Davidson on reference

J. Robert G. Williams

April 17, 2012

> [Reference is] a theoretical construct, whose function is exhausted in stating the truth-conditions for sentences (Davidson, 1977, p.223)[1]

Reference leads a strange double life in Donald Davidson's philosophy of language. On the one hand, it is at the heart of his account of the meaning of words and the learnability of language. On the other, he is concerned to combat overly 'realist' views about the relation. And he explicitly endorses the seemingly incredible position that reference is wildly *inscrutable*. There is no fact of the matter *at all* about what our words refer to.

Section 1 reviews the role of reference within Davidson's T-theoretic account of language. Section 2 examines the case for the inscrutability of reference based on Davidson's account of radical interpretation. Section 3 concerns the explanatory role of reference, which we cover in two parts: the first concerning explanations that *directly* appeal to reference; the second concerning explanations that appeal to *beliefs* about reference.

1 Reference in T-theories.

1.1 Fitting an infinite capacity in a finite head.

Each one of us is in a position to understand infinitely many, never-before-encountered sentences of our language. How, even in principle, could a finite being manage to acquire such an infinite capacity?[2]

Suppose that the infinitely-many sentences have a very simple vocabulary and syntax. There are one name "Asa", one (monadic) predicates "runs", and the sentential connectives "not: ..." and "...and...". They combine into grammatical sentences just as you would expect. This generates infinitely many sentences, e.g.:

- Asa runs

[1] All page references to articles by Davidson will be to the collected editions (Davidson, 1984)
[2] (Davidson, 1965, 1967)

- Not: Asa runs
- (Asa runs and Not: Asa runs).
- Not: (Asa runs and Not: Asa runs).
- …

A T-biconditional links the truth of any one of these object-language sentences to a way the world might be. For example:

- "Asa runs" is true iff Asa runs.
- "Not: Asa runs" is true iff Asa doesn't run.
- "(Asa runs and Not: Asa runs)" is true iff Asa runs and Asa doesn't run.
- "Not: (Asa runs and Not: Asa runs)" is true iff it's not the case that Asa runs and Asa doesn't run.
- …

Users of the language are in a position to pick out biconditionals of this form for each sentence they understand. This itself is an infinitary capacity, and surely not unconnected to the infinite capacity to understand the sentences (we'll look at what the relation is further in §3.2). Let's see how a finite creature might in principle get into a position to know all these biconditionals.

The obvious solution is for the creature to have a small range of core information from which the whole range of claims we've identified may be derived—plus the ability to derive any given claim on the list when prompted. This is the job description for Davidson's T-theories. The core information are the axioms of the T-theory; and it is here that the notion of reference is introduced. A sample T-theory for our toy language might be:[3]

1. "Asa" refers to Asa.
2. For any name N, "N runs" is true iff the referent of N runs.
3. For any sentence S, "Not: S" is true iff S is not true
4. For any sentences, S, T, "(S and T)" is true iff S is true and T is true.

Someone who knew these axioms, with minimal reasoning, could work out each T-biconditional on our earlier list; and infinitely many more besides. Nor would they need creativity in figuring out *how* the derivation should go. There is a canonical method, depending only on the syntax of the sentence in question, for deriving the

[3]I follow (Lepore & Ludwig, 2005) in the construction of this toy T-theory.

corresponding T-biconditional. A derivation from the axioms of a T-biconditional that implements this method will be called a *canonical derivation* and the T-biconditional that results the *canonical T-biconditional for S* of the T-theory in question.[4]

We have seen one way of accounting for our (infinite) capacity to know each of the T-biconditionals on the list in question, by storing a small range of information and draw out their consequences. Davidson (1965) claims that in setting out this T-theory, we have also answered the original question of explaining how *in principle* we could acquire the infinite capacity to understand all the sentences—for he claims that learning the T-theory (or perhaps: learning the T-theory and that it is a T-theory; or learning the T-theory and knowing that it is a T-theory adequate by the lights of radical interpretation) suffices for understanding.

At the heart of Davidson's explanation of the learnability of language, then, are T-theories. And at the heart of T-theories lies reference.

1.2 Conditions under which reference is required.

Our toy language is extraordinarily simple, and natural languages are extraordinarily complex. Whether a Davidsonian T-theory can be constructed for natural language and if so, what shape it will finally take, is a major empirical issue.[5] This matters! Perhaps it will turn out that for English, reference is not the key notion in generating a T-theory; perhaps a specification of the intension of a name rather than its reference will be required; perhaps we will find a need for reference relations for expressions in other syntactic categories.[6] However, I'll work here under the assumption that reference remains at the heart of a T-theory for natural language as well as toy examples.

It is illuminating to study in very simple cases the conditions under which to answer the learnability question we require a referential axiomatization (or more generally, an axiomatization that attributes semantic properties to subsentential expressions). The toy language given above is so simple, that appeal to reference doesn't seem mandatory to explain an ability to derive each T-biconditional. For example, we could replace (1) and (2) above with the single axiom

(1+2) "Asa runs" is true iff Asa runs.

This would bring both quantitative and qualitative parsimony to the theory. Quantitatively, we switch two axioms for one of comparable complexity. Qualitatively, we eliminate appeal to any relation other than truth from our theory. From this

[4]For canonical derivations, see, inter alia, (Davies, 1981; Kölbel, 2001; Lepore & Ludwig, 2005)

[5]See for example (Larson & Segal, 1995; Lepore & Ludwig, 2007).

[6]The semantic theory presented in the influential textbook (Heim & Kratzer, 1998) is an interesting point of comparison and contrast, since, like Davidson (and unlike e.g. (Lewis, 1970), it emphasises the derivation of (something close to) T-sentences, but works from the start with semantic values for expressions of all types, and ultimately introduces intensions.

perspective, appeal to reference to explain the in-principle learnability of the language is otiose, and rather unmotivated.

As we add more names and predicates into our toy language, however, the two kinds of economy work against each other. With ten names, and ten monadic predicates there will be one hundred atomic sentences in the toy language ("Asa runs", "Asa walks", "Beth runs",…). Finitely axiomatizing the T-theory in our later "truth-only" style, via a T-biconditional for each atomic sentence, plus the recursive axioms, would require 102 axioms—one for each atomic sentence, plus the recursive clauses for connectives labelled (3) and (4) above. If we stuck with the approach of our original toy theory, with an axiom stating the referent of each name (plus an appropriate axiom for each predicate) we require only 20 axioms plus the axioms for the connectives.

There is a kind of structure to the knowledge obtained via learning the 22 lexical axioms.[7] If our agent forgets one piece of the basic information—for example, she forgets what "Asa" refers to—then she loses the ability to derive T-biconditionals for the ten atomic sentences in which "Asa" figures. Conversely, if she learns the referent of a new name, she is immediately in the position to understand ten new atomic sentences. An understanding obtained through learned a separate piece of information for each atomic sentence separately would not manifest these kinds of patterns.

We have three results. First, even with many names and predicates, appeal to reference is theoretically dispensable if one is solely concerned with specifying a finite route to gain an infinite capacity. Second, there's a clear sense in which the extra ideology of reference simplifies the learning task: far fewer pieces of information need be stored on a subsentential axiomatization. And finally, one who learned a language via the subsentential route would have a kind of structured ability that matches patterns familiar from our own case. The significance of this last point will depend on what *other* than answering the in-principle learnability question we wish to extract from studying T-theories—a question we turn to shortly.

There is a point at which the appeal to subsentential semantic properties becomes (prima facie) ineliminible.[8] When *quantifiers* are added to the language, one cannot even in principle get away with just talking of truth (or even reference and truth, as in our original theory). To relate the truth-conditions of "something runs" to information about the predicate "runs", we need more than just information about how that predicate combines with names—we need to know whether the predicate applies to arbitrary things in the domain of discourse (not just those that happen to be named). One important change that is required is the need for a second subsentential semantic notion (satisfaction; or, what amounts to the same thing, the relativization of truth to a variable assignment). But equally, one cannot even in principle derive the T-biconditionals for all sentences from the T-biconditionals of the atoms.

I note one last characteristic of Davidson's approach: compared to his rivals, he has only a limited role for reference. Davidson's preferred approach assigns referents to names

[7] Compare (Evans, 1981).

[8] I'm emphasizing quantification here; but see (Evans, 1975) for further considerations about the kinds of linguistic structure that motivate discerning reference within language (emphasizing in particular predicate modification).

(and other singular terms—descriptions and demonstratives, for example). He does not appeal to reference—or indeed any relation between word and 'semantic value'—for other linguistic items such as predicates (or connectives or complex expressions such as sentences). This contrasts with a Fregean approach,[9] which might replace (2) in our toy theory with:

(2*) "runs" refers to the function f such that, for all x, f maps x to the True iff x runs.

If we switch from (2) to (2*), we have to add extra information to tell us about how this information about the reference of parts of sentences relates to their truth, e.g.:

(5) for N a name, and F a predicate, "N F" is true iff the referent of "N" is mapped by the referent of "F" to the True.

From a Fregean perspective, the earlier axioms for predicates fuse two separable bits of information: the semantic significance of predicates (their reference), and the semantic significance of concatenation (in this case, function-application). It is a substantive issue whether or not the technical project of natural language semantics requires such a disentanglement, or can proceed without it.[10]

Following Davidson, we've been presenting the significance of T-theories as an answer to the question of how a finite agent can *in principle* acquire the capacity to understand a language. But it's natural to ask the follow-up: how do *we* manage to pull the trick off? Can T-theories illuminate our own case?

The obvious conjecture is this: the way we ourselves understand a language is by having internalized a T-theory. This involves inter alia, believing the referential axioms, and standing ready to derive T-biconditionals for arbitrary sentences in the language in question. On this view, our understanding of language is ultimately constituted by our *beliefs about what words refer to.* This does not seem to be Davidson's view (Lepore & Ludwig, 2005, pp.119-124)

[9] As set out e.g. in (Heim & Kratzer, 1998)

[10] Compare (Heim & Kratzer, 1998; Larson & Segal, 1995). Davidson's T-theories use a satisfaction relation on predicates, which requires a compositional axiom to link it to claims about truth. See (Davies, 1987) for dicussion (if one is to do without satisfaction, one must give an alternative semantics for quantifiers—perhaps such as the one (**?**) attributes to Frege).

The project of general semantics(Lewis, 1970) promised that the semantic significance of concatenation in general, for arbitrary syntactic categories, would be function-application (compare the analysis of connectives as referring to truth functions, which then operate on the semantic values of sentences—truth values). This project hit problems, as discussed in the appendices to the collected edition. (Heim & Kratzer, 1998) show how to cover a large variety of concatenation with just a few kinds of compositional principles. There seem to be interesting generalizations (and cross-linguistically stable ones) about the semantic significance of concatenation, that would be hidden from view if we packed them into axioms. Moreover, disentangling composition from lexical axioms in the right way may allow us to articulate theses about different ways in which semantic theory is learned: for example, perhaps our understanding of concatenation is *innate*, whereas clearly lexical meaning must be learned.

It is often suggested that the T-theoretic model illuminates our actual understanding—that our understanding of language consists of a *structured ability*, and the ability to understand a sentence is derived from *subabilities* systematically associated the sentences' components.[11] If so, then although the *theory* appeals to reference, all we can say in the abstract is that there is some 'subability' that plays the same role—and it's very much an open question what sort of thing this is.[12] The worry with this is that it abstracts away from the very feature of learning T-theories that gave them their explanatory power. In deductive reasoning, we have an extant model for how a set of inference rules, applied to a finite stock of beliefs, can put us in a position to acquire any of an infinite stock of beliefs. But it's not clear what non-doxastic model is available for putting together a finite stock of abilities to "derive" some set of "structured abilities". Some (even if not Davidson himself) have thought that best version of the Davidsonian project will involve attributing attitudes to reference to language users as part of linguistic competence. We look at this view—the *strong cognitive conception of understanding*—in §3.2 below.

2 The inscrutability of reference.

In virtue of what do our utterances mean what they do? It is tempting to see a T-theory as one step towards an answer. What it is for "Asa runs" to be true iff Asa runs, we may say, is inter alia for "Asa" to refer to Asa, for "runs" to have the appropriate satisfaction conditions. Having reduced the question of truth to the notions of reference (of names) and satisfaction conditions (of predicates), we can then look outside language, to get a fix on what sort of non-semantic facts make it the case that a name refers to a particular individual, or a predicates is satisfied by some object. Some kind of systematic causal relation between tokenings of the name and states of the world involving the object, might be appealed to at this point. The general strategy—the building block approach—reduces truth to reference (et al) and then reference to causation (et al).[13]

Davidson views the building block approach as misguided. Interestingly, he says that it *would* work on condition that a reduction of the reference relation to non-semantic relations could be performed.[14] His case against it rests on pessimism over the availability of this final step. He doesn't lay out the grounds for his pessimism—as he says, many have been attracted to the view, and tried to make it work.[15] But I think we should have sympathy for Davidson's pessimism. The building block theory is based on a grand promissory note, of reducing to the non-semantic the semantic significance of expressions in all categories. It is hard enough to get a credible theory for names of medium sized dry goods and their bearers, but a building block metasemantics needs much more: to account for the reference of singular terms for unobservables, for

[11](Davidson, 1967, p.25) says that a success T-theory requires it recover "the structure of a very complicated ability—the ability to speak and understanding a language".

[12]Compare the 'mirror constraint' and the discussion of 'tacit knowledge' in (Davies, 1981; Wright, 1981; Evans, 1981; Davies, 1987; Miller, 1997).

[13]A locus classicus is (Field, 1972). The name is introduced in (Davidson, 1977).

[14](Davidson, 1977, p.219)

[15]One influential programme in the 1980s was the asymmetric dependence causal analysis of (Fodor, 1987).

abstracta; and beyond singular terms, for determiners, modifiers, connectives, and many other things.

The building block approach to the metaphysics of meaning adopts one familiar strategy: identification of the queried relation (reference) with some relation in the "reductive base". Davidson instead aims to provide a characterization of when a T-theory as a whole (in particular, a set of lexical axioms) is *selected* by an appropriate range of non-semantic facts. The account of what this "selection" amounts to we find in Davidson's discussion of radical interpretation.[16] In particular, we find the radical interpreter noting regularities in patterns of assent and dissent, and deriving from these a set of target T-biconditionals. The selected T-theory is, minimally, one that will enable one to derive the target T-biconditionals.

From a metaphysical point of view, it is less transparent what is going on here, than in the case of simple reductive identification. Whether the kind of indeterminacy or inscrutability of reference we're about to mention is even coherently formulable, depends on the view of the metaphysics of meaning that it is paired with; I will delay further discussion until we have inscrutability on the table.

2.1 Twisted T-theories and the grain of data.

As Davidson describes the process, the selection of a T-theory proceeds primarily by gathering data at the level of the T-biconditionals the selected T-theory must entail. Thus, on the basis of an observed regularity such as:

> Subjects hold-true "Asa runs" iff Asa runs.

The radical interpreter will attempt to find T-theories that entail:

> "Asa runs" is true iff Asa runs.

There will be many such data-points to account for. Many T-theories can be ruled out on this basis. For example, one that assigned Beth as the referent of "Asa", while giving "runs" its usual interpretation, would entail that "Asa runs" is true iff Beth runs, and fail to entail the target T-biconditional. On these grounds, the T-theory is knocked out of the running.

But odd-seeming theories are (seemingly) let through this filter.[17] For example, suppose that we consider the following T-theory for our toy language (with the standard clauses for connectives):

[16](Davidson, 1973, 1974, 1975, 1980). For a nice comparison and contrast between Field's building block approach and the Davidsonian one, see (McDowell, 1978). (Lepore & Ludwig, 2005) contains an excellent, detailed examination of the prospects and pitfalls of Davidson's account of radical interpretation.

[17]See (Davidson, 1979). For other discussions of permutation based arguments for inscrutability, see (Quine, 1964; Wallace, 1977; Putnam, 1981).

1*. "Asa" refers to the best friend of Asa.

2*. "N runs" is true iff the referent of N is the best friend of someone who runs.

Given that Beth is Asa's best friend, the first axiom assigns Beth, rather than Asa, as the individual to whom the name refers. But the second axiom, for runs, makes a compensating twist, and so, according to theory, we will derive:

> "Beth runs" will be true iff the best friend of Asa is the best friend of someone who runs.

The right hand side of this biconditional is obviously equivalent to the requirement that Asa runs (at least under the assumption that Asa has a best friend). So it looks like by twisting the reference scheme, and then compensatingly twisting the interpretation of predicates, we can derive targetted T-theorems in unexpected ways.

In general, for any set of "target" T-biconditional generated by a "sensible" T-theory θ, and an arbitrary individual β, we can design an alternative T-theory ϕ that both assigns β as the reference for "Asa" and still allows one to derive the T-biconditionals.

Here's how to do it. Suppose θ says that "Asa" refers to α. Now choose a 1-1 mapping of individuals onto individuals, f (a "permutation" of the universe), such that f maps α to β. The chosen function could, for example, map α to β, and β to α, and leave every other object fixed.

We can then piggyback on θ and f to specify the new T-theory ϕ. For each name N in the language, if θ includes an axiom that says that N refers to n, ϕ is to include an axiom that says that N refers to $f(n)$. If θ includes an axiom for predicate P that says that "N P" is true iff the referent of N is F, then ϕ will include an axiom that says that "N P" is true iff there is some x such that $f(x)$=the referent of N, and x is F.

In general, the T-biconditionals derivable in θ for atomic sentences will take the form:

> "N P" is true iff α is F.

And the corresponding T-biconditionals derivable in ϕ take the form:

> "N P" is true iff there exists some x such that $f(x) = f(\alpha)$ and x is F.

The two sentences are obviously equivalent, so the original T-biconditional will be derivable in the twisted theory (in this case, the equivalence presupposes the "side information" that f is a function defined for α).

A particularly dramatic kind of underdetermination of T-theory by data arises.[18] The data doesn't even eliminate T-theories according to which "Aristotle" refers to a small furry creature on Alpha Centauri, rather than the ancient Greek philosopher. When extended to language with demonstratives, the argument implies that the demonstrative "that", uttered while you poke your finger at a ball you are holding up and making salient, might just as well refers to the Titanic, rather than the poked ball. If 'selection' of a T-theory doesn't impose further constraints, then it appears that either the twisted theories are selected along with the sensible one, or no T-theories are selected at all.

Davidson appears to embrace the first horn of this dilemma. He takes the conclusion to be that there is simply no fact of the matter which T-theory "gets things right". And what we learn from permutation constructions is that reference is utterly indeterminate. For any x whatsoever, there is no fact of the matter about whether or not "Asa" refers to x.

2.2 The grain of the data

Do twisted T-theories really generate the data just as well as sensible ones? It is important to note that the argument is sensitive to exactly what is taken to be required for a T-theory to 'generate' some target T-biconditionals. A very strict understanding of this relation is possible. To motivate this, note that even in a sensible T-theory, if we can show:

> "Asa runs" is true iff Asa runs

Then if "derivability" is as strong as first-order consequence, then equally the following T-biconditional will be derivable:

> "Asa runs" is true iff Asa runs and everything is self-identical.

Intuitively the first "gives the meaning" of the sentence mentioned on the left, in the way the second does not.[19] As mentioned earlier, sensible T-theories will generate what appears to be meaning-specifying biconditionals 'canonically'—in a way that step-by-step reflects the syntactic build up of the sentence. From sensible axioms the first T-biconditional above is *canonically* derivable, while the second is not. Its proof requires some extra steps: proving the relevant tautology and adjoining it to the biconditional. Valid moves, true, but not ones that have a place in the canonical derivation.

[18] Well, at least it does if the above result can be generalized to total theory. As we'll see below, it's worth thinking hard about the details—and paying particular attention to the range of devices involved in natural languages, which may call for extra tweaks. In the case of intensional semantics, an extremely general version of permutation inscrutability, for multiplying intensional type theory, is given in (Williams, 2008b).

[19] See (Foster, 1976; Davidson, 1976)

What happens if in radical interpretation we require target T-sentences to be *canonically* generated from the T-theory?[20] What is canonically derivable from the twisted theories described earlier are T-biconditionals such as:

> "Asa runs" is true iff Asa's best friend is the best friend of someone who runs.

Or, for permutation f:

> "Asa runs" is true iff there is an x, such that $f(x) = f(\text{Asa})$, and x runs.

So the twisted theories simply fail the test of allowing the *canonical* derivation of the following target T-biconditional:

> 'Asa runs' is true iff Asa runs.

What the twisted T-theories canonically generate are biconditionals obviously equivalent to the targetted one. But that cuts no ice if canonically derivability is really the criterion they are to be judged by.

Does this mean that we don't need to worry about the inscrutability of reference in a Davidsonian theory, then? That would be far too quick. For the move to canonical derivability essentially fine-grains the data to which we are helping ourselves (the target T-biconditionals). The radical interpreter might find it comparatively easy to shift from a set of T-biconditionals to a T-theory, but how is she to get a fix on the T-biconditionals themselves, on the basis of the kind of hold-true data we are supposing? Thus, suppose it's a regularity that:

> Speakers hold-true "Asa runs" iff Asa runs.

The following will also be a regularity:

> Speakers hold-true "Asa runs" iff there is an x, such that f(Asa)=f(x), and x runs.

So which regularity should the radical interpreter take as her guide to picking a target T-biconditional? If the former, then the sensible interpretation of the language will be (determinately) selected. If the latter, then the twisted interpretation will be (determinately) selected. There is still, prima facie, just as much indeterminacy in the process of selecting a T-theory. It simply enters at an earlier stage.

[20] See(Williams, 2008a).

Theorists of radical interpretation (if not Davidson himself) could try to block the argument for radical inscrutability by putting pressure at this point. For example, perhaps not all regularities are created equal. Some regularities may be lawlike, others not. and perhaps only *lawlike* regularities should be data the radical interpreter uses to form target T-sentences.[21] For this to be relevant to the current case, one would have to argue that our first regularity is lawlike, and the second (permutation-involving) regularity is not. There are a few options to explore here. Even if the two regularities are conceded to be necessarily equivalent, it might be that lawlikeness is hyperintentional: only the *simple* formulations count as laws. Or one could question lawlikeness by pointing to issues with their alleged equivalence. For example, suppose that f(Asa)=Beth. Now consider a situation in which Asa exists but Beth does not. Relative to that situation, the condition laid down in the right hand side of the 'twisted' T-biconditional is false, though that laid down in the original is true.[22]

2.3 Simplicity.

We've seen one reason for thinking that radical inscrutability of reference does not follow even from Davidson's very liberal constraints on theory-selection. To exploit this line in defending the determinacy of reference would require an ultra-fine-grained story about the target data for radical interpretation.

The other way of blocking inscrutability appeals, not to fine-grained data, but to theoretical constraints the radical interpreter uses to pick a T-theory on the basis of that data. We might argue, for example, that among the various theories that would generate the right T-biconditionals (or something obviously equivalent to the right T-biconditionals) the radical interpreter should believe the simplest. This seems not at all ad hoc, but simply an application of usual canons of theory choice to the particular situation the radical interpreter faces.

Simplicity, it might be said, will favour sensible axioms such as

1. "Asa" refers to Asa
2. "N runs" is true iff the referent of N runs

over their twisted, and slightly more complex, counterparts:

[21] Compare Davidson's appeal to lawlike regularites in (Davidson, 1984, introduction).

[22] Compare (McGee, 2005; Williams, 2008b). Why then did I say earlier they seem 'obviously' equivalent? Well, they are equivalent under the 'side constraint' that f is a well-defined function. In any counterfactual scenario in which f maps Asa to anything, the two conditions are equivalent. Moreover, just as one can arguably know a priori the (contingent) truth that the standard metre is a metre; one can arguably know a priori the (contingent) truth that f is a function that maps Asa to Beth. So the two conditions may be a priori, though only contingently, equivalent.

Note that we would be unable to rule out on these grounds twisted T-theories that map Asa to Asa's singleton, or to necessary existents like numbers. So arguably the strategy does not go to the heart of the problem.

1*. "Asa" refers to f(Asa).

2*. "N runs" is true iff there is an x, such that f(the referent of N)=f(x), and x runs.

So the sensible theory will rank as a *better* theory accounting for the relevant data, when compared to its twisted rivals.

This is the kind of thing that David Lewis once advocated as a response to "Quinean" worries about indeterminacy within an interpretationist metasemantics.[23] And he added a useful caveat: facts about reference, he said, would on this account be "no more objective" than our canons of simplicity. One can appreciate the concern: simplicity might very well be understood in ethnocentric terms—as what people like us find simple, or tractable, or useable.[24] But if that's the case, then it seems that even if indeterminacy is avoided, a relativity to our practical concerns or limitations is not.

One way in which simplicity might be relative, is that its evaluation may be sensitive to the language the interpreter happens to use. Suppose, for example, that we introduce a descriptive name for f(Asa)—Fasa, say. And suppose that we introduce a primitive predicate "fruns" synonymous with "y such that there is an x, such that f(y)=f(x), and x runs".[25] Then we could replace (1*) and (2*) above with:

1**. "Asa" refers to Fasa.

2**. "N runs" is true iff the referent of N fruns.

Now from a syntactical point of view, this new theory exactly parallels the original, sensible one. To the extent that we are using syntactic economy as a stalking horse for whatever criterion of theoretical simplicity we will ultimately appeal to, then (at least for radical interpreters with the enriched metalanguage) there is no discriminating between sensible and twisted T-theories on these grounds.

One reaction is to say that there is a *privileged* metalanguage in which to assess simplicity. This is the root, as I see it, of David Lewis's appeal to eligibility as a constraint on radical interpretation.[26] In essence, Lewis picks a canonical metalanguage (the language of "Ontolegese", where primitive predicates stand for metaphysically fundamental properties) and constrains interpreters to pick theories that (all else equal) are simplest when formulated in it. Davidsonians impatient with the Lewisian metaphysics (or even those happy with the metaphysics, but disliking its invocation as a way of "objectivizing" the notion of simplicity) might still be able to use the *general* idea if they can by some other means identify a *privileged* metalanguage from which to assess simplicity/complexity.

[23](Lewis, 1975). For more on Lewis' version of radical interpretation, see (Lewis, 1969, 1974, 1984, 1992).

[24]Compare the detailed discussion of the appeals to theoretical virtues in generative linguistics in (Ludlow, 2011).

[25]Compare the trouble for Humean theories of laws described in (Lewis, 1983)

[26](Lewis, 1983, 1984, 1992). I discuss and defend this reading in (Williams, 2007, ms.)

2.4 The formulation of inscrutability: relativizations

Suppose, pro tem, that total inscrutability of reference does follow from Davidson's overall position. In the next section, we will look at Davidson's thesis of the unexplanatoriness of reference, which might help to draw the sting from the initially counterintuitive thesis being put forward. But a preliminary question is how this thesis should even be formulated.

The threat of inscrutability arises because multiple theory are 'selected' by the radical interpretation procedure. But on the face of it we can't consistently endorse all of them—they say incompatible things! But that leaves us with a puzzle. If we're to use one of the T-theories in interpreting one another, aren't we committed to thinking it gets things *right*—but thinking it's *right*, and hence the incompatible coselected T-theories are *wrong*, seems exactly what the thesis of the inscrutability of reference was trying to avoid commitment to.

One option is to deny that the T-theories that are 'selected' are inconsistent. Relativizing the semantic predicates they involve is an obvious way to achieve this. There is no contradiction in saying that "Ada" refers$_1$ to Ada and not f(Ada); but refers$_2$ to f(Ada) rather than Ada. Davidson in fact endorses a version of this proposal. Semantic properties on his treatment are already relativized to a language, as well as a speaker and a time (the existence of homologues and homophones across different languages makes this necessary: 'gift' applies to poisons in Scandanavian languages, but applies to very different things in English). Drawing on this, Davidson suggests that rather than two candidate reference relations for a single language, we have two languages, each with a single candidate reference relation. You might have thought "English" (or "my idiolect of English") named a single language. For Davidson, it picks out an infinite cluster of languages, one for each precisification of the reference scheme.[27] The idea is striking. But it's unclear whether it answers the problem. What makes the word "Ada" in one such L, relate to Ada rather than fAda?[28] Not the empirical factors that feed into radical interpretation, ex hypothesi[29]—but then what are these mysterious 'non-empirical' factors that complete the job?[30] Indeed, the very supervenience of word-meaning on non-semantic facts, that Davidson took as initial motivation for his project, appears to be being denied—since the relevant non-semantic facts are the same for each of the "languages" being posited, and yet the reference relations differ.

Davidson's approach to relativization is to give an informative account of what the relativization is to—in this case, a language. Relativity to language and to context

[27]See (Davidson, 1979, p.239).

[28]Davidson says "all that we can say gets fixed relativization is the way that we answer questions about reference, not reference itself". I don't understand this qualification, but it seems relevant to the challenge being put forward here.

[29]"it is not entirely an empirical question which language a person speaks" (Davidson, 1979, p.240)

[30]Davidson appears to think that 'the language that P speak' fails to pick out a unique language, but we can describe one among the cluster by specifying which T-theory we are using. But we if picked out the L to which we relativize *as* the one spoken by population P *and* characterized by T-theory θ, then the claim that θ is the right T-theory for L will become tautologous, and the empirical constraints of radical interpretation utterly rundundent. What we want is a sense of how the work factorizes, into an emprical component provided by radical intepretation, and some non-empirical component somehow built into the languages themselves.

(speaker and time) are precedents for this approach. The other relativization strategies he considers (and rejects) share this characteristic. But there are other models for relativization within semantics. In quantification theory, we relativize semantic properties (satisfaction, in particular) to a sequence of objects. The relativization "disappears" at the level at which the theory makes contact with data—we say that S is true simpliciter just in case S is satisfied relative to every sequence.

We could relativize reference and predicate-satisfaction to *arbitrary* indices. Our metasemantics will say that every individually empirically adequate T-scheme should be assigned an index in this super-T-theory. We can then derive relativized T-biconditionals, and our rule will again be that a sentence is true simpliciter iff it is true on every index. If it is possible to derive that S is true-i iff p, for all i, then by the definition of truth simpliciter, we will be able to derive the unrelativized T-biconditional that S is true iff p. For the case of permutation inscrutability, where the truth conditions of whole sentences are (with the caveats noted earlier about fine-graining derivability) held fixed, then the relativization makes no difference at the level of whole sentences. In other cases of indeterminacy of interpretation, sentences might (in given conditions) come out true on one adequate interpretation, and false on another. The kind of approach just sketched (a broadly "supervaluational" one), predicts in this case they are neither true nor false.[31]

The abstract relativization move is worth considering, if only for its popularity elsewhere in philosophy. To mention one immediate worry: it's not obvious that it resolves the problem of "too many adequate theories"—if anything, it worsens it, by adding yet another seemingly adequate theory, generating the target T-sentences, to the mix. Does it have virtues the original cluster lacked? It certainly is *less committal*, not making any definitive predictions about what refers to what. Perhaps a theorist should favour theories, all else equal, that are cautious in this way.

Davidson's frequent invocation of instrumentalism suggests a third response, rather different from his official relativization strategy. The characteristic move of the instrumentalist about theory $\boldsymbol{T}$ is to shift attention from whether $\boldsymbol{T}$ is true (or whether we believe it) to whether $\boldsymbol{T}$ is useful (and what we can use it to do). In the context of T-theories, rather than asking in the first instance about whether "Ada" refers to Ada, we might instead ask: under what conditions is it ok to use a T-theory θ that says that "Ada" refers to Ada, in interpreting a speaker? Here are two possible answers:

(a) The use of θ is ok so long as θ fits the relevant data in the way described in radical interpretation.

(b) The use of θ is ok so long as (inter alia) "Ada" refers to Ada.

[31]For supervaluations, see (Fine, 1975; Keefe, 2000). Another way a Davidsonian could get this effect is by endorsing *disjunctions* of T-theories. If θ and ϕ are selected T-theories, then one could endorse $\theta \vee \phi$, and use reasoning by cases to derive the common consequences of each theory. Notice, however, that the derivations would not have the usual canonical form (requirement two such derivations under suppositions, plus the disjunction-discharge step). Moreover, only disjunctions of whole theories would work—disjoining the lexical atoms individually would not have the same power, since one loses the 'penumbral connections' between twisted reference schemes and twisted predicate axioms that is vital to pulling off the trick.

Someone who endorses (b) commits themselves to the view that T-theories used in interpretation must get the semantic facts right. Endorsing (a) does not have this commitment. For all (a) says, a T-theory I use to interpret you could be false, or meaningless—the question of θ's status simply does not arise if (a) alone is the criterion. The advocate of (a) will insist that there is no direct inconsistency between the view it is permissible to use θ in interpretating a speaker, and also permissible to use ϕ for the same task—even if θ and ϕ themselves are mutually contradictory. Two fictional tales can be equally good at helping a child to sleep—they don't have to be mutually consistent to play that role.

Of course, giving answer (a) cannot prevent us from asking the question whether a particular theory θ is in fact true. And it looks like (for the reasons set out above) that unless we somehow adjust what the T-theories say (by a relativization strategy, for example), then (b) tells us straight off that it cannot be that θ and ϕ are both permissible interpretative strategies. If mutually incompatible theories both "fit" the radical interpreters data, therefore, then (a) and (b) contradict each other. (a) will tell us that the theories are all ok to use in interpretation. (b) tells us that at least one of these theories is not ok to use in interpretation. One reaction would be to hold (b) fixed, and consider this a problem for those Davidsonians who would endorse (a). But a legitimate and interesting reaction for the Davidsonian, it seems to me, is to hold onto (a), and dismiss (b).

This kind of view has genuine instrumentalist precedent. Consider the following formulation the position on the use of a mathematics in science in (Field, 1980, 1989). Field's position is that it is acceptable to use mathematics in science so long as the mathematized science is conservative over true, nominalized science. But he rejects that permissible use requires the mathematized scientific theory itself be true ("Mathematics doesn't have to be true to be good"). Field's own original view is that we should in fact disbelieve existentially committing parts of mathematics. But the core of his view is that the issue of the truth of mathematics is *irrelevant* to its applicability. Note that analogues to the situation of incompatible T-theories can well arise. Two incompatible set theories, each conservative over nominalistic science, might equally well serve the purposes of formulating an instrumentally useful mathematized science.

Davidson explicitly endorses both the relativization strategy, and the analogy to instrumentalism. In the terms just sketched, we can view this as him endorsing (a), and thinking of this as *sufficient* to account for the utility of T-theories; but then rather than rejecting (b) (as an error-theorist like Field advocates) he attempts to hold onto it by trying to make the T-theories compatible via relativization. Rather than pursue relativization, perhaps he would have been better giving up on (b) altogether, and simply insist that semantics does not have to be true to be good.

3 Explanations and reference.

Suppose Ada raises her arm. This may (in context) be explained by pointing to the beliefs and desires that Ada has—that she wanted to be excused, and believed that the

best way to do this was by attracting attention by raising his arm. Beliefs, and desires, *prima facie*, have a wide explanatory role—they explain a wide variety of events in the outside world. Plausibly, they *immediately* explain actions, and mediately, *via* the actions, many other events are brought about. Presumably, an agents' beliefs are also involved in immediate explanations of various other aspects of her psychology (for example, emotional states) which themselves go on to have behavioural and worldly consequences.

It is less clear that semantic properties of sentence (truth, reference, and the rest) feature in such immediate explanations of worldly happenings. If they do, the most likely candidate explananda would be to do with communication. The fact that I uttered "Ada runs" may be explained by the truth of that sentence (given the fact that I'm a reliable detector of the truth of such sentences, a truthful interlocutor, etc). And the fact someone formed the belief that Ada runs may be explained by the fact that the truth conditions for "Ada runs" are that Ada runs, plus the fact she heard me utter those words, trusts me as informant, etc. A first question for anyone thinking about the explanatory role of semantic properties concerns the *breadth* of that role—the varieties and kinds of events that are (in part) *immediately* explained by citing semantic properties.

Say that a property P is involved in an explanation *directly* if one of the things that explains the event in question is something's having P. Say that it is involved merely *indirectly* if one of the things that explains the event is an agent's *believing* something has P. It is not obvious that we need to invoke semantic properties *directly* in linguistic explanations at all. Prima facie, my uttering "Ada runs" is just as well explained by the fact that I believe the sentence true (plus my sincerity) as it is by the fact that they are true (plus my truthfulness). A second question for anyone thinking about the explanatory role of semantic properties is whether the role is direct or indirect.

The third question—and our focus here—is whether there's difference on either dimension between the explanatory role of sentential semantic properties like truth, and the explanatory role of subsentential reference.

3.1 Reference-invoking explanations.

Let's make the assumption that the only happenings immediately explained by semantic properties involve speech acts—either on the producer or consumer side (uttering a true sentence; or forming a true belief). We assert by uttering a sentence with assertoric force ("the door is shut''), question by uttering a sentence with interrogative force ("is the door shut?''), command by uttering a sentence with imperatival force ("shut the door''). In each case, you might well think we should cite the truth-conditions of the sentence in explaining why the speaker uttered the sentence or why their audience reacted as they did. Thus: Ada utters "the door is shut" inter alia because she wished to convey whether the door was shut, and "the door is shut" means that the door is shut.

Is reference involved in such explanations? Prima facie it is involved *mediately*—facts about reference explain facts about truth-conditions, which in turn explain other things.

On the assumption that the only locus for semantic explanations involve speech acts, and the *only* relevant speech acts are sentential, then (strangely) reference-involving explanations will always pass through a bottleneck: they only explain anything at all *via* grounding truth-conditional properties. Reference would be extraordinarily *narrow* in explanatory power—immediately explaining *only* the obtaining of other semantic properties.

Davidson's radical intepretation allows him to add a distinctive twist. After all, here we have T-sentences as the basic data, and the success of a T-theory is tested by how well it matches them. Does this mean that truth conditions explain reference, rather than vice versa? That might make reference utterly explanatorily idle. But Davidson's does not straightforwardly endorse this. He distinguishes 'the order of explanation that is appropriate once the theory is in place' from 'the explanation of why the theory is correct'. [32] In the second sense, truth has priority over reference, for the interpretationist. But that still leaves *another* sense of explanation in which reference explains truth. But *what* is this sense? To be sure, when learning a language, perhaps one proceeds from conjectures about truth conditions, to establishing a theory invoking reference, and then (later, with the theory in place) to derive novel predictions about truth. One has those final beliefs about truth only *because of* the beliefs about reference possessed at the intermediate phase. But here we have beliefs explaining beliefs; not reference explaining truth.

Davidson (1990, p.300) claims there is an *exact* analogy between appeal to reference in semantics, and appeal to unobservables in physical explanations. So perhaps his answer to the above would be to claim that reference explains truth in a sense of 'explanation' utterly familiar from the physical sciences. But *pace* Davidson, there are real differences between the cases. Posit atoms for one explanatory purpose (reducing thermodynamics) and one can generate new surprising predictions in a quite different area (brownian motion). Davidson says "there can be no question about the use of these theoretical concepts [reference and satisfaction] beyond the question of whether they yield a satisfactory account of the use of sentences" (ibid). But, I think, one should not say that "there can be no question of the use of concepts of atomic particles, beyond the question of whether they yield a satisfactory account of thermodynamic properties". Physical posits break free of the particular theories for which they are originally introduced. It's really not obvious whether reference does so, and this puts pressure on the claim that it is *explanatory* of truth-conditions in any strict sense.

A bottlenecked (at best) explanatory role for reference has consequences for our assessment of the tenability of radical inscrutability of reference. After all, why are revisionary philosophical theories to be regarded with suspicion in the first place? Not simply because they require us to change our beliefs (we might welcome the shock of the new). We are and should be wary of them because we want assurance that the revisions being proposed won't wreak havoc on our wider theory of the world—that the baby isn't being thrown away with the bathwater. A *confined* revisionism, where we have an assurance that we can change our views on one subject matter while keeping views on *other* subject matters untouched, would be far less worrying. If reference truly does have

[32](Davidson, 1990, p.300). In earlier writings, e.g. throughout (Davidson, 1984/1970), Davidson is happy to talk of *explaining* the truth-conditions of sentences.

(at best) the extraordinarily narrow explanatorily role outlined above, then it looks like we're in this position. What we think about reference is a "don't care" issue so far as total theory is concerned, so long as the appropriate truth-conditions are generated. And, of course, the twisted reference schemes that generate the inscrutability of reference do all agree (more or less) at the level of truth-conditions of whole sentences.

It is interesting that Field (1975)—even in a phase where he was an explicit advocate of the building block metaphysics of meaning that Davidson deprecated—explicitly endorsed the narrow, bottlenecked conception of the explanatory role of reference. He distanced himself from what he saw as Davidson's instrumentalism. But he did think that facts about reference were *conventional*, in the sense that there was no *deep* rationale for our focusing on the reference relation we do (which Field took to match up with certain causal relations) rather than some permuted variant of it. Someone who worked with what Field regarded as a determinately *different* subsentential semantic relation arising from a permutation of the original, would be able to explain all that needs to be explained equally well. He compared our focus on reference over its twisted cousins as comparable to the convention of driving on the right side of the road. It gets the job done—perhaps in a rather elegant way—but it's not the only way to do it.

An absent, or bottlenecked explanatory role for reference would make the inscrutability of reference liveable-with. But we should subject the starting points to scrutiny. Can semantic properties explain things other than speech acts? And must a putative explanatory role for reference in communication always go via truth-conditions? Fragmentary speech acts (discussed further below) might provide cases where we reference has an *immediate* explanatory role.

3.2 Explanations involving beliefs about reference?

Even the bottlenecked explanatory role for reference presupposed that truth or truth-conditions have direct explanatory work to do. But it is not obvious that the kind of explanatory contexts that have been discussed—communicative ones—require direct appeal to facts about truth or truth-conditions as opposed to *opinions* on the part of speakers and audience on the matter. Perhaps the role of semantics in explanations is always 'indirect' in this sense. Even if this is not the case, the explanatory profiles of reference vs truth, and beliefs about reference vs. beliefs about truth, may differ. Even if reference itself is explanatorily idle, in at least some cases there's a strong case that one *believes* about reference, explains what one believes about truth-conditions of whole sentences. Second language learning is a case in point. Some go further, and say that first-language linguistic competence is constituted, in part, by (perhaps knowledgable) beliefs about reference.[33]

The role of beliefs about reference in linguistic competence is of independent interest. And it is particular relevant to Davidson's project, since it makes a difference to the tenability of the inscrutability theses he endorses. If one believes p, and there's no fact of the matter about p, there's no fact of the matter about whether one's belief is

[33]Compare the slogan that understanding is knowing the meaning of one's words. Pettit (2002) contains an interesting discussion about whether *knowledge* is appropriate here.

correct. The fully informed and reflective will know this full well. It seems an uncomfortable—even Moore-paradoxical—state to be in.[34] But if beliefs about reference are constitutive of linguistic competence, and reference is inscrutable, then competence itself commits us to this uncomfortable situation.

Call the view that person-level beliefs about semantic properties are (in part) constitutive of competence the *cognitive* view of understanding. Cognitive views come in various strengths. A *weak* cognitive view says that only beliefs about the truth-conditions of sentences are required for linguistic understanding.[35] A *strong* view says that beliefs about the reference of names are required for understanding. The strong version needn't be ultra-strong; it needn't attribute knowledge of an entire semantic theory, or every axiom within it (which would be phenomenologically very implausible, if we're talking about person-level, consciously-accessible beliefs).

One case for a cognitive view of (first-language) competence is based on the *verbal rationality* of linguistic acts.[36] I utter 'Ada runs'. My action is an intentional one, not something that just 'happens to me'. Characteristically, then, we should expect practical reasoning explanations to be available, showing how my beliefs and desires interact so as to make the action 'the thing to do'. A natural suggestion is that this involves a wish to communicate that Ada runs, the belief I can communicate this by uttering something that means that Ada runs, plus the belief that 'Ada runs' means that Ada runs. A nice aspect of this story is that it can show why certain inappropriate speech acts are nevertheless rationally performed—I might say that 'Asa runs' when wanting to convey Ada runs, mistakenly believing that 'Asa runs' means that Ada runs. The action would be rationalized as normal, and we can pinpointed the faulty belief which makes it misfire.

Suppose we accept that person-level beliefs about truth-conditions are involved in first-person competence.[37] This doesn't immediately take us to a strong cognitive conception. To be sure, if one systematically believes the right things about the truth-conditions of sentences, this must be explained. But the mechanism providing the explanation might be causal, rather than person-level and inferential.[38]

[34]In fact, the issues here run deep, since it turns on what rational constraints believing p to be indeterminate place on one's attitudes to p. See (Williams, forthcoming) and the references therein.

[35]Such a view is proposed in Wright (1987). It's worth considering whether Davidson held a version of the weak theory. at (Davidson, 1974, p.142), he writes that any competent interpreter "a specifiable infinite subset of the truths of the theory", and seems to have in mind the T-theorems that connect most directly to patterns of usage. For a different reading of the passages, see (**?**, p.122)

[36]What follows is heavily influenced by the excellent (Heck, 2006). For deployment of this kind of consideration against Davidson, see especially (Rumfitt, 1995, §XI). Hornsby (2005) argues for a model of linguistic rationality that does not taken the form sketched below.

[37]Incidentally, one issue here is whether 'competence' is the right word to use, or whether we should really be thinking of this as a higher achievement—*mastery* of the language. Perhaps young children are competent with a language, in virtue of simple immersion, without their linguistic actions being fully verbally rational.

[38]For discussion, see Wright (1981); Evans (1981); Davies (1981) and Miller (1997). As an example of the 'subpersonal' strategy, consider the conjecture that human psychology involves a *semantic module*—a representation of semantic axioms of a T-theory, formulated in the 'language of thought', as discussed in (Larson & Segal, 1995). The module stands ready to process the canonical derivations, to output T-biconditionals (perhaps into consciouslly-accessible beliefs) for any given sentence. The representations of reference relations would not deserve the name 'beliefs' (as part of a mental module, they would be unavailable for general processing, and we would not have person-level introspective access to them).

It's not clear that the weak cognitive conception just sketched is stable. Consider again the patterns in our understanding of sentences. If I mistakenly think that 'Ada runs' means that Asa runs, I'll also think that 'Ada sits' means that Asa sits. But consider someone who didn't operate this way: who confidently believed the first while retaining the belief that 'Ada Gs' is true if and only Ada Gs for all other G. The weak cognitive conception would depict this as a mere processing error, and think there need be no rational criticism of the linguistic beliefs that constitute understanding. Consider also dictionary-learning of new words. A strong cognitive view of understanding has a straightforward account: one learns new facts about what words refer to, or what predicates are true of, which are directly relevant to one's understanding. The weak cognitive view depicts the process as strangely indirect: involving a post-hoc management of the subpersonal capacities that dispose one to acquire person-level beliefs about truth conditions.

If there are (verbally rational) subsentential speech acts, then the motivation for a strong cognitive conception may transfer directly. Stainton (1998) gives an example from which the following is adapted. We are at a round-table meeting, and I am describing seating arrangements. I do so by pointing at various chairs, while uttering, successively, "the boss", "anyone who needs to go early", "a representative from the faculty board", "Ada". The audience forms appropriate beliefs. Clearly I have managed to communicate propositional contents, but I have done so by uttering fragments of sentences. The motivations at this point for a strong cognitive conception can parallel those for the weak one. This sort of language use looks verbally rational. To secure verbal rationality, we need semantic beliefs to play a role in the practical reasoning explanation of the speech act. Suppose that I wrongly believe that the chairperson is called 'Ada'. My pointing to the chair and uttering "Ada" is a rational act, with a practical reasoning explanation routing through the (mistaken) belief that:

> 'Ada' refers to the Chairperson.

This argument turns crucially on the principle that my utterances are truly *subsentential* speech acts. There is an alternative diagnosis: even though a fragment is all that is phonetically realized, it may be that a sentence is uttered. The obvious parallel is to ellipsis. For example the material in square brackets in the following might not be pronounced:

> Ada was walking and Jean was [walking] too.

The ellipsis hypothesis (Stainton, 1998) is that the kind of speech acts we have been considering are elliptical for full sentences, and are *syntactically* sentential, that what is uttered is:

> 'Ada [sits here]'

Only a sentential semantic beliefs would be required to account for the verbal rationality

of the speech act. Whether the ellipsis hypothesis is correct is a contested, empirical question.[39]

A strong cognitive view of understanding, as mentioned, seems *prima facie* in tension with the view that reference is radically indeterminate. It's worth recalling, however, that Davidson's own articulation of the inscrutability of reference was rather nuanced: he did think there were correct things to think about appropriately *relativized* reference relations, and perhaps belief in *these* could subserve the verbal rationality of language use. I earlier suggested that he might have been better off embracing a more full-blooded instrumentalism.[40]

Such a position, on the face of it, would be in severe tension with cognitive views of understanding—linguistic competence would commit one to a range of semantic beliefs, to which the informed and reflective would on the contrary advocate an instrumentalist (perhaps error-theoretic) attitude. How could we square this circle? Perhaps the instrumentalist could appeal to some surrogate attitude: *as-if belief* or *pretence* rather than *belief* as the appropriate cognitive attitude to take to T-theories. Whether appeal to pretence can really provide practical reasoning explanations is another matter—a wish for a cup of tea; plus the pretence that the liquid in the toy cup is tea, does not rationalize taking a sip.

References

Davidson, Donald. 1965. 'Theories of meaning and learnable languages'. *In:* Bar-Hillel, Yehoshua (ed), *Proceedings of the 1964 inrternational congress for logic, methodology and philosophy of science*. Amsterdam: North Holland. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.3–16.

Davidson, Donald. 1967. 'Truth and meaning'. *Synthese*, **17**, 304–23. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.17–36.

[39]Stainton argues that the ellipsis hypothesis is false (Stainton, 1998, p.854). preferring the view that semantic information received from the linguistic fragment is processed *pragmatically* to determine a propositional content for the speech act as a whole. To give a flavour of the debate, consider the following. According to Stainton, paradigm examples of ellipsis do not occur in discourse-initial position, yet fragmentary speech acts can do so (Stainton's example is starting a conversation in a shop with"five red apples, please"). Stainton holds that ellipsis proper requires prior linguistic cues that allow the audience to 'fill in' the unpronounced material. The fragmentary utterances don't fit this description. For further discussion and debate, see Merchant (2005) and the references therein.

[40]I floated earlier the suggestiont that semantic properties had only indirect explanatory roles—via propositional attitudes in which they are embedded. This fits an anti-realist/instrumentalist position much better than does the view that they are themselves directly involved in explaining worldly happenings. (Compare Wright (1992) on 'wide cosmological role' as a test for realism). But there is a tension here: Davidson bundles together reference, truth and the content of attitudes to be determined by the radical interpreter. And the case for contentful beliefs and desires figuring in direct, immediate explanations of worldly happenings is much stronger (though compare Stich, 1991). What (Lewis, 1994) calls the *headfirst* approach to content, on which at least the coarse-grained content of beliefs and desires are settled independently of linguistic content, fits more nicely here.

Davidson, Donald. 1973. 'Radical interpretation'. *Dialectica*, **27**, 313–28. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.125–140.

Davidson, Donald. 1974. 'Belief and the basis of meaning'. *Synthese*, **27**, 309–23. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.141–154.

Davidson, Donald. 1975. 'Thought and talk'. *In:* Guttenplan, Samuel (ed), *Mind and language*. Oxford: Oxford University Press. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.155–170.

Davidson, Donald. 1976. 'Reply to Foster'. *In:* Evans, Gareth, & McDowell, John (eds), *Truth and meaning: Essays in semantics*. Oxford: Clarendon Press. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.171-80.

Davidson, Donald. 1977. 'Reality without reference'. *Dialectica*, **31**, 247–53. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.215–226.

Davidson, Donald. 1979. 'The inscrutability of reference'. *The southwestern journal of philosophy*, 7–19. Reprinted in Davidson, *Inquiries into Truth and Interpretation* (Oxford University Press, Oxford: 1980) pp.227–242.

Davidson, Donald. 1980. 'Towards a unified theory of thought, meaning and action'. *Grazer philosophische studien*, **11**, 1–12. Reprinted in Davidson, *Problems of Rationality* (Oxford University Press, Oxford: 2004) pp.151-166.

Davidson, Donald. 1984. *Inquiries into truth and interpretation*. Oxford: Oxford University Press.

Davidson, Donald. 1984/1970. Semantics for natural language. *In: Inquiries into truth and interpretation*. Oxford university press.

Davidson, Donald. 1990. The structure and content of truth. *The journal of philosophy*, **87**(6).

Davies, Martin. 1981. *Meaning, quantification, necessity*. London: Routledge and Kegan Paul.

Davies, Martin. 1987. Tacit knowledge and semantic theory. *Mind*, **96**(384), 441–462.

Evans, Gareth. 1975. 'Identity and predication'. *Journal of philosophy*, **LXXII**(13), 343–362. Reprinted in McDowell, (ed) *Gareth Evans: Collected Papers* (Clarendon Press, Oxford).

Evans, Gareth. 1981. 'Semantic theory and tacit knowledge'. *In:* Holtzman, & Leich (eds), *Wittgenstein: To follow a rule*. London: Routledge and Kegan Paul. Reprinted in McDowell, (ed) *Gareth Evans: Collected Papers* (Clarendon Press, Oxford) pp.322-42.

Field, Hartry H. 1972. 'Tarski's theory of truth'. *Journal of philosophy*, **69**, 347–375. Reprinted in Field, *Truth and the Absence of Fact* (Oxford University Press, 2001) pp. 3-29.

Field, Hartry H. 1975. 'Conventionalism and instrumentalism in semantics'. *Noûs*, **9**, 375–405.

Field, Hartry H. 1980. *Science without numbers: A defence of nominalism*. Library of philosophy and logic. Oxford: Blackwell.

Field, Hartry H. 1989. *Realism, mathematics and modality*. Oxford: Basil Blackwell.

Fine, Kit. 1975. 'Vagueness, truth and logic'. *Synthese*, **30**, 265–300. Reprinted with corrections in Keefe and Smith (eds) *Vagueness: A reader* (MIT Press, Cambridge MA: 1997) pp.119-150.

Fodor, Jerry. 1987. *Psychosemantics: The problem of meaning in the philosophy of mind*. Cambridge, MA: Bradford.

Foster, J. A. 1976. 'Meaning and truth theory'. *Pages 1–32 of:* Evans, G., & McDowell, J. (eds), *Truth and meaning: Essays in semantics*. Oxford: Clarendon Press.

Heck, Richard. 2006. 'Reason and language'. *In:* MacDonald, Cynthia (ed), *Mcdowell and his critics*.

Heim, Irene, & Kratzer, Angelika. 1998. *Semantics in generative grammar*. Oxford: Blackwell.

Hornsby, Jennifer. 2005. 'Semantic knowledge and practical knowledge: I'. *Proceedings of the aristotelian society: Supplementary volume*, **supp. LXVIII**, 107–130.

Keefe, Rosanna. 2000. *Theories of vagueness*. Cambridge: Cambridge University Press.

Kölbel, Max. 2001. 'Two dogmas of Davidsonian semantics'. *The journal of philosophy*, **XCVIII**(12).

Larson, Richard K., & Segal, Gabriel. 1995. *Knowledge of meaning*. Cambridge, MA: MIT Press.

Lepore, Ernest, & Ludwig, Kirk. 2005. *Donald davidson: Meaning, truth, language and reality*. Oxford: Oxford University Press.

Lepore, Ernest, & Ludwig, Kirk. 2007. *Donald davidson's truth-theoretic semantics*. Oxford university press.

Lewis, David K. 1969. *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Lewis, David K. 1970. 'General semantics'. *Synthese*, **22**, 18–67. Reprinted with postscript in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 189–229.

Lewis, David K. 1974. 'Radical interpretation'. *Synthese*, **23**, 331–44. Reprinted in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 108–18.

Lewis, David K. 1975. 'Language and languages'. *Pages 3–35 of: Minnesota studies in the philosophy of science*, vol. VII. University of Minnesota Press. Reprinted in Lewis, *Philosophical Papers I* (Oxford University Press, 1983) 163-88.

Lewis, David K. 1983. 'New work for a theory of universals'. *Australasian journal of philosophy*, **61**, 343–377. Reprinted in Lewis, *Papers on Metaphysics and Epistemology* (Cambridge University Press, 1999) 8–55.

Lewis, David K. 1984. 'Putnam's paradox'. *Australasian journal of philosophy*, **62**(3), 221–36. Reprinted in Lewis, *Papers on Metaphysics and Epistemology* (Cambridge University Press, 1999) 56–77.

Lewis, David K. 1992. 'Meaning without use: Reply to Hawthorne'. *Australasian journal of philosophy*, **70**, 106–110. Reprinted in Lewis, *Papers on Ethics and Social Philosophy* (Cambridge University Press, 1999) 145–151.

Lewis, David K. 1994. 'Reduction of mind'. *Pages 412–31 of:* Guttenplan, Samuel (ed), *A companion to the philosophy of mind*. Oxford: Blackwell. Reprinted in Lewis, *Papers on Metaphysics and Epistemology* (Cambridge University Press, 1999) 291–324.

Ludlow, Peter. 2011. *The philosophy of generative linguistics*. Oxford university press.

McDowell, John. 1978. 'Physicalism and primitive denotation: Field on Tarski'. *Erkenntnis*, **13**, 131–52. Reprinted in McDowell, *Meaning, Knowledge and Reality* (Harvard University Press, 1998) pp.132-156.

McGee, Vann. 2005. "Inscrutability and its discontents". *Noûs*, **39**, 397–425.

Merchant, Jason. 2005. "Fragments and ellipsis". *Linguistics and philosophy*, **27**, 661–738.

Miller, Alexander. 1997. 'Tacit knowledge'. *Pages 146–174 of:* Hale, Bob, & Wright, Crispin (eds), *A companion to the philosophy of language*. Oxford: Blackwell.

Pettit, Dean. 2002. 'Why knowledge is unnecessary for understanding language'. *Mind*, **111**, 519–550.

Putnam, Hilary. 1981. *Reason, truth and history*. Cambridge: Cambridge University Press.

Quine, W. V. 1964. 'Ontological reduction and the world of numbers'. *Journal of philosophy*, **61**. Reprinted with substantial changes in Quine, *The Ways of Paradox and Other Essays: Revised and enlarged edition* (Harvard University Press, Cambridge, MA and London, 1976) pp.212—220.

Rumfitt, Ian. 1995. 'Truth conditions and communication'. *Mind*, **104**, 827–862.

Stainton, Robert J. 1998. 'Quantifier phrases, meaningfulness "in isolation" and ellipsis'. *Linguistics and philosophy*, **21**, 846–865.

Stich, S.P. 1991. Narrow content meets fat syntax. *In:* Loewer, & Rey (eds), *Meaning in mind: Fodor and his critics*.

Wallace, J. 1977. 'Only in the context of a sentence do words have any meaning'. *In:* French, P.A., & T.E. Uehling, Jr. (eds), *Midwest studies in philosophy 2: Studies in the philosophy of language*. Morris: University of Minnesota Press.

Williams, J. Robert G. 2007. 'Eligibility and inscrutability'. *Philosophical review*, **116**(3), 361–399.

Williams, J. Robert G. 2008a. 'Permutations and Foster problems: two puzzles or one?'. *Ratio*, **XXI**(1).

Williams, J. Robert G. 2008b. 'The price of inscrutability'. *Nous*, **42**(4), 600–641.

Williams, J. Robert G. forthcoming. Indeterminacy and normative silence. *Analysis*.

Williams, J. Robert G. ms.. Lewis on reference.

Wright, Crispin. 1981. 'Rule-following, objectivity and the theory of meaning'. *Pages 99–117 of:* Leich, C., & Holtzman, S. (eds), *Wittgenstein: To follow a rule*. London: Routledge.

Wright, Crispin. 1987. 'Theories of meaning and speaker's knowledge'. *Pages 204–38 of: Realism, meaning and truth*. Oxford: Blackwell.

Wright, Crispin. 1992. *Truth and objectivity*. Cambridge, Mass.; London: Harvard University Press.