean examples and the indeterminacy of Putnam's cats*/mats*-example.

2 Putnam's permutation argument

The main premises of Putnam's argument against empiricist semantics are these:

- (P1) Theoretical and operational constraints can at most determine truth-conditions for whole sentences.
- (P2) Fixing the truth-conditions of all the sentences of a language still leaves sub-sentential reference radically indeterminate.
- (P3) Radical indeterminacy is absurd.

²Though it may well be arguable that neither of them has made a compelling case for indeterminacy of reference; see for some searching criticisms of Quine's argument for indeterminacy [6], [7], [25] and, e.g., [5], [9], [10] for criticisms of Putnam's argument to be discussed in this paper.

Putnam advances no support at all for the last premise; presumably he takes it to be obvious (cf. also [1, p. 313]). I will not touch upon the alleged absurdity of radical indeterminacy in this paper, and propose that we grant Putnam that at least the indeterminacy of the cats*/mats*-example is absurd. On (P1) I will focus in section 4, and I will try to show that it is false. Here I shall consider (P2).

Putnam's argument for (P2) is beyond doubt. In fact, it is proved in the appendix of [11]. Rather than go into the details of the proof, I will illustrate it with the help of a little example. This will also prove useful in making the distinction between the Quinean and the Putnamian variety of indeterminacy more precise.

Let w be a possible world with exactly eight individuals, $c_1, \dots, c_4, d_1, \dots, d_4$, and let the descriptive vocabulary of language \mathcal{L} consist of the predicates 'cat' (C), 'dog' (D), 'black' (B), and 'white' (W), and contain names for at least the individuals of w (' c_1 ', ' c_2 ', etc.). Under the standard interpretation of \mathcal{L} the predicates have the following extensions in w : $I(C) = \{c_1, \dots, c_4\}$, $I(D) = \{d_1, \dots, d_4\}$, $I(B) = \{c_3, c_4, d_3, d_4\}$, $I(W) = \{c_1, c_2, d_1, d_2\}$. Let f be a permutation function on the domain of w that maps c_1 onto d_3 , c_2 onto d_4 , c_3 onto d_1 and c_4 onto d_2 , and vice versa. Then consider the interpretation I' that interprets \mathcal{L} 's predicates in w as follows: $I'(C) = \{d_1, \dots, d_4\}$, $I'(D) = \{c_1, \dots, c_4\}$, $I'(B) = \{c_1, c_2, d_1, d_2\}$, $I'(W) = \{c_3, c_4, d_3, d_4\}$, and that renames the individuals in the obvious way (thus under I' ' c_1 ' denotes d_3 , etc.). As can easily be verified, every sentence of \mathcal{L} is made true by the model $\langle w, I \rangle$ exactly if it is made true by $\langle w, I' \rangle$ (try for instance 'Some dogs are white', 'There is at least one black cat', ' c_1 is a white cat').

The permutation procedure is quite general and can be carried out for all possible worlds. Assuming the possible worlds to be well-ordered in some fashion, we can define an interpretation I^* so that, for every possible world w_i , and some permutation function f_i on the domain of w_i , it assigns as extensions to \mathcal{L} 's predicates the f_i -images of what the predicates denote under I in w_i . Doing so yields two interpretations that may be radically at variance with each other, but that nevertheless make the same sentences true/false in every possible world. What this shows is that, if (P1) is true, so that I and I^* cannot be prised apart by theoretical and operational constraints (it's not empirically decidable whether an \mathcal{L} -speaker interprets \mathcal{L} according to I or to I^*), the indeterminacy entailed by empiricist semantics is indeed of a quite extreme kind.

3 Indeterminacies of reference

Suppose Putnam's permutation argument is correct. Then how much more radical than Quine seems to believe is the indeterminacy entailed by empiricist semantics? More generally: How do the rather mild indeterminacy of Quine's examples and Putnam's radical indeterminacy differ from each other? Quine's examples together with the illustration of Putnam's proof and the cats*/mats*-example given in the introduction already give *some* intuitive feel for the distinction between Putnamian and Quinean indeterminacy of reference. But for the discussion in the next section it cannot harm to have a more definite grasp of it. To that end I will attempt to define both kinds of indeterminacy in a more or less formal way. This requires some preliminary remarks however.

1. It is my intention to compare Quinean and Putnamian indeterminacy in the framework of Putnam's proof for (P2). That inevitably puts the discussion in an explicitly non-Quinean vernacular of possible worlds, possible objects and truth-conditions. So it is well conceivable that, already for that reason alone, Quine would not at all agree with the definition I will give of the 'Quinean' variety of indeterminacy. However, since my overall purpose is to inquire what kind of indeterminacy is entailed by empiricist semantics, rather than exegesis, I don't see this as all too problematical.

2. In Putnam's proof permutation functions play a role similar to that of the proxy functions in Quine's presentations of his argument for indeterminacy of reference. However, from Quine's writings on the matter one gets the impression that for him proxy functions may but need not be permutations. And indeed, of some of his examples it is hard to see how they could be permutations. E.g., how could the proxy function that maps each object of a certain domain onto its singleton—to give a prominent example of Quine's more recent work—be a permutation of that domain? Suppose the domain includes the cat Nana. Since Nana is not a set, it cannot be within the range of our function. But perhaps Quine has just not elaborated his example enough and is the function he had in mind really a more complicated one part of which is that *some* objects are mapped onto their singletons. However, as far as I can see whether a function is a permutation or not is not really essential for Quine's purposes. What does seem essential, and what he often emphasizes, is that the function be one-to-one. So instead of guessing what exactly Quine thinks the constraints to be imposed on proxy functions are, we will simply allow one-to-one functions in general, and not just permutations, as candidates of functions that can be used to obtain diverging interpretations in the manner illustrated in the previous section.

3. The Quinean examples of reinterpretations putatively giving rise to referential indeterminacy are, in Fodor's words, 'a mixed lot' ([7, p. 59]). I must admit that I cannot make equally good sense of each of them.³ What I will therefore do is select, mainly for illustrative purposes, some of the examples that I do find intelligible, and forget about the others, hoping that nothing essential turns on my selection. The proxy functions I will use are, apart from the function mentioned in the foregoing point, the function mapping each object in a given domain onto its cosmic complement (another one of Quine's present favorites) and the function mapping each object onto the set containing that object plus the empty set, which is, to my knowledge, not to be found in Quine's writings but is clearly Quinean in spirit.⁴

4. In Putnam's proof there is a separate function $f_i(x) = y$ for every possible

³One of the problematical aspects of especially some of the earlier and better-known examples of reinterpretations, like for instance, for 'rabbit', *temporal rabbit-stage* and *undetached rabbit-part*, is that there seems to be no way to understand exactly what they are supposed to come to that is consistent with Quine's requirement that proxy functions be one-to-one. Since any rabbit has many temporal stages as well as undetached parts the rabbits cannot be put in one-to-one correspondence to the temporal rabbit-stages nor to the undetached rabbit-parts. That in Quine's recent writings one searches in vain for examples like these may indicate that he himself has become suspicious of them.

⁴This last function was suggested to me by Lieven Decock.

world w_i . For convenience let us ‘glue’ such a collection of functions together to one function $f(i, x) = y$ that yields for each possible object x (i.e., for every object that exists in at least one possible world) and each possible world w_i in which x exists an image y such that for no $z \in w_i : f(i, z) = y$ unless $x = z$.⁵ What Putnam proves, formulated by means of this 2-argument function, is that if, for instance, for some interpretation I , $I(P) = \{a_1, a_2, \dots\}$ in w_i , $I(P) = \{a_4, a_{16}, \dots\}$ in w_j , and so on, for all possible worlds and all predicates, then there is an interpretation I^* such that $I^*(P) = \{f(i, a_1), f(i, a_2), \dots\}$, in w_i , $I^*(P) = \{f(j, a_4), f(j, a_{16}), \dots\}$ in w_j , and so on, and I and I^* make exactly the same sentences of the language true if $f(i, x)$ is suitably chosen.

The idea now is to distinguish Putnamian from Quinean indeterminacy in terms of the $f(i, x)$ -functions that relate diverging interpretations to each other (in the way f relates I and I' to each other in the illustration of Putnam’s proof); more precisely, the plan is to come to a formal distinction between, on the one hand, the $f(i, x)$ -functions that yield interpretations diverging to the extent *rabbit*, *cosmic rabbit-complement*, *temporal rabbit-stage*, etc., diverge as different interpretations of ‘rabbit’, and, on the other hand, those functions yielding interpretations more radically diverging.

There is a useful suggestion as to what the characteristic feature of the former might be in a remark in Quine’s [24, p. 72], made after having discussed some of his typical examples of proxy functions:

The reason such reinterpretations preserve truth values is that they preserve *sameness* of reference from mention to mention throughout discourse. Sameness of reference is what variables mark, and it is all that ontology contributes to science and truth

Sameness of reference throughout discourse would, in the context of our possible worlds presentation, of course have to become sameness of reference across possible worlds. But that indeed seems to me to characterize the functions yielding only divergencies of the Quinean sort; such functions are uniform, in the sense that for all i, j, x, y, z , if $f(i, x) = y$ and $f(j, x) = z$, then $y = z$. It must be noted, however, that in our setting the notion of sameness of reference is problematical for one of our examples of $f(i, x)$ -functions. Whereas the functions $f(i, x) = \{x\}$ and $f(i, x) = \{x, \emptyset\}$ —mapping x in every world in which it exists onto its singleton and onto the set containing x itself and the empty set respectively—literally preserve sameness of reference across worlds, the function $f(i, x) = w_i - x$, mapping each x in every world in which it exists onto its complement cannot do so: x ’s cosmic complement varies from one world to another. There is a way to overcome that problem, *viz.*, to understand the notion of uniformity as ‘preservation of *type* of reference across possible worlds’. However, although in the present context it seems quite intuitive what is meant, it may not at all be easy to make the notion of type involved precise. Here I will not even try to do so, also because I think it may not be worth the effort since the problem with the complement-function can be conceived of as an artifact of

⁵This can easily be done. Identifying the f_i -functions with sets of ordered pairs $\langle x, y \rangle$ such that $f_i(x) = y$, the procedure is to first turn every ordered pair $\langle x, y \rangle$ into an ordered triple $\langle x, y, i \rangle$, and then to take the union of all the sets of thus obtained triples. The resulting set obviously corresponds to a function $f(i, x) = y$.

our possible worlds presentation and does not arise for Quine.⁶

The passage of Quine's [24] just cited gives reason to believe uniformity is a necessary condition for Quinean divergence. But, to obtain a pair of interpretations diverging no more than the alternative interpretations in Quine's examples diverge, is it also sufficient to require that $f(i, x)$ be uniform? As far as I can see it is. That such a requirement rules out the divergence of the cats*/mats*-example and of the example used to illustrate Putnam's proof is immediately evident: in some worlds w_i in which c_1 exists we can set $f(i, c_1) = d_3$, but, since c_1 and d_3 do not necessarily co-exist, not in all; for similar reasons we cannot define $f(i, x)$ so that it uniformly swaps the cats and the cherries. But it seems to be more generally the case that the relation between the possible objects the $f(i, x)$ -function is defined on and their images under that function will have to be of the intuitively 'close' kind we encounter in Quine's examples: If for all x, y such that $f(i, x) = y$, y is to be the same (/of the same type) in every w_i in which x exists, then the $f(i, x)$ -function will have to be so specified that, for all x and y , x and y necessarily co-exist (/belong to necessarily co-existing types), and that just seems to guarantee that we do not get anything even near the Putnamian kind of divergence of interpretations by means of a uniform function.

We now appear to have two characteristics of Quinean proxy functions: they are uniform and they assign to each object another object necessarily co-existing with it. It is important to note that the two need not coincide. What is true is that, as we saw in the previous paragraph, uniform $f(i, x)$ -functions will assign necessarily co-existing objects. But the converse does not hold. There exist functions that assign to each object a necessarily co-existing object but that are not uniform. Here are some examples:

$$f(i, x) = \begin{cases} w_i - x & \text{if } |w_i| < \alpha \\ \{x\} & \text{if } \alpha \leq |w_i| < \beta \\ \{x, \emptyset\} & \text{otherwise,} \end{cases}$$

where $|w_i|$ denotes the cardinality of w_i and α, β are (finite or infinite) cardinal numbers;

$$f(i, x) = \{\alpha x\}_\alpha \Leftrightarrow |w_i| = \alpha,$$

where ' $\{\alpha\}$ ' stands for α ' $\}$'s;

and:

$$f(i, x) = \{x, \{\emptyset\}_n\},$$

where n equals the number of w_j 's such that $w_j \prec w_i$, given some well-ordering \prec on the class of possible worlds. These functions are increasingly 'non-uniform', the last one even varying from one world to another. If some function like these were used to obtain an interpretation I^* , on the basis of some other interpretation I , then, it seems to me, the resulting divergence would still not be as radical as the divergence of interpretations in Putnam's cats*/mats*-example—the denotation of 'rabbit' would under I^* still be very 'rabbity'—but

⁶One might *prima facie* think that it is a problem for Quine too, for it might seem that no cosmic complement of anything can stay the same for the dure of a discourse, however short (the universe, when we finish our talk, will not be as it was when we began). But since Quine conceives of objects as space-time regions, and thus can also identify their complements with space-time regions, this point of critique does not apply. Thanks to Lieven Decock for discussion on this point.

it would also not be of the modest Quinean type. In the next section this type of divergence will be seen to be of considerable interest to the matter of our concern.

These considerations give rise to the following definitions (for simplicity I restrict the definitions to unary predicates; the generalization to predicates with higher arity is straightforward):

- ▷ I and I^* are *Q-divergent interpretations* for \mathcal{L} if there is a function $f(i, x)$ such that:
 - (i) in any possible world w_i and for any descriptive predicate P of \mathcal{L} , I^* assigns to P the $f(i, x)$ -images of the objects I assigns to P ;
 - (ii) $f(i, x)$ is not trivial (i.e. it is not the case that for all $i, x : f(i, x) = x$);
 - (iii) $f(i, x)$ is uniform across possible worlds (in the sense explicated).
- ▷ I and I^* are *Q/P-divergent interpretations* for \mathcal{L} if there exists a function $f(i, x)$ such that clauses (i) and (ii), but not (iii), of the foregoing definition hold for it, *and*
 - (iv) for all i, j, x, y : if $f(i, x) = y$ and $x \in w_j$ then $y \in w_j$.
- ▷ I and I^* are *P-divergent interpretations* for \mathcal{L} if there is a function $f(i, x)$ such that clauses (i) and (ii), but not (iii) and (iv), of the foregoing definitions hold for it.

Before we can go on to define the forms of indeterminacy corresponding to these notions of divergence, we must make precise what it is for interpretations to be empirically indistinguishable:

- ▷ Interpretations I_1, \dots, I_n for \mathcal{L} are *empirically indistinguishable* if each one of them fully satisfies all theoretical and operational constraints imposed on interpretation.

Notice that empirical indistinguishability hasn't been generally defined for interpretations that satisfy all constraints equally well (which may mean: equally poorly). The reason for this is that interpretations which do equally, but not optimally, well are uninteresting from the current perspective, since they will be jointly dismissed for being unable to accommodate certain observations and/or for not conforming to some of the theoretical demands. A second point that should be stressed is that, since what operational and theoretical constraints are has only been loosely stated in this paper, and since it is especially controversial what is to be included among the theoretical constraints, the definition is still ambiguous. However, as far as I can see Putnam wants to be as liberal as possible on this point. That is to say, his claim is that whatever theoretical constraints the empiricist semanticist is to invoke, he is committed to a quite radical indeterminacy of reference.⁷ So we can just as well assume all parties

⁷This gesture appears more generous than it really is, given that virtually all philosophers nowadays agree that many of the theoretical constraints do not cut any methodological ice at all and serve at most rhetorical purposes (cf. for instance [4, p. 232], [8, p. 91]). Moreover, as a little demonstration in [11, pp. 37f] shows, simplicity, which many do regard as a theoretical constraint in good standing, does not rule out aberrant interpretations of the kind we encounter in the Quinean and Putnamian examples of indeterminacy of reference; what counts as the

involved to agree on this matter and pass over the intricate debate on which theoretical constraints are and which are not methodologically acceptable.

Here, then, are the definitions corresponding to the above defined types of divergence of interpretations:

- ▷ \mathcal{L} is *weakly referentially indeterminate* if there exist at least two empirically indistinguishable Q-divergent interpretations for \mathcal{L} .
- ▷ \mathcal{L} is *moderately referentially indeterminate* if there exist at least two empirically indistinguishable P/Q-divergent interpretations for it.
- ▷ \mathcal{L} is *radically referentially indeterminate* if there exist at least two empirically indistinguishable P-divergent interpretations for \mathcal{L} .

Two brief comments: (1) As defined here, these types of indeterminacy only entail genuine indeterminacy if empiricist semantics is correct; a proponent of some other kind of semantics may well be in the position to argue that, although some language is weakly/moderately/radically referentially indeterminate in the sense defined here, it is not really indeterminate given that empirically indistinguishable interpretations can be distinguished by other than empirical means. (2) It is reasonable to expect that the notion of indeterminacy of reference can be further unraveled than has been done here. Especially radical indeterminacy, as it stands, seems to me to be a sort of crude ‘catch-all’-category. E.g., in the cats*/mats*-example as well as in the illustration in section 2, we so to speak swap natural kinds. But Putnam’s proof does not require this. So we might try to distinguish various types of radical indeterminacy by somehow making formal the extent to which an $f(i, x)$ -function respects the ‘joints of nature’ of each of the possible worlds. However, although this seems to be a worthwhile project, I have at present no clue how to pursue it in a systematic fashion.

4 What the proof does *not* prove

Putnam’s argument for (P2), we saw, is unassailable. So if it can be argued that (P1) is correct, then the empiricist is indeed committed to radical indeterminacy of reference. In this section (P1) will be examined, and it will be shown false: though theoretical and operational constraints may not be strong enough to fix sub-sentential reference, they are not so weak that they can only fix the truth-conditions of sentences.

The problem the empiricist faces according to Putnam is that an empiricist semanticist not only lacks the resources to tell whether a speaker uses ‘cat’ to denote cats rather than cat-complements or cat-singletons, but that he is even unable to tell whether the speaker uses ‘cat’ to denote cats rather than cherries: The speaker’s behavior would be no different if he interpreted his words in the ‘standard’ way as it would be if he used another, be it Q-divergent be it P-divergent (in our terminology), interpretation. Now Putnam proved something about the preservation of truth-values of sentences in every possible world under

simpler of two competing interpretations depends very much on what the primitive terms of one’s ‘home language’ are. This point is reminiscent of Goodman’s that ‘bent’ predicates like ‘grue’ and ‘bleen’ cannot be dismissed on the ground that they implicitly refer to some particular time, for were these predicates primitive in our language instead of ‘green’ and ‘blue’ these latter predicates would have to be defined by reference to a particular time.

P-divergent reinterpretations of the descriptive vocabulary. But how does that result relate to behavior? How do we come from *preservation of truth-values in every possible world under P-divergent reinterpretation* to '*preservation*' of behavior under P-divergent reinterpretation?

That the inference in any case is not straightforward is already clear when we consider that, in general, our actions will not be based on complete information, known to be true, but on the information actually possessed and on the various degrees of plausibility we attach to each piece of information.⁸ The following may illustrate why this fact poses a potential problem for Putnam. Let I and I^* be P-divergent interpretations for a language \mathcal{L} . Call a speaker that interprets \mathcal{L} according to I (I^*) an I -speaker (I^* -speaker). Then suppose an I -speaker and an I^* -speaker both attach a high degree of plausibility to every sentence in a certain set T , each of which is—we assume—true (under I , and hence also under I^*). Suppose further that T informs the I -speaker about how to solve particular problems. If, then, the I -speaker bases his strategy for solving one of those problems on T (and as he has a high degree of confidence in it he will presumably do this), then, provided he manages to act in accordance with T , his action will be successful. This success is no miracle; it is perfectly explained by the availability of information which is both true and relevant to the problem. Now consider the situation for the I^* -speaker. By assumption T has a high degree of plausibility for the I^* -speaker; we also know that T , under I^* , is true. However, it seems that, under I^* , T may not be about problems and how to solve them at all, and even if it should be, it seems that these need not be the same or even of the same kind as those T under I is relevant to. May we not expect this to have an effect on the I^* -speaker's actions?

Putnam is aware of this problem. When considering the question whether evolutionary arguments have any bearing on the problem of reference and, in particular, whether radical indeterminacy of reference would have caused the

⁸Sometimes Putnam seems to interpret his permutation argument in a way which does make the inference straightforward. For instance, in his [18, p.264; emphasis added] he contends that 'the point of the [permutation] argument is that your behavior would be exactly the same, and the truth-values of all your sentences would be exactly the same, *on the hypothesis that you are mistaken about the reference of your terms* and each token of an arbitrary term (e.g. 'hydrogen atom') *really* [denotes] not the things in what you take to be its extension (the hydrogen atoms) but their [images in the permuted universe]'. In terms of the cats*/mats*-example cited in the introduction, this suggests not that women may use 'cat' to denote cats* but rather that although women use 'cat' to mean 'cat' this does not guarantee that they are referring to cats; it might be the case that, unbeknownst to them and to all of us, when a woman uses the word 'cat' she really refers to cats* (for instance because God has willed it so—cf. [12, p.x]). And it seems arguable that for your behavior what counts is what you believe your words to refer to, not what they actually refer to. But this way of putting the problem is not just at odds with how it is presented in [11], but also with the often stated claims that the permutation argument is just a more rigorous and more radical version of Quine's indeterminacy-argument and that it is meant to reduce Quine's semantic view to absurdity (cf. for instance [14], [16, p.280], [17, p.303], [18, p.251]): Quine's claim never was that it might be that when you use the word 'cat' it 'really' refers to cat-singletons whereas when I use the same word it 'really' refers to cat-complements but that, in case we should both happen to *believe* the word to refer to cats, this difference would never show. On Quine's view it makes no sense to assume that there is such a thing as '*really* referring to something'. And, of course, an absurdity derived from a number of Quinean premises *plus* one explicitly non-Quinean premise can hardly count against Quine's position. However, it seems that Putnam also employs the permutation argument in his battle against semantic externalism (see e.g. [14], [15]). That may well explain the above interpretation of his argument. To our discussion this interpretation is evidently irrelevant.

extinction of our kind, he contends that as long as ‘sufficiently many of our directive beliefs [beliefs of the form ‘If you do x , you will get y ’] are true under the non-standard interpretation... we will certainly be successful, and we will certainly *survive*... and have offspring...’ ([11, p. 40]). Now suppose that I am an I -speaker and my *Doppelgänger* on Twin Earth is an I^* -speaker. Of course, every true/false directive belief I have will be true/false under both I and I^* ; so my *Doppelgänger* and I will have the same number of true and false directive beliefs. But it is not hard to think up, for any directive belief, two divergent interpretations such that under one it clearly has ‘survival-value’ whereas under the other it has no such value at all.

What this shows, Putnam says, is that evolution must have produced

...in us representation systems whose sentences or sentence-analogues have certain *truth conditions* (and certain *action conditions*, or ‘language exit rules’). *But the truth-conditions for whole sentences were... shown not to determine the reference of sentence parts* (nor does adding the ‘language exit rules’ help, for these are preserved under [I^*]). It follows that it is simply a mistake to think that evolution determines a *unique* correspondence (or even a reasonably narrow range of correspondences) between referring expressions and external objects. ([11, pp. 40f])

But how can Putnam be so sure there is a mistake here? All he has shown is that *truth conditions* underdetermine reference, not that *truth conditions cum action conditions* underdetermine reference.⁹ Put differently, Putnam ([11, p. 40]) says that ‘[f]rom the point of view... of ‘evolution’, all that is necessary is that sufficiently many of [our] beliefs be true under *any* interpretation that connects those beliefs with the relevant *actions*’, but he fails to show that there *is* more than one interpretation which does connect beliefs with actions in the appropriate way.

Even if this were shown, however, Putnam’s ‘solution’ still would not be conclusive because it doesn’t account for considerations concerning plausibility-orderings. Suppose one of my directive beliefs says: ‘If I do a , I will get b ’. Suppose further that the action condition is fulfilled—I want b . Does it follow that I will do a now? Not necessarily; it might be the case that I have another directive belief which says: ‘If I do c , I will get b ’. What I will actually do will depend on many things; it may, e.g., depend on the respective degrees of confidence I have in those beliefs, on the estimation of which action is the more convenient or more efficient or safer one, etc., which will in turn heavily depend on what other sentences I accept and on the plausibility attached to each of them. And since Putnam’s permutation proof is only about preservation of truth-values under P-divergent reinterpretations, and is silent on preservation of plausibility-orderings under such reinterpretations, it doesn’t follow that, even if my *Doppelgänger* and I agree upon many sentences which have the same action conditions under both I and I^* (suppose evolution has taken care of that), we will act similarly in case a certain action condition is fulfilled. Sentences which we both accept and which have the same truth-conditions as well as the same action conditions, do not necessarily have the same ‘action-readiness conditions’;

⁹Surely it is correct that the truth conditions of the language exit rules will be preserved under P-divergent reinterpretations, but—for reasons explicated in the text—this is in itself no guarantee that our beliefs will be connected to the ‘right’ actions under such reinterpretations—and that’s what matters.

even if we have the same goal, we may opt for wholly different actions, because the sentences we agree on have a different plausibility-ordering.

It's not hard to see that this may make all the difference from the point of view of evolution. Suppose my *Doppelgänger* and I have for each goal necessary to attain in order for us to survive exactly two directive beliefs which tell us what to do, and that from every pair of alternatives only one is true. Then the circumstance that the other one is false does not matter as long as the right option is suggested by the factors just mentioned. We can imagine that the falsity of the other option will never show, just because we will never choose it—e.g., because, although we do think the belief to be true, we believe the action it tells us to perform to be much too risky or cumbersome or whatever. But, clearly, that the relevant information under *I* always suggests the right action, doesn't imply that the relevant information under *I** will also do so. Hence it cannot be true that all that evolution does is 'produce in us representation systems whose sentences have certain truth conditions and certain action conditions'. At least we should add '... and certain actions-readiness conditions', for these will in the end determine what we will actually do. In other words, it is not enough that both under *I* and under *I** certain sentences are related to certain actions in that these sentences suggest those actions as solutions to certain problems; it should be shown that, in most relevant cases, *I*-speakers and *I**-speakers will *do* the same thing.¹⁰

But we should ask ourselves whether there is any reason to doubt that P-divergent reinterpretation will leave plausibility-orderings unaffected. Is there any reason to doubt that my *Doppelgänger* and I will attach the same (or in any case almost the same) plausibility to every sentence of our language? If there is no such reason, then the above considerations are quite pointless of course. There is plenty such reason however. Putnam himself seems to be aware that doubt might arise concerning this point. Having defined 'cat*' to refer to cherries except in those worlds in which both some cat is on some mat and no cherry is on any tree, he remarks:

If... a person... looks at something and sees it is neither a cat nor a cherry, then [he] can tell it is *not* a *cat**; but if the thing is either a cat or a cherry then... the person needs to be informed of the truth-values of 'A cat is on a mat' and 'A cherry is on a tree' to decide if [he] is... seeing a *cat**, and these truth-values go beyond what [he] can learn by just examining the object presented to [him] for inspection. ([11, p.36])

¹⁰Things get even more complicated when we consider the interplay between belief and action in science and scientific methodology. For survival it may suffice to have a certain, 'fixed', body of effective directive beliefs. In science the search for better and better (faster, easier, safer, etc.) devices for attaining our goals is one of the main rationales for our inquiries. Suppose we have the directive belief 'If we do *a*, we will reach goal *b*'. A common strategy to find a more efficient device than *a* is to search for devices which are similar to *a* but which have some advantage(s) over *a* (cf. [2], [3]). When searching for such alternative devices we will have to rely on many background theories, telling us, e.g., which ones are similar to *a*, what exactly *a*'s weaknesses are (which are to be avoided), what its strong points are (which preferably should be retained), etcetera. So this time it wouldn't even be enough for Putnam to explain why 'sufficiently many of our directive beliefs are connected with the relevant actions'. What needs explanation is the success of the 'dialectical process' (the term is Boyd's), i.e., the constant 'positive' interaction between scientific theories on the one hand and methodological principles on the other in the sense that changes in our scientific beliefs are typically paralleled by changes in our methodological principles which in turn often lead to improvements of our theories, and so on.

It seems to need no comment that this difficulty makes it more than likely that a person's inspection of a hundred cats will not issue in even approximately the same increase or decrease of his confidence in, say, 'All cats are limber' or 'Some cats are cute' when he is an *I*-speaker as when he is an *I**-speaker. As Putnam pointed out, the latter, but not the former, will need additional information, beyond his observations, to know whether these observations inductively support certain statements; and nothing in the way Putnam has set up his proof for (P2) prevents that this additional information may in some cases be easy to obtain, in others hard, and in again others even impossible.

Am I not deliberately overlooking what Putnam has to say following the passage just cited?

[O]ne can reinterpret 'sees' (say, as sees*) so that the two sentences [1] John (or whoever) sees a cat; and [2] John sees* a cat*, will have the same truth-value in every possible world So whenever a person sees a cat, he *is* seeing* a cat*; the experience we typically have when we see a cat *is* the experience we typically have when we see* a cat*, and so on. ([11, p. 36])

Well, as a rejoinder to my point concerning plausibility-orderings this would entirely miss the mark. No doubt we can reinterpret 'seeing' as 'seeing*', together with our reinterpretation of all the other predicates, so that all sentences, including those in which 'seeing', 'to see', etc., occur, retain their truth-conditions. But we cannot thereby make seeing* as epistemologically significant as seeing (as perception, that is). To put the point another way, when an *I**-speaker is seeing* a cat*, he may still be unable to decide whether the experience he has is evidentially relevant to certain statements, like—most notably—those ascribing some property generally to cats. However ingeniously 'seeing*' is defined, we cannot make the *I**-speaker *see* a cat* when he is seeing a cat—that depends on what the world is like—nor can any reinterpretation make him believe that his experience is evidentially relevant to, e.g., 'Cats have funny little ears'.

Putnam may seem to have an easy way out. He can say that in the preceding paragraphs it is assumed both that perception is epistemically relevant to *I*- and *I**-speakers in the same way and to the same extent and that plausibility considerations play the same role in decision making for both *I*- and *I**-speakers. But that can be denied. It may be that an *I**-speaker, after having seen* a hundred cats* or so, has much more confidence in 'Cats are furry animals' than he had before, even though he hasn't seen a single cat* (i.e., cherry) in the meantime. The way he comes to hold beliefs is simply unrelated to perception, or only loosely related, or related in some hitherto undiscovered way. Alternatively, or in addition, it could be claimed that if an *I**-speaker thinks he has the choice between two actions, *a* and *b*, to reach a certain goal, and he deems it likely that *a* is the best action to perform given his goal, then that does not necessarily mean he will opt for *a*. Perhaps he bases his actions on something other than plausibility considerations.

A response along one of these lines may be effective, but only against a certain type of opponent. It will fail against a naturalist, who will point out that the hypotheses just considered are badly out of tune with what science has to tell us about how we come to our beliefs and how we act upon them. Not that, in the opinion of the naturalist, we must dogmatically adhere to what scientists come up with. The lesson of Quine's naturalism, as I understand it, is that no

part of our belief system is beyond doubt in principle, but also that putting in doubt the system as a whole is an infelicitous practice, springing from a wrong conception of what philosophy should accomplish. Thus, to put the point in the picturesque way employed earlier, from a naturalist perspective it might be unsettling if it could be shown that there exist (at least) two possible worlds such that (i) they cannot be distinguished from each other on an empirical basis, but (ii) they nevertheless differ in that in one men and women interpret their language in P-divergent ways whereas in the other language is interpreted uniformly or at most in Q-divergent ways, *and* (iii) both belong to the class of possible worlds in which all the scientific findings relevant to epistemology hold. Merely establishing the existence of a pair of worlds satisfying (i) and (ii), but not (iii), cannot convince the naturalist of Putnam's claim that empiricist semantics leads straightforward to radical indeterminacy.

Given his qualms about naturalism (cf. for instance [13]), we may expect Putnam to find this move objectionable. But then the discussion is no longer about reference but about epistemological naturalism. Notice that, at any rate, the appeal to naturalism does not beg the question against Putnam: science might have discovered that, say, ESP at least sometimes works. If it had, then a person's increase in confidence in sentences like 'Cats are furry', after having seen a certain number of cats, but without having seen any cherries, couldn't serve so simply to put aside interpretations on which he uses 'cat' to denote cats*. Clearly, it cannot be excluded that science will as yet discover ESP-like phenomena. As long as it hasn't, the above arguments are good enough to rebut Putnam's claim considered in this paper.

Exactly how strong are the arguments put forward in this section? Do they, if correct, rule out the possibility of radical indeterminacy, or do they perhaps even undermine indeterminacy theses in general? They appear to be pointless against weak indeterminacy: it seems that any observation evidentially relevant to 'All cats are furry' must also, and to the same extent, be evidentially relevant to 'All cat-singletons are singletons of something furry', or to 'All cosmic cat-complements are complements of something furry', or to any other reinterpretation of the sentence by means of some Quinean proxy function. Thus data concerning how a speaker amends his degrees of belief in response to certain happenings cannot help us to distinguish the hypothesis that the speaker uses 'cat' to denote cats from any competing hypothesis according to which he uses 'cat' to denote a specific sort of proxies of cats (cat-complements, cat-singletons, etc.).

It should also be clear that, although they do bear against the kind of indeterminacy of the cats*/mats*-example, the arguments presented are not decisive against radical indeterminacy in general. After all, it hasn't been shown that there may not be *some* pair of P-divergent interpretations that is empirically indistinguishable even on naturalist premises, i.e., even if we take into account all relevant scientific findings concerning our cognitive capacities. So the most we can conclude is (1) that the cats*/mats*-example elaborated in [11] does not offer such a pair of interpretations, and (2) that Putnam hasn't shown there exists any other pair of such interpretations.

What we do have are the means to generate, given some empirically adequate interpretation, another interpretation, that Q/P-diverges from it, and that, at least as far as the arguments given in this section go, is indistinguishable from it, even from a naturalist perspective. For suppose some speaker of our language

uses the word ‘cat’ to denote cat-complements if the world contains less than λ individuals, singletons of cats if the world consists of λ or more, but less than μ individuals, and sets having as members a cat plus the empty set if the world contains μ or more individuals (with λ and μ large infinite cardinals, say), and other predicates in a correspondingly diverging fashion. Then he faces the problem that he cannot know whether the sentence (+) ‘Cats are furry’ says that all cat-complements are complements of furry things, or rather that all singletons of cats are singletons of furry objects, or that all temporal cat-stages are temporal stages of something furry. Yet he is not thereby in the same predicament as the person in Putnam’s example, given that, for the relation of inductive support between (+) and particular observations, it doesn’t matter which of the foregoing possibilities states the correct interpretation of (+). And since, as was seen earlier, it neither matters whether he interprets ‘Cats are furry’ as saying that cats are furry or as saying that cat-complements are complements of furry things or . . . (etc.), seeing a hundred cats will have no other effect on the probability assigned to (+) if the speaker uses ‘cat’ to refer to cats than if he interprets ‘cat’ in the way just defined. The same holds for reinterpretations on the basis of any other non-uniform function that assigns to each object in every world in which it exists an object necessarily co-existing with it. Thus the considerations regarding (the non-preservation of) plausibility-orderings proffered against the radical indeterminacy of the cats*/mats*-example, cannot serve to make a case against moderate indeterminacy. So unless the empiricist has some other argument at his avail that tells against the possibility of moderate indeterminacy, he will have to acknowledge that the kind of indeterminacy his semantics entails is of a somewhat more radical nature than he believed.

Let us take stock. We have seen that Putnam’s argument against empiricist semantics fails. Operational and theoretical constraints can do more than (P1) alleges. It has also been pointed out that there is no easy way to fix the argument, at least not if it is directed against a naturalist opponent. On the other hand we have just seen that the kind of indeterminacy following from an empiricist approach to semantics may well be somewhat stronger than the weak indeterminacy of Quine’s examples. As intimated earlier, it may be possible to further analyze the notion of indeterminacy of reference and to come to a more fine-grained classification of indeterminacies than this paper could give. We might then be able to determine more precisely the indeterminacy to which the empiricist is committed. But, as the saying goes, that is topic for further research.

References

- [1] D. L. Anderson, ‘What is the Model-Theoretic Argument?’, *Journal of Philosophy* 90 (1993), pp. 311–322.
- [2] R. Boyd, ‘On the Current Status of Scientific Realism’, in: R. Boyd *et alia* (eds) *The Philosophy of Science*, Cambridge MA, 1991, pp. 195–222.
- [3] R. Boyd, ‘Lex Orandi est Lex Credendi’, in: P. Churchland and C. Hooker (eds) *Images of Science*, Chicago, 1985, pp. 3–34.

- [4] W. Callebaut (ed.), *Taking the Naturalistic Turn*, Chicago, 1993.
- [5] M. Devitt, ‘On Determining Reference’, in: A. Burri (ed.) *Language and Thought*, New York, in press.
- [6] G. Evans, ‘Identity and Predication’, *Journal of Philosophy* 72 (1975), pp. 343–363.
- [7] J. Fodor, *The Elm and the Expert*, Cambridge MA, 1994.
- [8] M. Friedman, ‘Philosophy and the Exact Sciences’, in: J. Earman (ed.) *Inference, Explanation and Other Frustrations*, Berkeley, 1992, pp. 84–98.
- [9] I. Hacking, *Representing and Intervening*, Cambridge, 1983.
- [10] D. Lewis, ‘Putnam’s Paradox’, *Australasian Journal of Philosophy* 52 (1984), pp. 221–236.
- [11] H. Putnam, *Reason, Truth, and History*, Cambridge, 1981.
- [12] H. Putnam, ‘Introduction’, in his *Realism and Reason*, Cambridge, 1983.
- [13] H. Putnam, ‘Why Reason can’t be Naturalized’, in his *Realism and Reason*, Cambridge, 1983, pp. 229–247.
- [14] H. Putnam, ‘Model Theory and the ‘Factuality’ of Semantics’, in: A. George (ed.) *Reflections on Chomsky*, Oxford, 1989, pages 211–232.
- [15] H. Putnam, ‘Is the Causal Structure of the Physical Itself Something Physical?’, in his *Realism with a Human Face*, Cambridge MA, 1990, pp. 80–95.
- [16] H. Putnam, ‘Realism without Absolutes’, in his *Words and Life*, Cambridge MA, 1994, pp. 279–294.
- [17] H. Putnam, ‘The Question of Realism’, in his *Words and Life*, Cambridge MA, 1994, pp. 295–312.
- [18] H. Putnam, ‘Comments and Replies’, in: P. Clark and B. Hale (eds) *Reading Putnam*, Oxford, 1994, pp. 242–295.
- [19] W. V. Quine, *Word and Object*, Cambridge MA, 1960.
- [20] W. V. Quine, *Ontological Relativity and Other Essays*, New York, 1969.
- [21] W. V. Quine, ‘Philosophical Progress in Language Theory’, in: H. Kiefer and M. Munitz (eds) *Language, Belief, and Metaphysics*, Albany, 1970, pp. 3–18.
- [22] W. V. Quine, *Theories and Things*, Cambridge MA, 1981.
- [23] W. V. Quine, *Pursuit of Truth*, Cambridge MA, 1992.
- [24] W. V. Quine, *From Stimulus to Science*, Cambridge MA, 1995.
- [25] B. Stroud, ‘Quine on Exile and Acquiescence’, in: P. Leonardi and M. Santambrogio (eds) *On Quine*, Cambridge, 1995, pp. 37–52.