De Ray: On the Boundaries of the Davidsonian Semantic Programme

Cameron Domenico Kirk-Giannini
Rutgers University
camerondomenico.kirkgiannini@rutgers.edu

Ernie Lepore
Rutgers University
lepore@ruccs.rutgers.edu

Greg Ray (2014) believes he has discovered a crucial oversight in Donald Davidson's semantic programme, recognition of which paves the way for a novel approach to Davidsonian semantics. We disagree: Ray's novel approach involves a tacit appeal to pre-existing semantic knowledge which vitiates its interest as a development of the Davidsonian programme.

1. Introduction

Donald Davidson (1965, 1967, 1976) placed three constraints on an acceptable theory of meaning for a natural language *L*:[1]

(D1) It must show how the meanings of sentences of *L* depend on the meanings of their constituent parts.
(D2) Knowledge of it must suffice to put the theoretician in a position to interpret the utterances of speakers of *L*.[2]
(D3) It must be finitely axiomatizable, so that its construction demonstrates how finite agents like ourselves can understand a potential infinitude of sentences of *L*.

Given these constraints, the most natural candidate for a theory of meaning for *L* would be a compositional, finitely axiomatizable theory which issued directly in theorems of the form '*s* means in *L* that *p*', with '*s*' replaced by a structural description of an object-language sentence, '*L*' replaced by a name for the object language, and '*p*'

[1] Here we paraphrase Ray (2014 p. 83), who paraphrases various remarks of Davidson.

[2] We understand (D2) as requiring that knowledge of the propositions expressed by the sentences of the theory be sufficient to put any individual competent in (first-order) logic in a position to interpret the utterances of speakers of *L*.

Mind, Vol. 126 . 503 . July 2017
doi:10.1093/mind/fzv186 Advance Access publication 23 September 2016

replaced by a sentence in the language of the theory which translates the sentence described by '*s*'. We will refer to theorems of this form as *M-sentences* and to theories that derive them as *M-theories*. Davidson (1967, 1976) famously abandoned the project of developing an M-theory directly, instead constructing a theory that derived theorems of the form '*s* is true in *L* if and only if *p*' (with '*s*', '*L*' and '*p*' as before). We will refer to such theorems as *interpretative T-sentences* and to theories that deliver them as *interpretative T-theories*. According to Davidson (1976), a sentence of the form ⌜Some interpretative T-theory states that Θ⌝, where Θ is the conjunction of the axioms of an interpretative T-theory, would satisfy (D1)–(D3).[3] Davidson's strategy was thus one of indirection; his goal was to move from an interpretative T-theory for *L* to a complete set of M-sentences for *L*.[4]

Even if Davidson's proposal satisfies (D1)–(D3), it might be thought prolix. Greg Ray (2014) claims that Davidson's decision to pursue an indirect route to M-sentences was based on a restricted view of the theoretical alternatives available to him, offering a theory of meaning that he claims satisfies (D1)–(D3) without the circumlocutions of an interpretative T-theory. If Ray's project succeeds, it constitutes a significant development within the Davidsonian semantic programme. If it fails, on the other hand, its failure goes some way toward vindicating

[3] Following Quine (1940), we use ⌜…μ…⌝ so that it is synonymous with the expression '"'⌢…⌢ μ ⌢…'"'—that is, the expression obtained by concatenating an opening quotation mark with the expressions to the left of μ (whatever they may be, so long as they do not contain a variable ranging over expressions), the expression μ, the expressions to the right of μ (whatever they may be, so long as they do not contain a variable ranging over expressions), and, finally, a closing quotation mark. ⌜Some interpretative T-theory states that Θ⌝ thus denotes a sentence in which Θ is used, whereas, for example, ⌜Some interpretative T-theory contains an occurrence of 'Θ'⌝ denotes a sentence in which Θ is mentioned. (In cases involving corner quotes surrounding more than one variable ranging over expressions, this rule extends in the natural way: the corner-quoted expression is synonymous with the result of replacing these variables with their values, reproducing the remaining parts of the original expression (*sans* corner quotes), and enclosing the result in regular quotation marks. In cases of quotation within quotation, the rule must be modified to accommodate the introduction of additional quotation marks around doubly-quoted material in accordance with the typographical conventions of this publication.)

[4] As Lepore and Ludwig (2005) point out, Davidson's (1976) proposal is not wholly adequate, because it does not provide the theoretician with the resources to distinguish interpretative T-sentences from other non-interpretative theorems. For example, Davidson's proposal allows the derivation of '"Snow is white" is true in English iff snow is white and $(2 + 2 = 4$ or $2 + 2 \neq 4)$', in which the sentence on the right of the biconditional is an obviously incorrect interpretation of the sentence mentioned on its left. We follow Lepore and Ludwig in charitably amending Davidson's proposal so that it is not subject to this problem (see §7 for a full statement of the modified proposal).

Davidson's conception of his own theoretical alternatives. In what follows, we develop what we take to be a decisive objection to Ray's theory: that it tacitly appeals to pre-existing semantic knowledge in violation of (D2).

Our discussion is structured as follows. §§2 and 3 present a (prima facie) promising semantic theory and then uncover its illicit appeal to pre-existing semantic knowledge, revealing that it is not in the spirit of Davidson's programme. §§4 and 5 consider and reject Ray's proposal, arguing that it differs only superficially from the theory presented in the preceding two sections. §6 responds to Ray's attempts to deflect criticism of his theory. Finally, §7 addresses the concern that Davidson's own proposal might be vulnerable to our argument against Ray.

2. An M-Theory on the cheap?

Devising an M-theory which satisfies (D1)–(D3) might seem to be a relatively simple exercise. For example, it might seem possible to start with a translational theory in the spirit of Katz and Postal (1964)—that is, a theory which issues in theorems stating synonymies between object language and metalanguage sentences—and add to it an axiom licensing the derivation of M-sentences. In this section, we devise such a translational theory and show that it fails to satisfy (D2).

The bulk of the effort in developing a translational theory consists in the construction of a function mapping object language sentences to synonymous metalanguage sentences. Once this function (call it $\mathbb{F}$) has been constructed, it is not difficult to generate theorems like:

(TS) 'Helmwige vole' in French translates into English as 'Helmwige flies'.

Taking a regimented version of French containing two names, two predicates, and a sentential connective as the object language ($\mathbb{O}$) and English as the metalanguage ($\mathbb{M}$), and assigning a syntactic type to each lexical item (type $\mathscr{N}$, $\mathscr{P}$, or $\mathscr{C}$) and sentence (type $\mathscr{S}$) of $\mathbb{O}$, we can define $\mathbb{F} : \mathbb{O} \rightarrow \mathbb{M}$ as follows:

$\mathbb{F}$('Helmwige') = 'Helmwige' (type $\mathscr{N}$)
$\mathbb{F}$('Grimgerde') = 'Grimgerde' (type $\mathscr{N}$)
$\mathbb{F}$('reste') = 'stays' (type $\mathscr{P}$)
$\mathbb{F}$('vole') = 'flies' (type $\mathscr{P}$)
$\mathbb{F}$('et') = 'and' (type $\mathscr{C}$)
For any $\mathscr{N}$, $\mathscr{P}$, $\mathbb{O}$-sentence $\mathscr{N} \frown \mathscr{P}$: $\mathbb{F}(\mathscr{N} \frown \mathscr{P}) = \mathbb{F}(\mathscr{N}) \frown \mathbb{F}(\mathscr{P})$.

For any $\mathscr{S}_1$, $\mathscr{C}$, $\mathscr{S}_2$, $\mathbb{O}$-sentence $\mathscr{S}_1 \frown \mathscr{C} \frown \mathscr{S}_2$: $\mathbb{F}(\mathscr{S}_1 \frown \mathscr{C} \frown \mathscr{S}_2) = \mathbb{F}(\mathscr{S}_1) \frown \mathbb{F}(\mathscr{C}) \frown \mathbb{F}(\mathscr{S}_2)$.

$\mathbb{F}$, together with (TA) below, allows us to derive (TS):

Translation Axiom (TA): For any French sentence ϕ and English sentence ψ, ⌜'ϕ' in French translates into English as 'ψ'⌝ is true in English iff $\mathbb{F}(\phi) = \psi$.

Alas, knowledge of this theory is not sufficient for understanding its object language. For there are not two languages involved in the presentation of the theory, but *three*: an object language and two distinguishable metalanguages (an upper one in which the theory is stated and a lower one into which object-language sentences are translated).[5,6] Anyone who understood the upper but not the lower metalanguage could know everything stated by the theory without understanding any French. Were the three languages transparently distinguished—for example, were German the lower metalanguage and English the upper metalanguage—the derived theorems would be useless to a monolingual English speaker attempting to interpret French, as witnessed by (TS*):

(TS*) 'Helmwige vole' in French translates into German as 'Helmwige fliegt'.

An interpreter who knew neither French nor German could know what is expressed by (TS*) without understanding 'Helmwige vole'. So the theory fails to satisfy (D2). All this is, of course, old news.[7] But it might seem possible to use $\mathbb{F}$ to generate theorems, knowledge of which *would* suffice to interpret French. Suppose we replace (TA) with (MA):

Meaning Axiom (MA): For any French sentence ϕ and English sentence ψ, the sentence ⌜'ϕ' means in French that ψ⌝ is true in English iff $\mathbb{F}(\phi) = \psi$.

[5] In fact, if the language in which (TA) is stated is allowed to differ from the language of the corner-quoted material it contains, there are four languages in play. Given that our arguments can be made with reference to only three languages, however, we will henceforth ignore the fourth as a matter of convenience.

[6] Some readers may be inclined to think of translational semantic theories as having two object languages rather than two metalanguages. We have preferred to speak of two metalanguages in order to make the similarities between Ray's theory and standard translational theories as explicit as possible.

[7] See, for example, Vermazen (1967), Lewis (1970), and Lepore and Loewer (1981).

Using (MA), we can derive an M-sentence for 'Helmwige vole' along the following lines:

(M1) '"Helmwige vole" means in French that Helmwige flies' is true in English iff $\mathbb{F}$('Helmwige vole') = 'Helmwige flies'. [Instance of MA]
(M2) $\mathbb{F}$('Helmwige vole') = 'Helmwige flies'. [Premiss]
(M3) '"Helmwige vole" means in French that Helmwige flies' is true in English. [From M1, M2]
(M4) If '"Helmwige vole" means in French that Helmwige flies' is true in English, then 'Helmwige vole' means in French that Helmwige flies. [Premiss]
(MS) 'Helmwige vole' means in French that Helmwige flies. [From M3, M4]

It would seem, then, that we can deduce every true M-sentence for our fragment of French using a theory significantly simpler than an interpretative T-theory. This simpler theory, which we will call the *modified translational theory* (in contradistinction to the *simple translational theory* from which it was derived), appears to satisfy (D2). Because $\mathbb{F}$ is defined recursively, it also satisfies (D3). And, though admittedly $\mathbb{F}$ could be defined so as not to respect the intuitive boundaries of semantic constituents,[8] if it is not, the resulting theory satisfies (D1).

3. The translation problem

Unfortunately, any advantage secured by employing (MA) rather than (TA) is illusory. Because (D2) requires that knowledge of the propositions expressed by the theory (which is written in the upper metalanguage) be sufficient to put the theoretician in a position to interpret the utterances of speakers of the object language, and since the propositions expressed by a theory are invariant under translation of that theory into other languages, any theory satisfying (D2) must also satisfy it when translated into an arbitrary target language. This fact provides us with a convenient test: if translating the upper metalanguage of a theory into another language results in a theory that does not satisfy (D2), then the original theory also fails to satisfy (D2). Let us apply this test to our modified translational theory, translating its upper metalanguage from English into German while leaving French

[8] We can imagine, for example, defining $\mathbb{F}$ so that $\mathbb{F}$ ('Grimgerde reste et Helmwige vole') = $\mathbb{F}$ ('Grimgerde') $\frown$ $\mathbb{F}$ ('reste et Helmwige') $\frown$ $\mathbb{F}$ ('vole'), with 'reste et Helmwige' serving as an illicit semantic primitive.

as its object language and English as its lower metalanguage. Doing so yields (M3*) rather than (M3):

(M3*) '"Helmwige vole" means in French that Helmwige flies' ist im Englischen wahr.

The disquotational premiss (M4*) corresponding to (M4) is:

(M4*) Wenn '"Helmwige vole" means in French that Helmwige flies' im Englischen wahr ist, dann 'Helmwige vole' means in French that Helmwige flies.

But (M4*) is not well-formed in any natural language. What is needed is something that does *not* follow from (M3*), namely:

(M4**) Wenn '"Helmwige vole" means in French that Helmwige flies' im Englischen wahr ist, dann bedeutet 'Helmwige vole' im Französischen, dass Helmwige fliegt.

The *translation problem* is constituted by that fact that no one could construct (M4**) from (M3*) without a prior understanding of the theory's lower metalanguage, English. In other words, our modified translational theory for the object language *presupposes that the theoretician possesses a distinct M-theory for the lower metalanguage*, and this presupposition violates Davidson's (D2).[9]

The superficial difference between our modified translational theory and the simple translational theory of §2 is that in derivations involving instances of (MA), object-language expressions appear inside quoted lower metalanguage expressions instead of beside them. This embedding, combined with the assumption that the lower metalanguage is a fragment of the upper metalanguage, serves to obscure the fact that the modified translational theory, like any translational theory, essentially consists in a syntactic mapping from the object language to the lower metalanguage. For this reason, the difference between the two theories, which originates in the difference between (TA) and (MA), is irrelevant to the question of whether they satisfy Davidson's constraint.

We take it as given that the simple translational theory does not satisfy (D2). It is even more damning for our modified translational theory, then, to note that if we presuppose that the theoretician is competent in the lower metalanguage, we can derive (MS) from the

[9] Speaks (2006) makes this point in the context of his criticism of another proposed revision of Davidson's semantic programme (Klbel 2001). Speaks, however, argues that *no* Davidsonian theory satisfies (D2). We disagree for reasons which will be discussed in §7.

simple translational theory. For if the theoretician is antecedently competent in the lower metalanguage, nothing prevents her from employing the following disquotational premiss, allowing her to derive (MS) from (TS) by *modus ponens*.

(DISQ) If '"Helmwige vole" in French translates into English as "Helmwige flies"' is true in English, then 'Helmwige vole' means in French that Helmwige flies.

4. Ray's proposal

In the previous two sections, we considered two translational theories of meaning and found that both fail to satisfy the crucial Davidsonian constraint (D2). We are now in a position to introduce and evaluate Ray's proposal. In this section and the next, we argue that Ray's proposal does not differ from the modified translational theory in any respect relevant to its satisfaction of (D1)–(D3). Ray's theory is consequently subject to the translation problem, indicating that, like the translational theories, it fails to satisfy (D2).

Ray's central insight is that it is possible to design a theory that generates theorems of the form

'*s* means in *L* that *p*' is true

(we follow him in calling these *MnT-sentences*) rather than M-sentences or interpretative T-sentences. His hope is to achieve what Davidson thought impossible: using only standard first-order logic, to produce a theory which 'will literally give us means-that information' (Ray 2014, p. 87). His theory comprises (i) a set of axioms—basic and recursive—from which an MnT-sentence for every sentence in the object language can be derived, and (ii) a method for deriving an M-sentence from its corresponding MnT-sentence.

Ray begins with a finite set of meaning postulates for proper names and predicates. For predicates, there are postulates of the form:

(PRED) For all i, ⌜Of any v_i, 'x_i est rouge' means [in French] that v_i is red⌝ is true [in English].

For proper names, there are postulates of the form:

(NOM) 'Jacques' means [in French] Jacques.

MnT-sentences for atomic formulae are then constructed by (i) pairing each atomic formula $\phi(\alpha)$ in the object language with an atomic formula $\Psi(A)$ in the lower metalanguage such that ⌜'α' means [in

French] A⌝ is true in English and, for all i, ⌜Of any v_i, '$\phi(x_i)$' means [in French] that $\Psi(v_i)$⌝ is true in English, and (ii) slotting the pairs thus obtained into the schema for MnT-sentences, yielding ⌜'"$\phi(\alpha)$" means [in French] that $\Psi(A)$' is true in English⌝.[10]

Once MnT-sentences for atomic formulae have been constructed, the generation of MnT-sentences for complex formulae proceeds straightforwardly. For example, Ray gives us the following recursive axiom for conjunction (modified so that French is the object language):

(CONJ) For all ϕ, ψ, Θ, ⌜'ϕ [et] ψ' means [in French] that Θ⌝ is true [in English] iff there are Φ, Ψ such that

(4) Θ is ⌜Φ and Ψ⌝, and
(5) ⌜'ϕ' means [in French] that Φ⌝ is true [in English], and ⌜'ψ' means [in French] that Ψ⌝ is true [in English].

Ray suggests that a theoretician in possession of this recipe for MnT-sentences could reason as follows:

(Mn1) '"Helmwige vole" means in French that Helmwige flies' is true in English. [From Ray's theory]
(Mn2) If '"Helmwige vole" means in French that Helmwige flies' is true in English, then 'Helmwige vole' means in French that Helmwige flies. [Premiss]
(MS) 'Helmwige vole' means in French that Helmwige flies. [From Mn1, Mn2]

If this proposal is cogent, it demonstrates how to construct a simple, finitely axiomatizable theory which has among its consequences all the true M-sentences for its object language and which appeals only to standard first-order logic.

5. Ray's proposal evaluated

Ray's proposal relies on the same deductive strategy for moving from MnT-sentences to M-sentences as the two translational theories already discussed; in particular, it relies on the disquotational premiss:

(Mn2 = M4)[11] If '"Helmwige vole" means in French that Helmwige flies' is true in English, then 'Helmwige vole' means in French that Helmwige flies.

[10] Some inessential details have been left out of this treatment of Ray's theory.

[11] We use this notation throughout to draw attention to the fact that multiple arguments rely on the same crucial premiss.

The argument of §3 can thus be reapplied to Ray's proposal without revision. This alone establishes the falsity of Ray's crucial claim (2014, p. 88) that (Mn2) expresses something that anyone who knows the metalanguage is in a position to know—unless that claim is interpreted on the assumption that the lower and upper metalanguages are identical.[12]

This result may be surprising; Ray's theory certainly *seems* different from the modified translational theory of §§2 and 3. Is its problematic reliance on a disquotation principle eliminable? In the remainder of this section, we motivate a negative answer to this question by considering the respects in which Ray's proposal differs from the modified translational theory and arguing that none of them constitute relevant advances over it.

Ray's proposal differs from the modified translational theory in two respects: first, it contains different axioms for proper names and predicates; second, it derives complex MnT-sentences from atomic MnT-sentences using recursive axioms like (CONJ) rather than generating both complex and atomic M-sentences wholesale using a translation function and a meaning axiom. We address each dissimilarity in turn.

Whereas the modified translational theory contains axioms of the form '$\mathbb{F}(\phi) = \psi$' for both predicates and proper names, Ray's theory has meaning postulates like (PRED) for predicates and (NOM) for proper names. Perhaps Ray's theory is distinguished from the modified translational theory of §2 by the ubiquity of 'means' in these postulates and the fact that expressions of the lower metalanguage appear in some of them surrounded by one fewer set of quotation marks than expressions of the object language.

In fact, replacing both 'means' and 'means that' with 'translates into', and enclosing lower metalanguage expressions in an extra set of quotation marks in all of Ray's axioms for predicates and proper names, creates no barrier to the construction of MnT-sentences for atomic formulae; we can revise part (i) of Ray's construction procedure to pair each atomic formula $\phi(\alpha)$ in the object language with an atomic formula $\Phi(A)$ in the metalanguage such that ⌜'α' in French translates into 'A' in English⌝ is true in English, and for all i, ⌜Of any v_i, '$\phi(x_i)$' in French translates into '$\Phi(v_i)$' in English⌝ is true in English. This demonstrates that Ray's choice to use 'means' rather than 'translates into' in his axioms plays no essential role in his theory. Given that the function from object-language expressions to

[12] We address the status of this assumption further in §6.

lower metalanguage expressions employed by our simple translational theory is translational, moreover, we regard the difference between axioms of the form $\ulcorner$'ϕ' in French translates into 'Φ' in English$\urcorner$ and axioms of the form $\ulcorner\mathbb{F}(\phi) = \psi\urcorner$ as immaterial. So we have yet to discover a relevant difference between Ray's theory and the modified translational theory of §2.

Whereas the modified translational theory generates both atomic and complex M-sentences wholesale using a translation function and the meaning axiom (MA), Ray's theory derives complex MnT-sentences from atomic MnT-sentences using recursive axioms like (CONJ). Perhaps this is where Ray's theory differs importantly from the modified translational theory. But no: because the translation function $\mathbb{F}$ is translational, it is the case for all ϕ, Φ that $\mathbb{F}(\phi) = \Phi$ iff $\ulcorner$'ϕ' means in French that $\Phi\urcorner$ is true in English. (CONJ) is thus just a less general version of the second recursive clause in the definition of $\mathbb{F}$ presented in §2. We conclude that Ray's proposal is not a relevant advance over the modified translational theory developed above.

6. Ray on Ray

Ray anticipates some of the general themes of the criticisms advanced above. In each case, however, he crucially underestimates the force of the considerations against his theory. Two of his comments require special discussion.

Ray dismisses comparisons of his theory to translational theories:

> [Such comparisons are] a mistake in the first instance because the meaning theory outlined is unlike a translation manual—to miss this is to fail to appreciate the difference in use–mention level as between the object language and metalanguage deployed in the theory. A translation manual has two object languages that it states a relation between the sentences of, but our theory has but one target language. So the charge misclassifies our theory. (Ray 2014, p. 98)

We wish to register four points in connection with this passage. First, Ray's appeal to the use–mention distinction works against him, since Ray's theory, like a translational theory, contains uninterpreted mentioned expressions not only of the object language but also of the lower metalanguage. This feature, which gives rise to the translation problem, makes it more philosophically revealing to group Ray's theory with the translational theory than to maintain that labels like 'object language' and 'target language' track some deep dissimilarity.

Second, Ray's distinction between theories with 'two object languages' and theories with 'but one target language' seems to commit him to regarding the simple and modified translational theories from §2 as distinct in some philosophically significant respect. For reasons we have already articulated, we believe that no such distinction can be maintained. Third, if Ray is correct in asserting that his theory does not '[state] a relation between the sentences of [two languages]', that is because it does less, not more: his theory tells us only which MnT-sentences are true, and nothing about either synonymy or meaning. Finally, our criticism of Ray's theory is that it resembles a translational theory *in one philosophically important respect*: it too is vulnerable to the translation problem. We are not committed to there being any other similarities between Ray's theory and translational theories; in particular, we are not committed to any claim about how many 'target languages' Ray's theory has.

This brings us to the central point of controversy between us and Ray: the translation problem and the significance of the tacit presupposition of competence in the lower metalanguage that it reveals. On this subject, Ray writes:

> An interesting [objection to translational theories that might apply to my proposal] is the charge that a translational theory relies on something crucial that is does not state—namely, that the theorizer understands what in our case would be the lower metalanguage … But the motivating charge founders. It is granted that the interpreter understands the metalanguage of her meaning theory, but Davidson objects when a sentence of that metalanguage is mentioned in the (translational) theory—as though its *mention* implies that it must be a sentence of some *other* language than the metatheory. But this is wrong. To speak of the truth of English sentences is still to speak English and not some other (meta-)language. Davidson's charge seems rooted in a naïve belief in the absoluteness of the object/meta-language distinction as we see it in the abandoned language-levels theory of Russell's theory of types. (Ray 2014, p. 98)

Here Ray does not appreciate the severity of the translation problem. The problem is *not* that it is impossible to identify the lower and upper metalanguages. Making the point that this *is* possible does nothing to defuse the problem, which is that even when the languages are identical, translation of the theory from the upper metalanguage into another language reveals that the theory can only be used to interpret an object language in the presence of further knowledge not stated by the theory. The fact that when the upper and lower metalanguages are identified the theoretician can be expected to have this extra knowledge does nothing to mitigate the force of the objection.

7. Is Davidson's theory subject to the translation problem?

So far, we have seen that Ray's proposal fails to satisfy Davidson's constraint (D2) on an acceptable theory of meaning. Of course, this result has little dialectical force if traditional Davidsonian theories of meaning are subject to the same problem. In the remainder of this discussion, we will argue that they are not, and, in the context of an imagined response by the proponent of Ray's theory, that there is no simple modification of his proposal that would result in it satisfying all of (D1)–(D3). Our argument for this last claim will reveal what we take to be one of the central insights of Davidsonian semantics.

According to the Davidsonian proposal (at least as developed by Lepore and Ludwig 2005, pp. 120–1),[13] the knowledge sufficient for interpreting an object language is (i) knowledge that some particular truth theory for that language (considered as a set of sentences)[14] is interpretative, (ii) knowledge of the meanings of that theory's axioms (considered as sentences), and (iii) knowledge of a canonical proof procedure for the language of the truth theory. Here, for example, are the axioms for an interpretative truth theory for our simple fragment of French.

(R1) The referent of 'Helmwige' = Helmwige.
(R2) The referent of 'Grimgerde' = Grimgerde.
(B1) For all names α, $\ulcorner \alpha$ vole$\urcorner$ is true iff the referent of α flies.
(B2) For all names α, $\ulcorner \alpha$ reste$\urcorner$ is true iff the referent of α stays.
(RC1) For all sentences ϕ, ψ, $\ulcorner \phi$ et $\psi \urcorner$ is true iff ϕ is true and ψ is true.

The theory is presented in English, but of course this choice is arbitrary; we stipulate that the theoretician know what each of the axioms (considered as sentences) means. For example, she knows that (R1) means that the referent of 'Helmwige' = Helmwige, and so on for the

[13] Everything we say here about the Lepore and Ludwig proposal is, we believe, in the spirit of Davidson's proposal (1976).

[14] This and subsequent related qualifications are necessary to avoid the objection, due to Foster (1976, p. 19), that in certain puzzle cases it is possible for an individual to know the meanings of the axioms of a truth theory (as stated in one language) and that the theory (as stated in a distinct language) is interpretative, without being able to interpret the utterances of speakers of the theory's object language.

other four axioms. This stipulation is perfectly legitimate, because the Davidsonian project is to explain how a speaker could understand what every expression of the language means *given that she already knows what each primitive expression means*: each axiom corresponds to the interpretation of a primitive expression.

With a small number of inference rules (call the set of these rules $\mathbb{R}$),[15] we can, following Lepore and Ludwig (2007), define a canonical proof as follows.

> A canonical proof is a finite sequence of sentences of our theory's lower metalanguage such that:
> (a) its last member is a T-sentence containing no semantic vocabulary introduced by the theory on its right-hand side; and
> (b) each of its members is either (i) an axiom, or (ii) derived from earlier members by one of the inference rules in $\mathbb{R}$.

Given a canonical proof of some T-sentence, the interpreter can derive the corresponding M-sentence along the following lines (an enthymematic version of this type of argument can be found in Ray 2014, p. 85):

> (DA) For all ρ and ϕ, if ⌜'ρ' is true in French iff ϕ⌝ in English is a canonical T-form theorem of interpretative T-theory Φ, then ⌜'ρ' means in French that ϕ⌝ is true in English.[16] [Premiss]
> (P1) '"Helmwige vole" is true in French iff Helmwige flies' in English is a canonical T-form theorem of interpretative T-theory Φ. [Premiss]
> (P2) If '"Helmwige vole" is true in French iff Helmwige flies' in English is a canonical T-form theorem of interpretative T-theory Φ, then '"Helmwige vole" means in French that Helmwige flies' is true in English. [Instance of DA]
> (P3) '"Helmwige vole" means in French that Helmwige flies' is true in English. [From P1, P2]
> (P4) If '"Helmwige vole" means in French that Helmwige flies' is true in English, then 'Helmwige vole' means in French that Helmwige flies. [Premiss]

[15] Which inference rules should be included in $\mathbb{R}$ depends on properties of the object language. For a detailed discussion of the rules required to define a canonical proof for an object language containing only proper names, predicates and truth-functional connectives, see Lepore and Ludwig (2007), in particular, the definition of a canonical proof (p. 36) and the example of a canonical derivation of a T-sentence (pp. 32–3).

[16] We assume that the interpreter is competent with 'means that' constructions in the lower metalanguage.

(MS) 'Helmwige vole' means in French that Helmwige flies. [From P3, P4]

Since Davidson's theory, like Ray's, utilizes premisses like (P4 = Mn2 = M4) in deriving its M-sentences, one might wonder whether it is vulnerable to the same criticism as Ray's. We argue that it is not; the semantic competence with the lower metalanguage that Davidson presupposes is benign.

The Davidsonian proposal differs from Ray's in that any interpreter who knows (i)–(iii) above will be in possession of knowledge analogous to that expressed by (P4), for she can be expected to know the meaning of the quoted material in (P1), having derived it from the axioms of the theory. Knowledge that a given T-theorem *t* is the product of a canonical derivation from axioms $\phi_1, \ldots, \phi_n$, combined with knowledge of what these axioms mean, plausibly secures knowledge of the meaning of *t*. This follows from the principle below, which we endorse:

(DERIV) For all persons *p* and sentences $\phi_1, \ldots, \phi_n$: if (i) *p* knows, for each of $\phi_1, \ldots, \phi_n$, what it means, and (ii) *p* knows that ψ, considered as a sentence, follows from $\{\phi_1, \ldots, \phi_n\}$ by a canonical derivation, then *p* is in a position to know what ψ means.

Thus any interpreter who knows both the theoretical and meta-theoretical propositions to which Davidson appeals has the competence to interpret the object language; in other words, Davidsonian semantic theories satisfy (D2).[17] So the imagined *tu quoque* objection to Davidson's programme fails.

It might be objected at this point that the Davidsonian theory escapes the translation problem only because the Davidsonian has stipulated that the theoretician know the meanings of the axioms of an interpretative T-theory for the object language. Why couldn't Ray or a translation theorist escape the translation problem by means of a similar stipulation? The answer to this question reveals the fundamental difficulty with Ray's approach.

Suppose we attempt to specify those lower metalanguage expressions the meanings of which the theoretician must be required

[17] This is our main point of disagreement with Speaks (2006), who argues that the Davidsonian proposal must presuppose competence with the entire lower metalanguage and therefore fails to satisfy (D2). On our Davidsonian proposal, what is presupposed is knowledge of the meanings of a finite number of sentences of the lower metalanguage; (DERIV) then guarantees that the interpreter has semantic knowledge sufficient to deduce the true M-sentences for the object language.

to know for Ray's theory to escape the translation problem. Suppose further that we charitably allow Ray's theoretician to know the meaning of the lower metalanguage M-sentence corresponding to every atomic formula in the object language. Consider Ray's rule for conjunction (modified so that German is the lower metalanguage):

(GCONJ) For all ϕ, ψ, Θ, $\ulcorner$'ϕ et ψ' bedeutet im Französischen, dass $\Theta\urcorner$ is true in German iff there are Φ, Ψ such that

(1) Θ is $\ulcorner\Phi$ und $\Psi\urcorner$, and
(2) $\ulcorner$'ϕ' bedeutet im Französischen, dass $\Phi\urcorner$ is true in German, and $\ulcorner$'ψ' bedeutet im Französischen, dass $\Psi\urcorner$ is true in German.

For concreteness, let ϕ = 'Helmwige vole', ψ = 'Grimgerde reste', Φ = 'Helmwige fliegt', and Ψ = 'Grimgerde bleibt'. Then by stipulation, the theoretician knows that '"Helmwige vole" bedeutet im Französischen, dass Helmwige fliegt' means in German that 'Helmwige vole' means in French that Helmwige flies and '"Grimgerde reste" bedeutet im Französischen, dass Grimgerde bleibt' means in German that 'Grimgerde reste' means in French that Grimgerde stays.

But what about '"Helmwige vole et Grimgerde reste" bedeutet im Französischen, dass Helmwige fliegt und Grimgerde bleibt' (call this sentence *S*)? Nothing our theoretician knows rules out an interpretation of the lower metalanguage on which *S* means in German that 'Helmwige vole et Grimgerde reste' means in French that Helmwige flies *or* Grimgerde stays; or that Helmwige flies and Grimgerde stays and arithmetic is incomplete; or that Helmwige flies and Grimgerde stays and Jebediah juggles; or indeed *anything at all*. The only constraint on the meaning of *S* is that Θ must contain expressions which figure in true MnT-sentences for 'Helmwige vole' and 'Grimgerde reste'. But this is consistent with *any* theory about the meaning of the expression derived by embedding these expressions in the context created by 'und'.

So knowledge of (GCONJ), combined with knowledge of the meanings of the M-sentences for 'Helmwige vole' and 'Grimgerde reste', does not position the theoretician to know the crucial information necessary to generate the disquotational premiss which will allow her to move from *S* to the corresponding M-sentence: the knowledge that *S* means in German that 'Helmwige vole et Grimgerde reste' means in French that Helmwige flies and Grimgerde stays.

How might Ray attempt to solve this problem? It will not do to say that the interpreter needs to know that

(K1) 'Und' in German translates into English as 'and'.

To make this move is to adopt a translational theory, and Ray agrees with us that translational theories do not satisfy (D2). It will not do to say that the interpreter needs to know that

(K2) 'Und' means and in German,

because this axiom would not tell her how to derive the meanings of complex German sentences containing 'und' from simple ones.[18] It will not do to try to solve the problem just mentioned by constructing an axiom analogous to the axiom for conjunction in a Tarskian truth theory for English,

(TCONJ) For all ϕ, ψ: $\ulcorner \phi$ und $\psi \urcorner$ is true in German iff ϕ is true in German and ψ is true in German,[19]

for this would result in

(GCONJ*) For all ϕ, ψ, Φ, Ψ: $\ulcorner \phi$ und $\psi \urcorner$ means in German that Φ and Ψ iff ϕ means in German that Φ and ψ means in German that Ψ,

which illicitly quantifies into contexts where expressions are used rather than mentioned. (In order to avoid this problem, we would have to retreat to something like (GCONJ), leaving us exactly where we began.) It will not do, finally, to say that the interpreter needs to know what every German sentence containing 'und' means; the number of such sentences is infinite, and requiring knowledge of the meanings of them all violates Davidson's finite axiomatizability constraint (D3).

In response to this problem, someone sympathetic to Ray's project may search for a finite set of axioms which jointly entail the upper metalanguage M-sentence corresponding to every conjunctive lower metalanguage sentence. We wish this friend of Ray's proposal luck; hers is the thorny problem of axiomatizing an intensional theory of

[18] One might also worry that the construction 'means and in German' is not well-formed. If 'and' is playing any grammatical role in (K2), one might argue, it is the role of a coordinating conjunction, and it is obligatory in English for a coordinating conjunction to join two independent clauses.

[19] The clauses coordinated by 'und' in (GCONJ) are dependent, whereas truth is most naturally predicated of independent clauses, which have a distinct word order in German. Our formulation of (TCONJ) ignores this complication, which is not of philosophical importance in this context.

meaning, which originally drove Davidson towards the more tractable logic of truth theories.[20]

These are, so far as we can see, all of Ray's options. So his theory (even supplemented by an arbitrarily large finite number of meaning postulates) isn't such that knowledge of it would confer the ability to interpret the object language. *There is no way for Ray to satisfy both (D2) and (D3)—he must pick one to violate.* This is the fundamental difficulty with Ray's approach.

8. Conclusion

We conclude that Ray has not provided a viable alternative to traditional Davidsonian semantic theories: because his proposal involves a violation of (D2), it stands squarely outside the boundaries of Davidson's semantic programme. Davidson's own proposal, as articulated by Lepore and Ludwig (2005, 2007), remains the only theory with a legitimate claim to satisfy all of (D1)–(D3). What Ray takes to be a demonstration that 'truth theory has no *fundamental* role to play in meaning theory' (2014, p. 80; emphasis in original) thus turns out to be the opposite; the failure of his proposal lends credibility to the claim that Davidson's truth-theoretic route to a theory of meaning is the consequence of necessity rather than mere theoretical oversight.[21]

References

Bar-Hillel, Yehoshua (ed.) 1965: *Proceedings of the 1964 International Congress for Logic, Methodology and Philosophy of Science.* Amsterdam: North-Holland.

Davidson, Donald 1965: 'Theories of Meaning and Learnable Languages'. In Davidson 1984, pp. 3–15. Originally published in Bar-Hillel 1965.

——1967: 'Truth and Meaning'. In Davidson 1984, pp. 17–36. Originally published in Synthese, 17, pp. 304–23.

[20] It is also worth noting that, if it is possible to devise such a meaning theory for the lower metalanguage, we have little reason to doubt the possibility of devising one for the object language directly. This possibility would, of course, render Ray's theory superfluous.

[21] The authors would like to thank Simon Goldstein, Andy Yu, two anonymous referees for *Mind*, and especially Greg Ray for helpful comments on earlier drafts of this work.

——1976: 'Reply to Foster'. In Davidson 1984, pp. 171–9. Originally published in Evans and McDowell 1976, pp. 33–41.
——1984: *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.
Evans, Gareth, and John McDowell (eds.) 1976: *Truth and Meaning: Essays in Semantics*. Oxford: Oxford University Press.
Foster, J. A. 1976: 'Meaning and Truth Theory'. In Evans and McDowell 1976, pp. 1–32.
Katz, Jerrold J. 1966: *The Philosophy of Language*. New York: Harper and Row.
Katz, Jerrold J. and Paul M. Postal 1964: *An Integrated Theory of Linguistic Descriptions*. Cambridge, MA: MIT Press.
Kölbel, Max 2001: 'Two Dogmas of Davidsonian Semantics'. *Journal of Philosophy*, 98, pp. 613–35.
Lepore, Ernest, and Barry Loewer 1981: 'Translational Semantics'. *Synthese*, 48, pp. 121–33.
Lepore, Ernest, and Kirk Ludwig 2005: *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Clarendon Press.
——2007: *Donald Davidson's Truth-Theoretic Semantics*. Oxford: Oxford University Press.
Lewis, David 1970: 'General Semantics'. *Synthese*, 22, pp. 18–67.
Quine, W. V. 1940: *Mathematical Logic*. Cambridge, MA: Harvard University Press.
Ray, Greg 2014: 'Meaning and Truth'. *Mind*, 123, pp. 79–100.
Speaks, Jeff 2006: 'Truth Theories, Translation Manuals, and Theories of Meaning'. *Linguistics and Philosophy*, 29, pp. 487–505.
Vermazen, Bruce 1967: 'Review of An Integrated Theory of Linguistic Description by Jerrold J. Katz and Paul M. Postal, and The Philosophy of Language by Jerrold J. Katz'. Synthese, 17, pp. 350–65.

