

HENRY JACKMAN

## RADICAL INTERPRETATION AND THE PERMUTATION PRINCIPLE

**ABSTRACT.** Davidson has claimed that to conclude that reference is inscrutable, one must assume that "If some theory of truth . . . is satisfactory in the light of all relevant evidence . . . then any theory that is generated from the first theory by a permutation will also be satisfactory in the light of all relevant evidence." However, given that theories of truth are not directly read off the world, but rather serve as parts of larger theories of behavior, this assumption is far from self-evident. A proper understanding of the role truth theories play in theories of interpretation makes the inscrutability of reference much less wide-spread than Davidson suggests, and, as a result, the radical interpretation methodology is much less likely to saddle its defenders with counterintuitive cases of indeterminacy than is commonly supposed.

Davidson has argued<sup>1</sup> that there is no fact of the matter as to what any speaker's words refer to because, even holding truth conditions fixed, a radical interpreter will always be able to come up with many equally good interpretations of the interpretee's language. This conclusion, which Davidson (following Quine) refers to as the "inscrutability of reference", has caused many to reject the radical interpretation methodology as fundamentally flawed.<sup>2</sup> Nevertheless, it isn't clear that such widespread inscrutability is a necessary consequence of the radical interpretation methodology. In particular, Davidson claims that the following assumption (which will hereafter be referred to as the "Permutation Principle") is "clearly needed if we are to conclude to the inscrutability of reference":

If some theory of truth (or translation or interpretation) is satisfactory in the light of all relevant evidence (actual or potential) then any theory that is generated from the first theory by a permutation will also be satisfactory in the light of all relevant evidence.<sup>3</sup>

However, given that theories of truth are not directly read off the world, but rather serve as parts of larger theories of behavior, this assumption is far from self-evident.

Since any permutation will preserve the truth of the *T*-sentences, the Permutation Principle would be justified if all that were required of a theory of truth was that its *T*-sentences be true.<sup>4</sup> However, a theory of truth is more than a theory about when what a person says is true: it is supposed to play a part in explaining the psychology of speakers, giving the content of both the beliefs and desires we attribute to them and the assertions

*Erkenntnis* 44: 317–326, 1996.

© 1996 Kluwer Academic Publishers. Printed in the Netherlands.

we take them to be making. It is not at all clear that any permutation of an acceptable truth theory will be psychologically adequate. Indeed, it is safe to say that many permutations will fail to be. An *empirically adequate*<sup>5</sup> truth theory need not be a *psychologically adequate* and hence *satisfactory* one. A proper understanding of the role truth theories play in theories of interpretation makes the inscrutability of reference much less wide-spread than Davidson suggests,<sup>6</sup> and, as a result, the radical interpretation methodology is much less likely to saddle its defenders with counterintuitive cases of indeterminacy than is commonly supposed.

Since we learn what words mean by seeing the sentences they are used in, all the *evidence* for a truth theory comes at the level of the *sentences*. We learn what words *refer* to by learning which sentences containing them are *true*. This evidential priority which truth has over reference suggests to Davidson an analogy between semantic concepts like reference and theoretical posits in the sciences:

The theory is correct because it yields the correct *T*-sentences; its correctness is tested against our grasp of the concept of truth as applied to sentences. Since *T*-sentences say nothing whatsoever about reference, satisfaction, or expressions that are not sentences, the test of correctness of the theory is independent of intuitions concerning these concepts. Once we have the theory though, we can explain the truth of sentences on the basis of their structure and the semantic properties of the parts. *The analogy with theories of science is complete: in order to organize and explain what we directly observe, we posit unobserved or indirectly observed objects and forces; the theory is tested against what is directly observed.*<sup>7</sup>

Davidson might here seem to be merely arguing that, since truth is the basic concept in our theory of interpretation, “there is no chance of explaining reference directly in nonlinguistic terms.”<sup>8</sup> We still refer to things, but “it is inconceivable that one should be able to explain this relation without first explaining the role of the word in sentences.”<sup>9</sup> However, Davidson goes on to claim that his argument shows not that there is some real, though conceptually dependent, relation of reference, but rather that we need neither the concept of reference nor reference itself.<sup>10</sup> It should be clear how his acceptance of the Permutation Principle could lead Davidson to such a conclusion.

The Permutation Principle entails that if there is *one* way of assigning entities to expressions that yields acceptable truth conditions for the sentences in a language, there will be *endless* other ways which do so as well.<sup>11</sup> The existence of these empirically equivalent reference schemes prevents the interpreter from uniquely identifying the reference of the interpretee’s predicates,<sup>12</sup> which leads to the inscrutability and hence unreality of reference because “What the interpreter cannot, on empirical grounds, decide about the reference of a schemer’s words cannot be an empirical feature of

those words.”<sup>13</sup> There will thus be no reason to call any one of the possible semantical relations “reference”;<sup>14</sup> there is no fact of the matter about what we refer to, and thus no substantial sense in which we can be said to refer to anything.

However, given how a theory of truth is “part of a more general theory that includes a theory of his beliefs, desires, intentions, and perhaps more”,<sup>15</sup> Davidson cannot consistently hold to both a robust notion of truth and an ‘instrumentalist’<sup>16</sup> conception of reference. Meaning and belief are, as Davidson puts it “interrelated constructs”,<sup>17</sup> and the interpretations we put on a speaker’s words will determine the contents of those beliefs corresponding to the sentences he ‘holds true’.<sup>18</sup> There is thus a parallel between the relation of reference to theories of truth and the relation of truth to theories of behavior: the semantic features of words “are abstracted from the semantic features of sentences, just as the semantic features of sentences are abstracted from their part in helping people achieve goals or realize intentions”.<sup>19</sup> Reference is two steps (and truth one step) away from what must ultimately be explained. Reference and truth are thus, in some sense, like atoms and molecules: the first is used to explain the behavior of the second, but both remain theoretical entities used to explain ‘observable’ behavior.<sup>20</sup> Truth may still be closer to what must ultimately be explained than reference, but *if* the observational/theoretical line must divide one of reference, truth and behavior from the other two, it should be clear that the line will fall between truth and behavior.

Unfortunately, in his discussions of reference, Davidson ignores the fact that the theory of truth is only part of a larger theory of behavior. In one revealing passage from “Reality without Reference”, he describes the issue as whether or not it is at the level of truth or reference that there is “*direct contact* between linguistic theory and events, actions, or objects described in non-linguistic terms”.<sup>21</sup> Even if we grant that Davidson has shown that “direct contact” is not made at the level of reference, he has given us no reason to think that we can have such contact at the level of truth. To say that we make direct contact with the world described in non-linguistic terms at the level of *T*-sentences is no longer to say that truth and reference are like molecules and atoms (i.e.: both theoretical entities with varying degrees of closeness to the observable world), but rather that, while reference is a theoretical notion, the truth conditions of sentences are ‘directly’ observable. This amounts to claiming that we won’t see anything like the sort of indeterminacy found *within* a given set of truth conditions on the level of truth conditions itself. While Davidson thinks that sentence-level indeterminacy will be reduced greatly though the application of the principle of charity,<sup>22</sup> charity is at least as effective

at eliminating the indeterminacy *within* truth conditions that Davidson is willing to allow.

Consider, for instance, the possibility of an atom for atom duplicate of our universe in which events necessarily progressed just as they did in ours.<sup>23</sup> We can call the counterpart of Earth in such a universe “Earth-2”, and if such a universe were to exist, the truth conditions of anything I said about objects on Earth could be given in terms of objects on Earth-2; the ‘Paris’ on Earth-2 is just like the Paris on Earth, and so the truth conditions of any claim I make using the word “Paris” could be given in terms of what happens on Earth-2 as well (provided, of course, that the other terms in the sentence are permuted as well).

Such a permutation would certainly yield an equally true collection of *T*-sentences,<sup>24</sup> but it isn’t clear that it would yield one that can be used to understand the speaker equally well. After all, how are we to make sense of a speaker who seems to be making assertions about items on Earth-2 whenever he sees their Earthly counterparts? One might think that the permutations gain in overall plausibility when we see that they affect not only the truth conditions of what we say, but also the contents of our beliefs, desires, perceptions and so on. However, while the permuted interpretations could extend to the contents of our propositional and perceptual states, this ultimately makes the permuted interpretation *less* plausible rather than more. Indeed, one could argue that the extension of the permutation to our propositional attitudes and perceptual states is precisely what makes them unacceptable. If someone is in a room with a cat on a mat and we interpret him as saying that there is a cat on a mat in the room’s Earth-2 counterpart, his behavior may be in some sense anomalous.<sup>25</sup> Claiming that upon entering the room he forms the belief that a cat is on a mat in the room’s Earth-2 counterpart may help make sense of his utterance, but his suddenly forming such a belief seems even more anomalous than the original assertion.<sup>26</sup> Saying that he forms the belief because, upon entering the room, he sees a cat on a mat in the room’s Earth-2 counterpart, makes sense of the belief formation, but, once again, the anomalousness seems even more grating now that it has been shifted. We started with a seemingly irrelevant statement and have ended up with something which is barely recognizable as a perceptual system.

A considerable loss of psychological explanation is produced by at least *some* theories that permute the interpretee’s beliefs desires and perceptions. A psychological explanation of a person that relies on their seeing a cat when they walk into a room which contains a cat meshes considerably better with, say, empirical work on vision than, say, one which treats them as seeing objects in a parallel universe.<sup>27</sup> We have, in the first case, an

explanation of *how* it is that the interpretee can see what he does; no comparable account is available for how the interpretee can see the cat in the permuted case.<sup>28</sup>

Furthermore, the claim that the content of my utterances could be given purely in terms of items on Earth-2 is clearly in conflict with Davidson's particular brand of externalism and recent accounts of radical interpretation. That is to say, Davidson gives an account of interpretation in which "causality plays an indispensable role in determining the content of what we say and believe".<sup>29</sup> As he puts it:

The basic connection between words and things, or thoughts and things . . . is established by causal interactions between people and parts or aspects of the world . . . The principle is as simple and obvious as this: a sentence someone is inspired (caused) to hold true by and only by sightings of the moon is apt to mean something like "There's the moon"; the thought expressed is apt to be that the moon is there.<sup>30</sup>

Such causal connections between the speaker and what he purportedly refers to seem to be just what is lacking in permutations such as Earth-2. Because of this, Davidson's adherence to the Permutation Principle and resulting claims about the inscrutability of reference are actually incompatible with much of what he says in his more recent discussions of radical interpretation.

This brings us to the subject of causation. While it has occasionally been suggested that 'causal theories of reference' could resolve problems of inscrutability, Davidson's response to such suggestions is to argue that the causal relations themselves can be permuted.<sup>31</sup> This might suggest that the permuted notion of sight can be salvaged by so permuting causation,<sup>32</sup> allowing both schemes have the interpretee seeing what causes his perceptions, etc. However, this would be to misunderstand the role causation is supposed to play in the argument. The appeal to causation is not here an appeal to an independent ingredient to be grafted on to the interpretation. Rather it is stressed here as an element of what is involved in explanation. Our intuitions about 'aboutness' and about what causes a speaker's utterances are both fallouts from our intuitions about intelligibility. We see people as talking about things that cause their perceptions because that is the way we make sense of them. Permuting the notion of causality that falls out of our notion of intelligibility will only be acceptable if intelligibility is preservable through the permutation.<sup>33</sup>

Causality is, then, not introduced as an independent constraint, but rather as the fallout of something like the principle of charity.<sup>34</sup> Reference relations are determined by non 'bent' notions of causality because that is the way *we* actually understand ourselves and each other. One might argue that the fact that we converge towards these nonpermuted manuals is of

little significance since such preferences on our part should not be, as Quine puts it, “mistaken for a substantive law of speech behavior”.<sup>35</sup> However, if one is a pragmatist about meaning (understanding meaning in terms of what the best interpretation preserves), as both Quine and Davidson are, one cannot so easily dismiss the fact that we do not, in fact, find the permuted manuals to be satisfactory. If one were a *realist* about meaning (that is, if one thought that independent facts about meaning determined what constituted the best interpretation) then one could argue that this fact about our preferences was of no real significance in determining what constituted the “best interpretation” of a particular speaker. But, of course, if one were such a realist, one could also view ‘indeterminacy’ as really just being a case of underdetermination. The indeterminacy thesis relies on treating meaning as constitutively tied to our interpretive behavior, and, once this connection is made, it is hard to see why the fact that we clearly prefer some interpretations over their ‘empirically equivalent’ rivals should be of no significance.

Our preference for non-permuted interpretation may still seem to be a contingent fact about us. As a result, it might seem possible for there to be interpreters, aliens perhaps, who had ‘permuted’ similarity spaces, and who would thus come up with queerly permuted manuals for us. However, it is not clear whether this possibility is even intelligible by Davidsonian lights. To say that there are thinkers who interpret perception, causality, etc., in a permuted way, is, for Davidson, to say that there are speakers whom we could *interpret* that way,<sup>36</sup> but if their speech behavior really was the product of a permuted set of similarity spaces, then we would inevitably interpret them as having a set similar to ours. We could have evidence that a speaker had a set of similarity spaces *different* from ours, but we could never have evidence that a set of speakers had a set of similarity spaces that corresponded to a *permutation* of ours.<sup>37</sup> As a result, not only could we never find the permuted interpretations acceptable by our own lights, but we can’t even make sense of the idea of a speaker by whose lights such a permutation was attractive.<sup>38</sup>

Davidson thinks that those who deny that the role of reference is exhausted by its contribution to the truth-conditions of sentences must be denying that theories of truth should be tested solely by evidence concerning sentences and their utterances.<sup>39</sup> However, given that the truth theory must be part of a psychologically adequate theory of behavior, one truth theory may be satisfactory, while an empirically equivalent alternate might fail to be so. The Permutation Principle runs together a theory’s being empirically adequate with its being satisfactory. Reference does more than merely contribute to truth-conditions in this sense: while two truth-theories

may both make a given set of sentences true under the same circumstances, one may be accepted and the other rejected on the basis of how the truth conditions are developed on the level of sub-sentential components. Saying this is not incompatible with believing that the truth theories are tested solely at the sentence level; it is only to insist that such testing does not drive any (theoretical/observational) wedge between truth and reference. One can accept that the theory is verified at the level of theorems, but if enough is built into the verification of these theorems (for instance, that they serve as part of a psychologically adequate theory of the person) one can give a comparatively robust account of the axioms as well.

We should thus reevaluate Davidson's claim that, while we must presuppose a "pre-analytic notion of truth", the same is not required of the concept of reference.<sup>40</sup> Since all the sets of *T*-sentences produced by permuting a satisfactory set are equally true, it couldn't be our pre-analytic notion of *truth* that allows us to pick out some of these sets as psychologically unacceptable. Rather, it seems that some are rejected because their acceptance would require attributing to the interpretee beliefs which, while true, are *about* the wrong things. However, rather than claiming that there must be a 'pre-analytic notion of aboutness' that allows us to pick out the psychologically acceptable truth theories, it may be more accurate to say that we must presuppose a pre-analytic notion of *intelligibility* out of which our near-pre-analytic notions of *both* truth and reference fall. For those who are attracted to the radical interpretation methodology, but less sanguine than Davidson about reconciling the Permutation Principle (and the widespread inscrutability that results from it) with our intuitions about meaning, this should be a welcome result.

#### NOTES

\* I'd like to thank Robert Brandom, Joe Camp, Mitch Green, John McDowell, Ram Neta, Jamie Tappenden, Leora Weitzman and members of the audience at the Central APA for comments on earlier versions of this paper, and the Canada Council for their generous support.

<sup>1</sup> In, among other places, Davidson 1977, 1979 and 1990b.

<sup>2</sup> The most prominent recent example of this reaction is Searle 1987.

<sup>3</sup> Davidson 1979, 230.

<sup>4</sup> Or perhaps true, counterfactual supporting and derived by some sort of "canonical proof" from the axioms of the theory (see Davidson 1984, xiv, 26, 174, for the former qualification, and 61, 138 for the latter), though these additions will not be relevant for the question at hand.

<sup>5</sup> A truth theory is empirically adequate if all of its *T*-sentences are true and counterfactual supporting.

<sup>6</sup> While the permutation principle leads to inscrutability, the falsity of the principle does

not entail that there could be *no* examples of inscrutability; it merely entails that such cases, were they to exist, would be considerably harder to find. In particular, most of the ‘examples’ of inscrutability in the literature seem dependent upon the principle, which is hardly surprising given how they are generated.

<sup>7</sup> Davidson 1990b, 300 (italics mine) see also Davidson 1977, 221–223.

<sup>8</sup> Davidson 1977, 220.

<sup>9</sup> Davidson 1977, 220.

<sup>10</sup> “We don’t need the concept of reference; neither do we need reference itself, whatever that may be.” (Davidson 1977, 224.)

<sup>11</sup> Davidson 1977, 224.

<sup>12</sup> Davidson 1979, 235.

<sup>13</sup> Davidson 1979, 235.

<sup>14</sup> Davidson 1977, 224.

<sup>15</sup> Davidson 1979, 239.

<sup>16</sup> The claim that Davidson’s argument against reference is unduly instrumentalist can also be found in Devitt 1984, chpt. 10.

<sup>17</sup> Davidson 1974a, 146.

<sup>18</sup> Davidson 1974a, 162.

<sup>19</sup> Davidson 1977, 220.

<sup>20</sup> Though the sense of explanation involved, and thus the type of dependence, may be quite different given that truth theories and physical theories are not theories of the same sort.

<sup>21</sup> Davidson 1977, 219, italics mine.

<sup>22</sup> Davidson 1979, 228.

<sup>23</sup> Such a possibility is discussed in Strawson 1959, 21–22, and its relevance to the Permutation Principle is discussed in Kirk 1986, 118–20.

<sup>24</sup> Which would also, of course, be counterfactual supporting and derivable by canonical proof.

<sup>25</sup> Assuming that he hasn’t been exposed to our theorizing about Earth-2, etc. If one were to find out about Earth-2, one could come to form beliefs about it based upon our perceptions on Earth and the knowledge that the two run in tandem. The interpretee discussed here will not, by hypothesis, have such information.

<sup>26</sup> The corresponding desires and their apparent lack of connection to the interpretee’s actions will present an even more serious problem. (Especially since the interpretee’s I-thoughts will have to be understood as being about his counterpart.)

<sup>27</sup> The relevance of such considerations in choosing manuals is also stressed in Putnam 1975 (especially p. 171)

<sup>28</sup> This becomes especially clear when we consider the fact that one can construct permutations of a truth theory which allow one to map a speaker’s words not on to the physical objects around him, but onto the natural numbers (see Quine 1990, 31–4). A theory of perception which claims that all we see are natural numbers, while perhaps solving some problems relating to the epistemology of mathematics, would not sit well with any sort of ‘naturalized epistemology’.

<sup>29</sup> Davidson 1986, 317.

<sup>30</sup> Davidson 1987, 450.

<sup>31</sup> See Davidson 1979, 237 (for a similar argument see Putnam 1983).

<sup>32</sup> Though we might wonder whether we are just changing the subject when we allow a notion of causation which takes us to enter into ‘causal’ relations with the natural numbers.

<sup>33</sup> The argument is thus not supposed to be that causation is an *independent* property which pins down the language we speak determinately (which I take to be Putnam’s target), but

only that *our* conception of causation substantially constrains how we map other languages onto *ours*.

<sup>34</sup> If charity required only that the interpretee's beliefs come out mostly true (as suggested in Wilson 1959, and endorsed in Quine 1960), then what the beliefs happen to be about, provided that truth is preserved, is of little importance. As a result, all the permutations would be equally acceptable. However, this original formulation of the principle, as Davidson himself has argued, is clearly inadequate; what is important is not so much that the interpretee come out as saying things that are *true*, but rather that he comes out as saying things that we would find it *intelligible* for him to say. (See also Grandy 1973). The conception of charity that requires only the preservation of truth rather than intelligibility is tied to a notion of psychological adequacy that focuses on the interpreters ability to *predict* how the interpretee will behave. All permutations of truth theory would be acceptable if the notion of "psychological adequacy" involved only behavior prediction, because attributing beliefs to a person about Earth-2 would allow one to predict their behavior as well as attributing the corresponding set of beliefs about Earth. However, if we want to explain *why* the agent has the beliefs he does, if the theory of interpretation is supposed to yield *understanding* rather than mere *prediction*, then a more sensitive approach to the truth-theory and a more substantial notion of charity are needed.

<sup>35</sup> Quine 1960, 74.

<sup>36</sup> See his 'On the Very Idea of a Conceptual Scheme' (Davidson 1974b), where Davidson argues that for there to be someone with a conceptual scheme radically different than ours, he would have to be interpretable as having such a radically different scheme (since there is no reason to think that 'someone' uninterpretable has a scheme at all).

<sup>37</sup> I have some reservations about this last argument, but I have similar reservations about Davidson's argument in 'On the Very Idea of a Conceptual Scheme', and, given that he does accept the conceptual scheme argument, I do not see how Davidson could be in a position to share any of my reservations about the argument above.

<sup>38</sup> In light of this, consider the following recent remarks about indeterminacy by Quine:

The indeterminacy of translation is unlikely to obtrude in practice, even in radical translation. There is good reason why it should not. The linguist assumes that the native's attitudes and ways of thinking are like his own, up to the point where there is contrary evidence. He accordingly imposes his own ontology and linguistic pattern on the native wherever compatible with the native's speech and other behavior, unless a contrary course offers striking simplifications. We could not wish otherwise. What the indeterminacy thesis is meant to bring out is that the radical translator is bound to impose fully as much as he discovers. (Quine 1990, 48-9.)

This last bit about the interpreter imposing fully as much as he discovers can be understood in two ways. The first, Humean, line treats the 'imposition' as a non-factual projection upon the more basic facts 'discovered', the second, Kantian, line treats both the 'impositions' and the 'discovery' as jointly constituting the relevant facts in the domain in question. Quine and Davidson seem to gravitate towards the first line of thought, while the position above is best understood as a version of the second.

<sup>39</sup> Davidson 1979, 235-6.

<sup>40</sup> Davidson 1977, 223.

## REFERENCES

- Davidson, D.: 1974a, 'Belief and the Basis of Meaning', in Davidson 1984.
- Davidson, D.: 1974b, 'On the Very Idea of a Conceptual Scheme', in Davidson 1984.
- Davidson, D.: 1975, 'Thought and Talk', in Davidson 1984.
- Davidson, D.: 1977, 'Reality without Reference', in Davidson 1984.
- Davidson, D.: 1979, 'The Inscrutability of Reference', in Davidson 1984.
- Davidson, D.: 1984, *Inquiries into Truth and Interpretation*, Oxford University Press, New York.
- Davidson, D.: 1986, 'A Coherence Theory of Truth and Knowledge', in E. Lepore (ed.), *Truth and Interpretation*, Blackwell, Oxford.
- Davidson, D.: 1987, 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, pp. 441–58.
- Davidson, D.: 1990a, 'Meaning Truth and Evidence', in R. B. Barrett and R. F. Gibson (eds.) *Perspectives on Quine*, Blackwell, Oxford, pp. 77–8.
- Davidson, D.: 1990b, 'The Structure and Content of Truth', *The Journal of Philosophy* **87**, 279–328.
- Devitt, M.: 1984, *Realism and Truth*, Princeton University Press, Princeton.
- Grandy, R.: 1973, 'Reference, Meaning, and Belief', *The Journal of Philosophy* **70**, 439–52.
- Kirk, R.: 1986, *Translation Determined*, Oxford University Press, Oxford.
- Putnam, H.: 1975, 'The Refutation of Conventionalism', in *Mind, Language and Reality*, Cambridge University Press, New York.
- Putnam, H.: 1983, 'Models and Reality', in *Realism and Reason*, Cambridge University Press, New York.
- Quine, W. V. O.: 1960, *Word and Object*, The MIT Press, Cambridge.
- Quine, W. V. O.: 1990, *The Pursuit of Truth*, Harvard University Press, Cambridge.
- Searle, J.: 1987, 'Indeterminacy, Empiricism and the First Person', *The Journal of Philosophy* **84**, 123–146.
- Strawson, P. F.: 1959, *Individuals*, Methuen, London.
- Wilson, N. L.: 1959, 'Substance Without Substrata', *Review of Metaphysics* **12**, 521–39.

Manuscript submitted July 6, 1995

University of Pittsburgh,  
 Department of Philosophy,  
 Faculty of Arts and Sciences,  
 1001 Cathedral of Learning,  
 Pittsburgh, PA 15260,  
 U.S.A.