

ON DETERMINING REFERENCE
Michael Devitt
In Sprache und Denken / Language and Thought
Alex Burri, ed.
New York: Walter de Gruyter (1997): 112-21

I am concerned with the general problem of determining reference and with Hilary Putnam's ingenious and much debated model-theoretic argument in particular. The paper is an application of the views on semantics I urge in Coming to Our Senses (1996) and "The Methodology of Naturalistic Semantics" (1994).¹

Reference is a relation between words in language (or thought) and the world; we may say that it relates a singular term to an object, a general term to a class of objects, and so on. To specify a word-world relation as reference is to specify an interpretation of the language (or thought) because the truth conditions, hence meanings (or contents), of sentences can be explained in terms of reference. If the specified word-world relation really is reference, then the interpretation is "standard" (and "intended"); if not, the interpretation is "nonstandard".

A problem arises about the determinacy of reference. We might pose this problem - rather misleadingly, as we shall see - in the following way: Why identify reference with one word-world relation - that of the standard interpretation - rather than with any of the others? Putnam suggests an answer: Reference is the one and only relation relative to which an "epistemically ideal" theory would come out true. According to Putnam, this is the only answer that the "metaphysical realist" is entitled to. This represents a problem for that realist because the model-theoretic argument shows that there are many such relations not just one and so reference is indeterminate. Thus, consider one (not very interesting) sentence in the ideal theory: 'Nana is a cat'. The relation that we would ordinarily identify as reference relates 'Nana' to my late cat, Nana, and 'cat' to all cats. Relative to that relation, the sentence comes out true. Putnam's argument is that it comes out true relative to many other relations including, for example, one that relates 'Nana' to my neighbor's dog, Fido, and 'cat' to all dogs. So the metaphysical

¹The paper was also stimulated by an interesting paper by Igor Douven's (1994), on which I was the commentator. On the surface, Douven argues that we can determine reference by considering the explanation of success. I think that this is a mistake (1991: 97-101, 113-17). However, another idea sometimes seems to underlie Douven's discussion: that of determining reference by considering the explanation of actions. Part of my own solution in what follows is along those lines. I am grateful to Georges Rey for comments on a draft of this paper.

realist's answer fails to determine reference (1978: 123-7; 1981: 29-48; 1983: ix-xii, 1-25; 1989: 213-21).

Some preliminary remarks are called for about the "epistemically ideal" theory and "metaphysical realism". An epistemically ideal theory is one that meets all operational and theoretical constraints. It is the theory that we would accept at the ideal limit of scientific inquiry. Putnam emphasizes that the precise details of these constraints are unimportant to his argument. What matters is that the constraints are epistemic (1989: 229n).

Putnam gives several versions of the doctrine "metaphysical realism". Here is an early one: The doctrine requires that

there be a determinate relation between terms in L and pieces (or sets of pieces) of THE WORLD . . . THE WORLD is independent of any particular representation we have of it . . . truth is . . . radically non-epistemic. (1978: 125)

In all versions, metaphysical realism has both a metaphysical component and a semantic component. The metaphysical component is a commitment to what I call "Realism"; roughly, physical entities like stones, trees, cats, electrons, muons, and curved space-time objectively exist independently of the mental. The semantic component is a commitment to a "correspondence theory of truth", one explained in terms of referential relations rather than in epistemic or deflationary terms. As a result of his argument, Putnam finds metaphysical realism "incoherent" (p. 124) and abandons it in favor of "internal realism" (p. 130), a form of antiRealism.

Putnam's discussion of metaphysical realism exemplifies two typical features of the current realism debate that I have argued against in Realism and Truth (1991): conflating the metaphysical and the semantic; and using semantics to settle metaphysical issues. On the basis of this earlier argument I shall take it that, contrary to what Putnam suggests, the model-theoretic argument has no direct consequences at all for Realism, the metaphysical component of metaphysical realism. What the argument does threaten is correspondence truth, the semantic component.²

Return to Putnam's argument. He anticipates a response that places a further constraint on reference. The response appeals to a causal theory of reference: 'Cat' refers to cats not dogs because it stands in the appropriate causal relation to cats not dogs. Putnam dismisses the response on the ground that it just adds more theory.

"How 'causes' can uniquely refer is as much of a puzzle as how 'cat' can on the metaphysical realist picture" (1978: 126); it is not "glued

²Lewis 1984 has a nice discussion of some other problems in interpreting Putnam's argument.

to one definite relation with metaphysical glue" (1983: 18). To suppose otherwise is to have "a magical theory of reference" (1981: 47).

In a critical study of Putnam's Meaning and the Moral Sciences, I argued against this "just-more-theory" dismissal, claiming that it begged the question (1983: 297-9).³ I claimed that the causal explanation of the reference of 'cat' is not undermined by raising an analogous question about the reference of 'cause'. And when that analogous question is raised, we give an analogous answer. And so on until the cows come home. All terms are glued to their referents by causal relations not by "metaphysical glue". Putnam himself nicely captures the underlying idea of this realist reaction:

A realist does not claim that reference is fixed by the conceptual connection (i.e. the connection in our theory) between the terms 'reference', 'causation', 'sense impression', etc.; the realist claims that reference is fixed by causation itself. (1983: xi)

This realist reaction has been urged independently by Clark Glymour (1982). It has been endorsed by David Lewis (1984) and James Van Cleve (1992), and even by David Leech Anderson (1993: 315) in the course of defending Putnam's model-theoretic argument.⁴ But it has been firmly rejected by Putnam, who accuses the realist of begging the question (1983: ix-xii; 295-6)! This leads Ernest Sosa to describe the situation as "an impasse" (1993: 606).

I don't think that Sosa is right because Putnam does not really address the realist reaction; he simply repeats his just-more-theory dismissal (as Van Cleve points out: 349). Nevertheless, I no longer think that the causal-theory response is entirely adequate.

To see why we need to get clearer about the determinacy problem. I shall start by "leaning on" our initial way of posing the problem: "Why identify reference with one word-world relation - that of the standard interpretation - rather than with any of the others?" This asks why the relation between 'Nana' and Nana, 'cat' and cats, and... [continue on through the paradigms of reference] rather than that between 'Nana' and Fido, 'cat' and dogs, and... [continue on in an apparently arbitrary manner] is to be identified as reference. This

³The argument is expanded in my 1991: 226-9.

⁴Which is not to say that these philosophers have endorsed a causal theory of reference. Anderson argues, rather strangely, that Putnam's real model-theoretic argument is not the argument that critics have taken it to be. He locates the real argument in Putnam's doubts about the possibility of a substantive and plausible causal theory of reference, doubts arising out of problems with causality. I discuss this apparently different argument briefly later.

way of posing the problem is misleading because it invites a naive response: "The question is silly and doesn't need a substantive answer. It is no more sensible than asking the analogous question about the relation **being the father of**: Why is the relation between Bill Clinton and Chelsie Clinton, Prince Philip and Prince Charles, and... [continue on through the paradigms of fathering] rather than the relation between Bill Clinton and Newt Gingrich, Prince Philip and Baroness Thatcher, and... [continue on in an apparently arbitrary manner] to be identified as **being the father of**? Or the analogous question about the property **cathood**: Why is the property of Nana, Jemima, and so on [continue on through the paradigm cats] rather than the property of Fido, Bill Clinton, and... [continue on in an apparently arbitrary manner] to be identified as **cathood**? We need say no more in answer to these questions than that the specified relation just is **being the father of** and the specified property just is **cathood**. Similarly, we need say no more in answer to the question about reference than that the specified relation just is reference. That relation and not any others simple is the one we call "reference" and that's that."

Despite any encouragement this response may get from misleading ways of posing the determinacy problem, the response seems naive because it largely misses the point of the problem. Identifying one word-relation as reference is indeed easy but we want to know how that relation differs from the others; for example, from one we might call "reference*", exemplified by the relation between 'Nana' and Fido, 'cat' and dogs, and so on? What is the difference between the standard interpretation and this nonstandard one? Similarly, the problems lying behind our analogous questions demand answers to: What is the difference between **being the father of** and a relation we might call "**being the father of***", exemplified by the relation between Bill Clinton and Newt Gingrich, Prince Philip and Baroness Thatcher, and so on? What is the difference between **cathood** and the property we might call "**cathood***", exemplified by Fido, Bill Clinton, and so on? To address these questions we need to say something about the natures of the properties and relations in question.⁵ And that is what the causal theory seems to do in response to the question about reference: Reference, unlike reference*, is determined by a certain causal relation, the one we are gesturing at with "appropriate". Clearly we need to do more than gesture and, as we shall see, it is hard to do all that is necessary. Still, if we had, we would seem to have addressed the point of the determinacy problem.

⁵Cf. my two-step methodology for explaining the nature of **being an F**: the preliminary first step identifies some apparently uncontroversial examples of Fs and non-Fs; the second examines the examples to see what is common and peculiar to Fs (1994: 561-5; 1996: 72-6). The naive response addresses only the first step.

And so we would have, but not adequately. The causal theory claims that 'cat' refers to cats because it stands in an appropriate causal relation to cats. But it stands in causal relations to many other objects, including all dogs; causality is ubiquitous. So if we can come up with a causal theory of reference according to which 'cat' refers to cats, we can surely come up with a causal - or perhaps we should call it a "causal*" - theory of reference* according to which 'cat' refers* to dogs. The former theory adverts to a causal relation that is appropriate for reference, the latter to a causal* relation appropriate for reference*. It is obvious, of course, that the two relations are different. So the causal-theory response is a clear advance over the naive one (pace Putnam). But until we have said something about the significance of that difference - about why reference is special - we have not gone nearly far enough beyond the naive response in addressing the point of the determinacy problem.

Perhaps awareness of this inadequacy underlies the view of Ernest Lepore and Barry Loewer that my causal-theory response "missed the force of Putnam's argument" (1988: 464). But it is worth noting that the unsolved problem is not best posed as they do: "Why identify reference with cause rather than cause*?" (p. 465). That way of posing the problem invites an extension of the naive response: "Reference simply is the relation exemplified by the relation between 'Nana' and Nana, 'cat' and cats, and so on, and that relation simply is determined causally not causally*." The problem is not that of justifying our practice of explaining reference causally not causally* but that of justifying the treatment of reference, explained causally, differently from reference*, explained causally*.

How would we say something about the significance of the difference between **being a father of** and **being a father of***, and between **cathood** and **cathood***? We would describe the causal roles that make talking of **being a father of** and **cathood** but not **being a father*** and **cathood*** serve our explanatory purposes; for example, we might point out that x's being the father of y explains certain similarities between them; and z's being a cat explains why it purrs. Similarly, to say something about the significance of the difference between reference and reference*, we need to describe the special place of reference in the causal nexus.

Why are we interested in reference in the first place? We are interested because we are interested in meaning, and reference is central to meaning. Why are we interested in meaning? What purposes does it serve to ascribe meaningful (contentful) thoughts and utterances to people? One important part of the answer is: to explain intentional behaviors or actions.⁶ Actions causally relate people

⁶I draw on my 1994: 548-54; 1996: 53-62. Another important part of the answer is: to use the thoughts and utterances of others as guides to reality.

to the world. The properties and word-world relations that we ascribe to thoughts have to be appropriate given the effect of thoughts. Suppose that a person is in the presence of a cat, Nana. She has a thought which, on the standard interpretation, refers to cats and means, **I LOVE THAT CAT**; and, on our nonstandard interpretation, refers* to dogs and means* whatever. In virtue of what is the standard interpretation right? It is right because the property it ascribes is part of the explanation of what the person in fact did: She stroked Nana.⁷ She might have had a different thought which, on the standard interpretation, would have meant **I HATE THAT DOG**. This would have led her to do something else: She would have chased Fido perhaps. In contrast, the nonstandard interpretation, which relates her actual thought to dogs and her possible one to whatever, has no role in the explanation of what she did or would do. That's why the nonstandard interpretation is not right. Reference is special because of its place in the explanation of actions.

In the course of a criticism of the idea that evolution determines reference, Putnam makes a paranthetical remark about "language exit rules" that seems to deny this special relation between reference and actions (1981: 41; see also 1983: x). He gives no argument for the denial. If he were pressed on it, I suspect that he would resort again to the just-more-theory dismissal. But the dismissal would be no better here than before. The claim is not that reference is distinguished from reference* by the conceptual connection between 'reference' 'causation' and the descriptions of actions but by the worldly connection between reference, causation, and actions.

The claim can be amplified in terms of what Putnam's calls "directive beliefs" (1981: 39). These beliefs are of the form 'If I do x, I will get y', and so are intimately connected with actions. Suppose that a person has such a belief as part of her ideal theory, and also has the related desire 'I want y'. As a result, she acts. As a result of that, she will get something (for her theory is ideal). What she gets, as a matter of fact, must be what her tokens of 'y' refer to. And so it is on the standard interpretation. But on our nonstandard interpretation, 'y' refers* to something else which she may not get.

The key idea of this response to the determinacy problem is that a term has certain effects partly in virtue of having a certain referent and not in virtue of having a certain referent*. But we also have to consider the causes of a term. Thoughts containing the term not only have effects they have causes. Consider the person's thought that refers to cats and means **I LOVE THAT CAT** on the standard interpretation. That the thought means this is part of the

⁷I reject the view that the explanation of behavior does not require the ascription of thoughts with "wide" referential contents but only that of states with "narrow" contents (1996: 245-312).

explanation of its formation by the causal action of Nana on the person via perception. If the person had had the thought that meant **I HATE THAT DOG**, that meaning would have been part of the explanation of its formation by something different, perhaps by the causal action of Fido. In contrast, the nonstandard interpretation which relates her actual thought to dogs and her possible one to whatever, has no role in the explanation of the formation of those thoughts. So we have another key idea for determining reference: Part of the explanation of a term's having certain causes is that it has a certain referent but part of the explanation is not that it has a certain referent*.

Let us sum up. Our problem is the determinacy of reference. A misleading way of posing the problem is: Why identify reference with one word-world relation rather than with any of the others? This invites the response that one of these relations is, as a matter of fact, the one we call "reference" and that's that. This response is naive because the problem of determining reference is not primarily to identify one word-world as reference but to say, first, how that relation differs from other word-world relations; and, second, what is significant about that difference, what is special about reference.

In response to the first part of this problem, I appeal to a causal theory of reference: A certain causal relation determines reference and not any of the other word-world relations. In response to the second part of the problem, I claim that reference is special in that it is in virtue of referential relations that thoughts have certain worldly causes and effects.

Putnam has two further objections to the causal-theory response to the first part of the problem. The more recent of these (in 1989) trades on an earlier discussion (in 1981) which he describes as follows:

I defined properties I called `cat*' and `mat*' in such a way that (i) in the actual world the things which are cats* are cherries and the things which are mats* are trees; and (ii) in every possible world, the two sentences `A cat is on the mat' and `a cat* is on a mat*' have the same truth value. (1989: 217-8)

On the strength of this Putnam claims that the two sentences describe "the same event". So "there is no way in which the event [of someone's using a token of a term] can be causally connected to a cat's being on a mat without being identically causally connected to a cat*'s being on a mat*" (1989: 219).

This is very hard to accept, but suppose that it were so. It would show that the event of some cat or other being on some mat or other had the same causal relations as the event of some cat* or other being on some mat* or other. It would not show that cats have the same causal relations as cats*. Yet that is what Putnam needs to show to rebut the causal-theory response for, according to that

response, a token of `cat' refers to cats not cats* because it has a certain causal relation to cats not cats*. And what he has not shown, but needs to, is implausible if not downright bizarre as we can see by noting that, in the actual world (and many others), cats* are cherries.

Putnam's other objection to the causal-theory response is deeper: The metaphysical realist's appeal to causality is not physicalistically respectable (1983: 205-28; 1984). In picking one aspect of the "total cause" as "the cause" of a token, we are distinguishing that factor as a "bringer-about" from others as "background conditions". But, Putnam argues, this distinction is not built into the world, it is "interest relative" and imposed by our minds.

The causal-theorist need not assume that the world singles out one aspect of the total cause as the cause. She can go along with Putnam's view that this is our work not the world's. Still, she does assume that there is some basis in nature for distinguishing the causal relation that `Nana' has to Nana from the one it has to Fido and so on, and that `cat' has to cats from the one it has to dogs and so on. I still find this assumption plausible despite Putnam's argument.

But, I confess sadly, I have nothing interesting to say about the nature of causality in support of the assumption. So, suppose that Putnam were right. What would be the consequences? We would be forced to choose between reference and the physicalism Putnam describes.⁸ And the choice is not simply between reference and this physicalism. Every property or relation that is explained causally would be open to the same physicalist criticism and would fall with reference. So, the choice is between the central properties and relations of all the "human" sciences - psychology, economics, history, and so on - and this physicalism. In the face of this, many may choose to abandon such a dire physicalism, seeking a more moderate unifying doctrine for science.

This is one left-over problem for the causal-theory response. Here is another. It is agreed that we need more than gestures toward a causal theory of reference to solve the determinacy problem, we need the details. It has proved very hard to supply them.⁹ And a major difficulty is a determinacy problem: specifying a causal relation between a word and the world that relates the word only to

⁸Not, mind, to choose between reference and the metaphysical doctrine I call "Realism". Physicalism, unlike Realism, is a reductive doctrine. Thus, dualism is inconsistent with physicalism but entails Realism; 1991: 24-5.

⁹Nevertheless, Anderson is surely too harsh in claiming that "for too long philosophers have gotten away with a wink and a nod parading as a substantive theory of reference" (1993: 321).

its referent. Thus, historical-causal theories face "the qua-problem" (Devitt and Sterelny 1987: 63-5, 72-5). In virtue of what is 'Nana' grounded in my late cat rather than a spatial or temporal part of her? In virtue of what is 'cat' grounded in cats rather than felines, mammals, and many other things. I doubt that historical-causal theories alone have the resources to solve this problem.

This brief discussion of the qua-problem is reminiscent of Quine's long discussion of indeterminacy (1960). But there is an important difference. I think that it is hard to find a physicalistically respectable theory that will determine reference, but I keep looking. Quine thinks that there is no point in looking because he already has an argument that reference is indeterminate.

The argument starts from the assumption that the only objective reality to be captured by ascriptions of meanings are verbal dispositions; for example, the disposition to assent to 'Gavagai' when provided with the stimulus caused by the presence of a rabbit.

He argues, I think convincingly, that the constraints of this meager reality are insufficient to determine reference. But, as many have pointed out, the starting assumption, which is both behaviorist and verificationist, badly needs an argument that it never gets. Why does physicalistic respectability demand the antimentalism of behaviorism? Why is the robust mentalism of functionalism not kosher?

Why suppose that meaning has anything to do with verification? Quine clearly thinks that his starting assumption is obvious. But to many others it seems like just another dogma of empiricism. And it is a dogma that sits oddly with Quine's naturalism.¹⁰

In conclusion, there are two parts to the problem of determining reference: explaining reference in a way that distinguishes it from other word-world relations; and demonstrating that reference so distinguished is special in some way. Causal theories promise a solution to the first part of the problem although they have not yet fulfilled that promise. I claim to have solved the second part by giving reference a certain place in the causal nexus. There is no good reason to accept that the constraints on reference are as meager as those of Putnam's model-theoretic argument or of Quine's

¹⁰In support of this paragraph, see, for example, Friedman 1975 and Antony 1987. Quine's antimentalism leads him to see semantic issues as arising primarily in the context of translation. So the semantic task is to explain the construction of a translation manual. My mentalism leads me to see semantic issues as arising primarily in the context of psychology. So the semantic task is to explain the nature of certain properties of thoughts, and of the utterances that express them, properties that we ascribe in order to explain behavior and to learn about a reality that is the object of those thoughts (1994, 1996).

indeterminacy argument. Hence there is no good reason to accept the conclusion of these arguments that reference is indeterminate.

University of Maryland, College Park

REFERENCES

- Anderson, David Leech 1993: What Is the Model-Theoretic Argument? Journal of Philosophy 90: 311-22.
- Antony, Louise 1987: Naturalized Epistemology and the Study of Language. In Abner Shimony and Debra Nails, eds., Naturalistic Epistemology. Dordrecht: D. Reidel: 235-57.
- Devitt, Michael 1983: Realism and the Renegade Putnam: A Critical Study of Putnam 1978. Nous 17: 291-301.
- 1991: Realism and Truth. Cambridge: Basil Blackwell (1st edn. 1984).
- 1994: The Methodology of Naturalistic Semantics. Journal of Philosophy 91: 545-72.
- 1996: Coming to Our Senses: A Naturalistic Program for Semantic Localism. New York: Cambridge University Press.
- and Kim Sterelny 1987: Language and Reality: An Introduction to the Philosophy of Language. Cambridge MA: MIT Press.
- Douven, Igor 1994: Could Reference Be Indeterminate? Paper delivered at the Eastern Division Convention of the American Philosophical Association in Boston.
- Friedman, Michael 1975: Physicalism and the Indeterminacy of Translation. Nous 9: 353-74
- Glymour, Clark 1982: Conceptual Scheming or Confessions of a Metaphysical Realist. Synthese 51: 169-80.
- Lepore, Ernest, and Barry Loewer 1988: A Putnam's Progress. In Peter A. French, Theodore E. Uehling, Jr. and Howard K. Wettstein, eds., Midwest Studies in Philosophy Volume XII: Realism and Antirealism. Minneapolis: University of Minnesota Press: 459-73.
- Lewis, David K. 1984: Putnam's Paradox. Australasian Journal of Philosophy 62: 221-36.
- Putnam, Hilary 1978: Meaning and the Moral Sciences. Boston: Routledge & Kegan Paul.

- 1981: Reason, Truth and History. Cambridge: Cambridge University Press.
- 1983: Realism and Reason: Philosophical Papers, Vol. 3. New York: Cambridge University Press.
- 1984: Is the Causal Structure of the Physical Itself Something Physical? In Peter A. French, Theodore E. Uehling, Jr. and Howard K. Wettstein, eds., Midwest Studies in Philosophy Volume IX: Causation and Causal Theories. Minneapolis: University of Minnesota Press: 3-16.
- 1989: Model Theory and the 'Factuality' of Semantics. In Alexander George, ed., Reflections on Chomsky. Cambridge: Basil Blackwell: 213-32.
- Quine, W. V. 1960: Word and Object. Cambridge, MA: MIT Press.
- Sosa, Ernest 1993: Putnam's Pragmatic Realism. Journal of Philosophy 90: 605-26.
- Van Cleve, James 1992: Semantic Supervenience and Referential Indeterminacy. Journal of Philosophy 89: 344-61.