

Beyond Logical Form

Brendan Jackson

(forthcoming in *Philosophical Studies*)

1. Introduction: Structural Entailment and Logical Form

Entailments like these are often called *structural entailments*:

1. Snowball is a black cat; so Snowball is a cat.
2. Lisa left school quietly; so Lisa left school.
3. Bart broke the window; so the window broke.
4. Homer ate pork chops; so pork chops were eaten.

Notice that each of (1)–(4) is an instance of a more general pattern. For example, we could replace ‘black’ in (1) with any of a wide range of other adjectives such as ‘furry’ or ‘hungry’ or ‘three-legged’, without rendering the entailment invalid or any less obvious. Similarly, there are a number of verbs that occur in entailments parallel to (3): ‘Moe boiled the water; so the water boiled’; ‘Bart blew up the school; so the school blew up’; ‘Homer sank the boat; so the boat sank’ and so on.

Structural entailment contrasts with *lexical entailment*, the latter of which includes the well-worn examples found in discussions of the analytic/synthetic distinction:

5. Bart is Lisa’s brother; so Bart is Lisa’s sibling.
6. Moe is a bachelor; so Moe is unmarried.
7. There are chairs in the room; so there is furniture in the room.

While the entailments in (5)–(7) may not be any less obviously valid than those in (1)–(4), their validity appears to depend on the particular meaning of one or more words in a way that the validity of structural entailments does not. If we replace ‘bachelor’ in (6) with almost any other

word of the same syntactic category—say, ‘husband’ or ‘bartender’—the entailment no longer holds. And (7) is no longer valid if we replace ‘chairs’ with most other plural count nouns (‘plants’, ‘children’, ‘elephants’, etc.) or ‘furniture’ with most other mass nouns (‘metal’, ‘water’, ‘wood’, etc.).

A further contrast between structural and lexical entailment can be brought out by borrowing a hypothetical scenario from Gilbert Harman.¹ Suppose we encountered a speaker who was unwilling to endorse (6) as valid and was generally unwilling to infer ‘unmarried’ sentences from ‘bachelor’ sentences. And suppose that upon interrogation we discovered that this speaker’s reluctance is due to his belief that a procedural flaw in the divorce courts has rendered almost all recent divorces invalid, so that many men whom we think of as bachelors are actually still married. Perhaps we should conclude that this speaker is confused about the meaning of ‘bachelor’; on the other hand, perhaps we should conclude that he understands the word but has highly idiosyncratic beliefs about bachelors. Pessimists about the analytic/synthetic distinction, like Harman, argue that there is no real fact of the matter about what we ought to say about such a speaker.

Contrast this case, however, with that of a speaker who is unwilling to accept (1) as valid—suppose she is willing to accept ‘Snowball is a black cat’ but rejects ‘Snowball is a cat’. Something has clearly gone wrong with this speaker. If her reluctance is limited to (1) and parallel sentences containing ‘black’ we would be reasonable to conclude that she is confused about the meaning of ‘black’, e.g. that she incorrectly thinks ‘black’ works like ‘fake’. But if, upon interrogation, we find that she is generally unwilling to accept the pattern exemplified in (1) as valid—she is unwilling to infer ‘man’ from ‘large man’, ‘dog’ from ‘shaggy dog’, ‘dancer’ from ‘good dancer’ and so on—we would not hesitate to conclude that she is confused in general about how standard adjectival modification works in English. Either way, her confusion is clearly confusion about meaning—however plausible it is in the lexical case, in this case it is highly implausible to maintain that she is a fully competent speaker who merely has idiosyncratic beliefs about the world.

It is tempting to construe these observations about the contrast between lexical and structural entailment as indicating that what is distinctive about the latter—as, indeed, the labels suggest—is that they are valid in virtue of structure or form rather than in virtue of lexical meaning. The study of the forms of sentences and arguments, as abstracted from their contents, is the traditional domain of formal logic, and many have thought that we can turn to logic in order to say what is distinctive about structural entailment. For example, in an introductory logic

¹ Harman 1999a, p. 127.

course ‘Snowball is a black cat’ would be translated by means of a conjunction, and perhaps (1) is really—in a sense of ‘really’ that is yet to be made clear—just an instance of conjunction elimination. This line of thought is no doubt familiar, and it is suggestive of a traditional view of structural entailment: that what distinguishes the structural entailment patterns is that they are valid in virtue of the underlying logical forms of sentences. I’ll refer to the view that an entailment is structural if and only if it is valid in virtue of logical form as the *Logician Thesis for Structural Entailment*, or the Logician Thesis for short.²

Though it is not often explicitly articulated, the Logician Thesis is often taken for granted in discussions in semantics and the philosophy of language. I am deeply suspicious of the Thesis, though, and my goal in this essay is to argue that we should reject it.

Arguing against the Logician Thesis requires first having a solid grasp on what it means, however, and as yet it is none too clear in what sense, e.g. ‘Snowball is a black cat’ is really a conjunction and (1) is really an instance of conjunction elimination. For clarification I’ll turn to theories of logical form developed within the context of Donald Davidson’s truth-conditional semantic framework. After outlining Davidson’s theory of logical form (section 2) I will argue that logical form so understood cannot support the Logician Thesis (section 3). Then I’ll examine the neo-Davidsonian theory of logical form developed by Richard Larson and Gabriel Segal, which might seem to offer a solution to the problem raised for the Thesis on Davidson’s view, and I’ll argue that on this view the Thesis encounters other problems (section 4). I’ll conclude by suggesting that the problems raised in the course of the discussion make the prospects for characterizing structural entailment in terms of logical form, Davidsonian or otherwise, look dim (section 5).³

² The label is meant to suggest the parallel between the thesis discussed here and logicism about mathematics. Just as the logicist about mathematics maintains that mathematical truths are, in some sense, really just logical truths, the supporter of the Logician Thesis maintains that structural entailment is, in some sense, really just logical consequence.

³ Jason Stanley (2000, pp. 391–92) draws a distinction between *revisionary* and *descriptive* conceptions of logical form. According to a revisionary conception, natural language is somehow inadequate for the purposes of rigorous inquiry in science or mathematics; claims about logical form are understood as claims about how natural language should be replaced with a notation that better suit the needs of such inquiry. According to a descriptive conception, by contrast, claims about logical form are meant to be taken as claims about the “real” or “underlying” forms of natural language sentences, forms that are often hidden by their surface forms. It is only logical form in this latter sense that is relevant to the Logician Thesis, and so in what follows I will not be concerned with revisionary conceptions of logical form. Also, in what follows, I won’t be concerned with Logical Form (LF) as it is employed in current research in syntactic theory. (Cf. May 1985, Chomsky 1995.) The project of describing LF is neutral with regard to the Logician Thesis, and it is extremely unlikely that this project will reveal the kind of logical structure needed to vindicate it.

2. Davidson on Logical Form

Our discussion will take place against the background of Davidson's views on the nature of semantic theory, and I will assume a basic familiarity with these views. I would like to briefly mention two features of Davidson's views that will be important in what follows, however.

A central Davidsonian tenet is that a Tarski-style truth theory for a language can suffice as a semantic theory for the language. According to Davidson, for a Tarski-style truth theory—or a T-theory, for short—for an object-language L to count as a theory of meaning for L it must at least entail an *interpretive T-sentence* for every sentence of L. A T-sentence is any instance of Tarski's schema T,

T. S is true in L iff p.

that results from replacing 'S' with a sentence of L and 'p' with a meta-language sentence with the same truth-value. A T-sentence is interpretive just in case the meta-language sentence that replaces 'p' on the right-hand side has the same meaning as the object-language sentence that replaces 'S' on the left-hand side. We'll call a Tarski-style truth theory for L that yields as theorems an interpretive T-sentence for every sentence of L an *interpretive T-theory*. Davidson develops a set of empirical and formal criteria for accepting T-theories, and argues that any theory that warrants acceptance on these grounds will, in fact, be interpretive and so will yield interpretive T-sentences for its object-language.

We will operate under the assumption that this is correct; a detailed examination of Davidson's criteria would require us to get into his theory of radical interpretation, and this would take us farther afield than is necessary for present purposes. In what follows one feature of Davidson's theory of radical interpretation will be important, however: on his view a central feature of empirical adequacy is that the T-theory must be supportable without appeal to claims about meaning—e.g. without claiming that it assigns the right meanings to words or that it accurately captures synonymy relations.⁴

An interpretive T-theory can serve as a semantic theory, in the sense that it can be regarded as specifying the meanings of object-language sentences. According to Davidson, there is also a deeper reason in virtue of which an interpretive T-theory counts as a semantic theory. An interpretive T-theory has a claim to being an adequate semantics, Davidson argues, because a subject who knew such a T-theory for language L—and who knew that it was interpretive—

⁴ Cf. Davidson 1984c, 1984d.

would know the meanings of sentences of L. Moreover, she “would know in detail how the truth values of sentences of L were owed to their structures, and why some sentences entailed others, and how words performed their functions by dint of relations to objects in the world.”⁵ Davidson’s claim that knowledge of an interpretive T-theory suffices for understanding is quite controversial, and one reason for the controversy will become important in section 4 below.⁶

Let’s turn, now, to Davidson’s theory of logical form. The theory has two components: an account of what it means to give the logical form of a sentence, and an account of what it means for an entailment to be valid in virtue of logical form.

Davidson’s account of what it means to give the logical form of a sentence arises from reflection on how Tarski’s method of defining truth for regimented languages is to be adapted to the treatment of natural languages like English:

The work of applying a theory of truth in detail to a natural language will in practice almost certainly divide into two stages. In the first stage, truth will be characterized, not for the whole language, but for a carefully gerrymandered part of the language. This part, though no doubt clumsy grammatically, will contain an infinity of sentences which exhaust the expressive power of the whole language. The second part will match each of the remaining sentences to one or (in the case of ambiguity) more than one of the sentences for which truth has been characterized. We may think of the sentences to which the first stage of the theory applies as giving the logical form, or deep structure, of all sentences.⁷

As the passage suggests, Davidson actually thinks of a semantic theory for a natural language as a T-theory accompanied by a set of axioms systematically linking the truth conditions of object-language sentences with sentences of “the gerrymandered part of the language”. Rather than trying to apply the T-theory to English sentences directly, the axioms allow us to correlate ordinary sentences with sentences of a regimented language to which Tarski’s method is directly applicable. In this passage Davidson indicates that the regimented language is to be contained in the language under study—e.g. as being something like “logicians’ English”. But this seems inessential, and elsewhere Davidson indicates that what matters is that the

⁵ Davidson 1984c, p. 128, 1984d, p. 172.

⁶ Cf. e.g. Soames 1989, 1992 and Higginbotham 1992 for discussion.

⁷ Davidson 1984c, p. 133.

language we use for the correlation be one to which Tarski's methods can be directly applied.⁸

Notice that, since the truth conditions of sentences are specified by the T-sentences for their regimented correlates, any theory that counts as interpretive must employ a correlation that is an accurate translation—i.e. one such that sentences and their regimented correlates are synonymous. Given Davidson's prohibition against appealing to meaning, however, this cannot be adopted as an explicit constraint on the choice of correlations. The choice is rather to be made on the basis of the same empirical and formal considerations that warrant acceptance of the theory as a whole; if Davidson is right then any theory that warrants acceptance on his grounds will, as a matter of fact, employ a correlation that amounts to a meaning-preserving translation manual from the object-language into the regimented language.

On Davidson's view, a semantic theory gives the logical form of a sentence by specifying its correlate in the regimented language associated with the theory. This claim is actually open to two interpretations, however. To illustrate, consider Davidson's suggestion that the logical forms of action sentences such as (8a) are given by formulas that quantify over events as in (8b):⁹

8. a. Brutus stabbed Caesar.
- b. $\exists e(\text{Stabbing}(\text{brutus}, \text{caesar}, e))$

((8b) can be glossed as 'There is a stabbing by Brutus of Caesar'.) The term 'logical form' is used sometimes in a *coarse-grained* and sometimes in a *fine-grained* sense. In the coarse-grained sense sentences with distinct meanings can have the same logical form; in this sense, for example 'Brutus stabbed Caesar' and 'Anthony hugged Cleopatra' have the same logical form. If we understand logical form this way then (8b) gives the form that the two sentences share in the sense that it is an instance of that form. We can think of the logical form itself as an abstraction from (8b), e.g. as what we get when we replace the non-logical constants in (8b) with schematic letters.¹⁰ In the fine-grained sense, by contrast, no two non-synonymous sentences have the same logical form. If we understand logical form this way then (8b) gives the logical form of (8a) by

⁸ Davidson 1984c, p. 136.

⁹ Davidson 1980a. I will suppress aspect as well as tense and other forms of context-sensitivity throughout.

¹⁰ Actually there are multiple ways to abstract from (8b) to arrive at a logical form in the coarse-grained sense. E.g. we could treat (8b) as an instance of $\lceil \exists v(\zeta(\alpha, \beta, v)) \rceil$ where ζ stands in for any three-place predicate, α and β stand in for any terms, and v stands in for any variable; we could also treat it as an instance of $\lceil \exists v\Phi \rceil$, where Φ stands in for any formula containing zero or more free occurrences of v . Strictly speaking, then, it is incorrect to describe (8b) as an instance of *the* logical form of (8a) in the coarse-grained sense.

being its logical form; the logical form of a sentence is the formula of the regimented language with which it is correlated. Davidson does not indicate which of the two senses he has in mind. For our purposes, however, it will do no harm to regard his view as an account of logical form in the fine-grained sense, and I will do so in order to simplify discussion.

Next we need to see what it means, on Davidson's view, for an entailment to be valid in virtue of logical form:

To see the structure of a sentence through the eyes of a theory of truth is to see it as built by devices a finite number of which suffice for every sentence; the structure of the sentence thus determines its relations to other sentences. And indeed there is no giving the truth conditions of all sentences without showing that some sentences are logical consequences of others...¹¹

In what sense does a theory show that some sentences are logical consequences of others? Elsewhere Davidson elaborates: "A theory of truth does not yield a definition of logical consequence or logical truth, but it will be evident from a theory of truth that certain sentences are true [and that certain entailments are valid] solely on the basis of the properties assigned to the logical constants."¹² These remarks suggest the following picture: for the entailment from sentence S_1 to sentence S_2 to be valid in virtue of logical form, according to a theory, is for the theory to show that S_2 is a logical consequence of S_1 . And for the theory to show *that* is for it to be a logical consequence of just the axioms of the theory that deal with the logical constants of the regimented language that if S_1 is true then S_2 is true. Following Gareth Evans, we'll call conditionals of this form—conditionals of the form 'If S_1 is true then S_2 is true' where S_1 and S_2

¹¹ Davidson 1984a, p. 61.

¹² Davidson 1984b, p. 71. Davidson goes on to say, "The logical constants may be identified as those iterative features of the language that require a recursive clause in the characterization of truth or satisfaction." This could be construed as a substantial principle about how to draw the distinction between logical and non-logical constants: an expression E is a logical constant just in case the axiom specifying E 's contribution to truth- or satisfaction-conditions is one of the recursive ones, i.e. one that uses the notion of truth or satisfaction on its right-hand side.

This is a problematic principle. One problem is that the identity predicate does not require a recursive clause, even though it is typically thought of as a logical constant. Another problem is that it may well be desirable to treat constants that are intuitively non-logical (e.g. adjectives) by means of recursive clauses. (Both of these objections are raised in Evans 1985a.) However, I don't think it's obvious that Davidson intends to commit himself to this problematic principle. The passage quoted is a tentative endorsement of the principle at best, and I can find no other passage in Davidson's writings in which it is repeated or defended. And whether or not he is committed to the principle, it is inessential to the overall theory of logical form. One is free to combine Davidson's general account of validity in virtue of logical form with one's preferred view about how to distinguish the logical constants. (Cf. n. 14.)

denote object-language sentences—*validating conditionals*.¹³ To a first approximation, then, the entailments that are valid in virtue of logical form are those whose validating conditionals are logical consequences of the axioms for the logical constants of the regimented language.¹⁴

This approximation won't quite do, however. Being valid in virtue of form is a property of entailments in the object-language, and the relevant validating conditionals must be ones linking the truth of object-language sentences. But such conditionals will not always be logical consequences solely of the T-theory's axioms for the logical constants. To see why it will help to consider an example:

9. a. Brutus stabbed Caesar quickly.
- b. $\exists e(\text{Stabbing}(\text{brutus}, \text{caesar}, e) \ \& \ \text{Quick}(e))$

Davidson proposes (9b) as the logical form of (9a); this and the suggested logical form for (8a) are intended to have the result that the structural entailment from (9a) to (8a) is valid in virtue of logical form. (8b) is clearly a logical consequence of (9b), and (C1) will follow logically from the standard axioms for the existential quantifier and the conjunction sign:

- C1. If ' $\exists e(\text{Stabbing}(\text{brutus}, \text{caesar}, e) \ \& \ \text{Quick}(e))$ ' is true then
' $\exists e(\text{Stabbing}(\text{brutus}, \text{caesar}, e))$ ' is true.¹⁵

To show that the entailment from (9a) to (8a) is valid in virtue of logical form, however, we need:

- C2. If 'Brutus stabbed Caesar quickly' is true then 'Brutus stabbed Caesar' is true.

The axioms for the logical constants in the T-theory do not license the move from (C1) to (C2); (C2) only follows given that (8a) is true just in case (8b) is true, and that (9a) is true just in case (9b) is true. These equivalences are given by the axioms that correlate sentences with their

¹³ Evans 1985a, p. 53.

¹⁴ There are substantial questions about the right way to distinguish logical from non-logical constants, and there is room for pessimism about whether any real distinction can be maintained. I'll leave these issues aside in the main text. Here let me note, however, that the Logicist Thesis requires *some* way of drawing the distinction, and it must not be too liberal if the Thesis is to be plausible—a view that counts 'brother' and 'sibling' as logical constants (to take an extreme example) will treat the lexical entailment from 'Bart is Lisa's brother' to 'Bart is Lisa's sibling' as valid in virtue of logical form.

¹⁵ Strictly speaking, more than just the axioms for the existential quantifier and conjunction are needed. A T-theory for a quantified language does not define truth directly, but rather defines it indirectly in terms of satisfaction, and the axiom relating truth to satisfaction will also be needed to arrive at (C1).

regimented counterparts. As we would expect, then, the correlation plays a crucial role in determining which entailments are valid in virtue of logical form. It is only by correlating object-language sentences with logical forms having the right kind of logical structure that a theory can generate validating conditionals like (C2) as theorems.

To recap, on Davidson's view the translation manual of a semantic theory maps object-language sentences onto their logical forms, and for the theory to treat an entailment as valid in virtue of logical form is for its validating conditional to be a logical consequence of the T-theory's axioms for the logical constants in conjunction with the axioms of the correlation component of the theory. In effect, the correlation "uncovers" logical structure in the sense that it allows us to see validating conditionals as following from the axioms for the logical constants even in cases involving object-language sentences that do not contain logical constants, as in the entailment from (9a) to (8a).

Let me conclude the section by highlighting a feature of Davidson's view that has so far been left in the background. It should be clear from our discussion that Davidson's view is what Evans calls an *immanent* rather than a *transcendent* characterization of logical form. Evans describes the distinction as follows:

One provides an immanent definition of some semantical term *W* if one does not define it absolutely but rather defines the notion 'e is *W* according to theory T'. One provides a transcendent definition when the definition contains no such relativity to a theory; when one says, rather, what a theory *ought* to treat as *W*.¹⁶

It only makes sense to talk about the logical form of a sentence, on Davidson's view, relative to a particular semantic theory—logical forms are assigned by the theory's correlation axioms. On Davidson's view logical forms can vary from theory to theory, along two dimensions. First, different theories can specify different correlations into the same regimented language. For example, instead of assigning (8b) to (8a) a theory might assign (8a) a logical form that quantifies over individuals as well as events:

8b'. $\exists e \exists x \exists y (\text{Stabbing}(x,y,e) \ \& \ \text{brutus} = x \ \& \ \text{caesar} = y)$

¹⁶ Evans 1985a, p. 50. Evans's use of these terms should be distinguished from Quine's 1970 use of the same terms to distinguish between language-relative and language-independent specifications of the logical constants.

A second way in which theories might disagree about logical form is in their choice of regimentations. For example, Davidson's proposal assumes a theory whose regimented language contains proper names. But we might instead follow Quine's suggestion of adopting a regimentation that eliminates proper names in favor of predicates; such a theory would assign (8b'')

8b''. $\exists e \exists x \exists y (\text{Brutizes}(x) \ \& \ \text{Caesarifies}(y) \ \& \ \text{Stabbing}(x, y, e))$

Both the choice of regimented languages and the choice of correlations are internal to particular semantic theories, and consequently Davidson's account of logical form must be construed in immanent terms.

If we want to give sense to the question of whether any of (8b)–(8b'') gives *the* logical form of (8a)—i.e. its logical form *simpliciter* rather than relative to one or another theory—the best we can do is treat it as the question of whether any one of the three will be assigned by a theory that meets Davidson's overall constraints on an adequate semantics. Davidson expresses optimism that his constraints are strong enough to rule out indeterminacy about assignments of logical form, but his optimism strikes me as unwarranted.¹⁷ Arguably mapping (8a) onto any of (8b)–(8b'') will yield an interpretive T-sentence for (8a). Since the constraints are designed only to guarantee that an adequate theory entails interpretive T-sentences, there isn't much reason to think that they will discriminate among (8b)–(8b''). If so then it will be a consequence of Davidson's view that there is a significant amount of indeterminacy about the logical forms *simpliciter* of sentences.

This indeterminacy might appear to scotch any attempt to support the Logicist Thesis within the context of a Davidsonian theory of logical form. This conclusion must not be drawn too hastily, however. Notice that, given appropriate adjustments to the logical form of (9a), the entailment from (9a) to (8a) will be valid in virtue of logical form on any of the three assignments of logical form to (8a). This shows that indeterminacy about logical forms does not necessarily translate into indeterminacy about which entailments are valid in virtue of logical form, and it suggests a way of understanding the Logicist Thesis. Even if Davidson's constraints do not guarantee that every interpretive theory assigns the same logical forms to sentences, they might be strong enough to guarantee that every interpretive theory does classify all and only structural entailments as valid in virtue of logical form. In the next section, however, I will take up this suggestion and argue against the possibility of using it to vindicate the Logicist Thesis.

¹⁷ Davidson 1984c, p. 136 n. 16.

3. Davidsonian Logical Form and the Logicist Thesis

3.1 Case study: the logical form of causatives

In this section I will argue that the Logicist Thesis should be rejected when understood within the context of Davidson's theory of logical form. Specifically, it is unlikely that every interpretive truth-conditional semantic theory will be forced to treat even core cases of structural entailment as valid in virtue of logical form. I'll first illustrate the point by looking at a specific case, and then I'll draw some general conclusions.

Let's return to the entailment pattern exemplified by (3):

3. Bart broke the window; so the window broke.

I'll label the pattern exemplified in (3) the *causative pattern* in virtue of the fact that sentences like 'Bart broke the window' are sometimes referred to as causative constructions. Terence Parsons suggests a way an interpretive theory might treat (3) as valid in virtue of logical form, using a regimented language that has two features we have not yet discussed. The first feature is that Parsons's regimented language includes a two-place predicate of events 'CAUSE(e_1, e_2)' that is glossed roughly as ' e_1 causes e_2 ' or ' e_1 makes e_2 happen'. The purpose of this predicate, as we will soon see, is to allow us to encode into logical form the fact that 'Bart broke the window' means roughly that Bart made the window break. Second, Parsons translates all verbs by means of one-place predicates of events rather than, as Davidson does, by means of predicates that take an event argument in addition to the more familiar arguments. On Parsons's view the relation, e.g. of 'Brutus' and 'Caesar' to 'stab' in (8a) is encoded in logical form by means of separate *thematic role* predicates, two-place predicates that relate events to their participants. On Parsons's view (8a) is mapped onto the logical form in (10):

10. $\exists e(\text{Stabbing}(e) \ \& \ \text{Agent}(e, \text{brutus}) \ \& \ \text{Theme}(e, \text{caesar}))$

The predicates 'Agent(e, x)' and 'Theme(e, x)' are used in (10) to reflect the fact that according to the original sentence the stabbing is *by* Brutus—he's doing the stabbing—and is *of* Caesar—he's the one getting stabbed.

Putting these elements together, Parsons suggests the following logical forms for the sentences in (3):

11. a. Bart broke the window.
b. $\exists e_1[\text{Agent}(e_1, \text{bart}) \ \& \ \exists e_2(\text{Breaking}(e_2) \ \& \ \text{Theme}(e_2, \text{the window}) \ \& \ \text{CAUSE}(e_1, e_2))]$
12. a. The window broke.
b. $\exists e(\text{Breaking}(e) \ \& \ \text{Theme}(e, \text{the window}))$

Given the intended interpretations of ‘CAUSE(e_1, e_2)’ and the thematic role predicates, (11b) can be roughly glossed as saying that Bart did something that caused the window to break. While one might hesitate to agree that this is really an accurate translation of (11a), it is close enough for our purposes, and we will assume that mapping (11a) onto (11b) yields an interpretive T-sentence for (11a).¹⁸ Moreover, given the translations above, the relevant validating conditional:

C3. If ‘Bart broke the window’ is true then ‘The window broke’ is true.

is a logical consequence of the usual axioms for the existential quantifier and the conjunction sign. So a theory that adopts Parsons’s translation scheme classifies (3) as valid in virtue of logical form, as desired.

There is no question that Parsons’s translation scheme provides an elegant way to capture the causative pattern in terms of logical form. The relevant question in the present context, however, is whether Parsons’s proposal, or something like it, is *required* in order to get an interpretive T-theory. The answer is pretty clearly no. To see why, suppose that instead of Parsons’s predicate ‘CAUSE(e_1, e_2)’ we encode the causal element in the meaning of (11a) by means of a sentence operator ‘CAUSE’ that applies to the open atomic formula ‘Breaking(e)’ as follows:

11b’. $\exists e[\text{CAUSE}(\text{Breaking}(e)) \ \& \ \text{Agent}(e, \text{bart}) \ \& \ \text{Theme}(e, \text{the window})]$

¹⁸ Objections have been raised against the possibility of treating transitive ‘break’ as synonymous with ‘cause to break’ (cf. e.g. Dowty 1979, Fodor 1970). Parsons (1990, ch. 6) responds to these objections, but I will not here be concerned with whether his response is successful.

The intuitive idea is that the ‘**CAUSE**’ operator turns an intransitive verb that takes only a theme argument into a transitive verb that takes both a theme and an agent argument. The formula ‘**CAUSE**(Breaking(e))’ can be understood roughly as ‘e is a causing-to-break’, and what (11b’) says, approximately, is that there is a causing-to-break of the window by Bart.

(11b’) does just as good a job as (11b) of translating (11a), so if the latter can be incorporated into an interpretive T-theory—as we are assuming—then so can (11b’). Nevertheless, (12b) is not a logical consequence of (11b’), and a theory that assigns (11a) the logical form (11b’) will not count (3) as valid in virtue of logical form. The problem is that the validating conditional (C3) follows only on the assumption that if ‘**CAUSE**(Breaking(e))’ applies to an event then there is an event to which ‘Breaking(e)’ applies. By appropriately stating the axiom for ‘**CAUSE**’ within our T-theory we can make sure that this is the case, and so we can make sure that (C3) is a consequence of our theory.¹⁹ But it won’t be a consequence that follows merely from the axioms for the logical constants.²⁰

In defense of his choice of mappings Parsons points to phenomena that he thinks no single-event analysis such as (11b’)—i.e. no mapping onto logical forms with only a single existentially bound event variable—will be able to adequately capture. If he is right then, contrary to what was just argued, the mapping suggested here cannot be part of an interpretive T-theory. One phenomenon Parsons cites is an alleged structural ambiguity in causatives containing prepositional phrases, e.g.:

13. Mary flew the kite behind the museum.

As Parsons points out, (13) can be understood either as indicating that Mary was behind the museum when she flew her kite or as indicating that her kite moved to a point behind the museum during its flight.²¹ Since Parsons treats prepositions as two-place predicates of events and objects, just like thematic role predicates, he is able to represent the two readings by means of the following logical forms:

¹⁹ Securing (C3) will also require a postulate linking the theme arguments associated with ‘**CAUSE**(Breaking(e))’ and ‘Breaking(e)’ so as to guarantee that if the window is the theme of a causing-to-break then it is also the theme of a breaking. Parsons makes use of similar postulates to secure other entailment relations; cf. Parsons 1990, pp. 78-82.

²⁰ If it could be argued that ‘**CAUSE**’ is actually a logical constant (as it is classified in Dowty 1979) then the entailment from (11b’) to (12b) would be valid in virtue of logical form. A view that counts ‘**CAUSE**’ as a logical constant is certainly non-standard, however, and it also runs the risk of being too liberal to be of use for defenders of the Logicist Thesis. (Cf. n. 14.)

²¹ There is at least one other possible reading of (13) where ‘behind the museum’ specifies which kite Mary flew. On this reading (13) would be an answer to the question ‘Which kite did Mary fly?’.

14. a. $\exists e_1[\text{Agent}(e, \text{mary}) \ \& \ \exists e_2(\text{Flying}(e_2) \ \& \ \text{Theme}(e_2, \text{the kite}) \ \& \ \text{CAUSE}(e_1, e_2) \ \& \ \text{Behind}(e_1, \text{the museum}))]$
- b. $\exists e_1[\text{Agent}(e, \text{mary}) \ \& \ \exists e_2(\text{Flying}(e_2) \ \& \ \text{Theme}(e_2, \text{the kite}) \ \& \ \text{CAUSE}(e_1, e_2) \ \& \ \text{Behind}(e_2, \text{the museum}))]$

In (14a) ‘behind the museum’ is represented as applying to the event of which Mary is the agent; in (14b) it is represented as applying to the event with the kite as the theme. A single-event analysis cannot capture the ambiguity in (13) in this way.

There isn’t much reason to think it needs to be captured in this way, however. Notice, first of all, that the same ambiguity can be observed in sentences that Parsons would give a single-event analysis:

15. Mary pushed the bicycle behind the garage.

One reading of (15) indicates that Mary did the pushing while behind the garage, while the other indicates that the bicycle moved to a point behind the garage; yet ‘push’ does not occur in the causative pattern (cf. ‘The bicycle pushed’) and would receive a logical form parallel to the one for ‘stab’, involving only a single event variable. Notice, moreover, that not all causatives with prepositional phrases are ambiguous in the way (13) is:

16. Mary dropped the bucket into the well.

The only reading available for (16) is one according to which the bucket moved into the well; it is not possible to understand the sentence as requiring for its truth that Mary went down the well. These are observations that Parsons must find some way of explaining—prima facie, his view predicts that (15) should be unambiguous and that (16) should be ambiguous. And while the semantics of prepositional phrases is notoriously tricky, it is at least clear in broad outline what ought to be said about examples like (15) and (16). Prepositions like ‘behind’, ‘under’, ‘near’, etc. have both a locative use (‘The book is under the table’) and a use indicating the goal of motion (‘The cat ran across the street and under the car’), and these two uses clearly correspond to the two readings of both (13) and (15). By contrast, ‘into’ has only the latter use (cf. ‘The bucket is into the well’). These observations suggest that Parsons’s data ought to be accounted for in terms of the meaning of the prepositional phrases, rather than in terms of differences in

which event variables the phrases take as arguments. If so then the data do not count against the single-event analysis in (11b').

Parsons points to a related phenomenon that might seem harder to account for on a single-event analysis. Consider the following sentences:

- 17. a. Mary felled the tree into the lake with the axe.
- b. The tree fell into the lake.
- c. ?The tree fell with the axe.

While (17a) clearly entails (17b) it does not entail (17c); in fact, if we read 'with the axe' in (17c) as indicating the instrument with which the tree was felled, as we do in (17a), (17c) seems quite awkward.²² However, on the single-event analysis the logical form of (17c) is a logical consequence of the logical form of (17a):

- 17. a'. $\exists e(\text{CAUSE}(\text{Felling}(e)) \ \& \ \text{Agent}(e,\text{mary}) \ \& \ \text{Theme}(e,\text{the tree}) \ \& \ \text{Into}(e,\text{the lake})$
 $\ \& \ \text{With}(e,\text{the axe}))$
- c'. $\exists e(\text{Felling}(e) \ \& \ \text{Theme}(e,\text{the tree}) \ \& \ \text{With}(e,\text{the axe}))$

This seems to show that the single-event analysis must be abandoned in favor of an analysis like the one Parsons provides.

This argument rests on a false assumption, however. While (17c') is, undeniably, a logical consequence of (17a'), this only predicts that (17a) entails (17c) on the assumption that (17c') is the logical form of (17c). But the correlation component of a semantic theory is not a mapping from the set of object-language sentences *onto* the set of formulas of the regimented part of the language; there will be lots of well-formed formulas to which no ordinary English sentence corresponds, even on Parsons's view. For example:

- 18. a. * Bart giggled Lisa.
- b. $\exists e_1[\text{Agent}(e_1,\text{bart}) \ \& \ \exists e_2(\text{Giggling}(e_2) \ \& \ \text{Theme}(e_2,\text{lisa}) \ \& \ \text{CAUSE}(e_1,e_2))]$

²² It is also possible to read 'with the axe' in both (17a) and (17c) as indicating that the axe fell along with the tree. On this reading (17c) is fully grammatical and does follow from (17a).

19. a. ? Bart sneezed with the pepper.
 b. $\exists e(\text{Sneezing}(e) \ \& \ \text{Theme}(e,\text{bart}) \ \& \ \text{With}(e,\text{the pepper}))$

Example (19) is especially relevant. Intuitively, (19a) is marked because the instrumental prepositional phrase ‘with the pepper’ requires an agent who uses the pepper to perform some action, but Bart isn’t an agent according to (19a)—sneezing is something that happens to him rather than something he does. As we noted above, (17c) is similarly odd on the intended instrumental reading of ‘with the axe’. The fact that (17c’) follows logically from (17a’) no more predicts that (17a) entails (17c) than the fact that (19b) is a well-formed formula predicts that (19a) is an acceptable sentence of English.

Neither of the problems Parsons raises for the single-event analysis is decisive. Barring further argument, then, both the single-event analysis suggested here and Parsons’s double-event analysis appear to do an equally good job of getting the right meanings for sentences that exemplify the causative entailment pattern. Yet the single-event analysis captures the entailment by means of the lexical axiom for the operator ‘**CAUSE**’, and consequently does not treat the pattern as valid in virtue of logical form. So in this case, at least, it appears that there is no need for an interpretive theory to treat the entailment as valid in virtue of logical form.

3.2 General conclusions

What the discussion of causatives illustrates is that in some cases it is possible to make a kind of trade-off between logical structure and lexical meaning: the constraints that Davidson intends to yield interpretive theories leave it open whether the causative pattern is to be represented as valid in virtue of logical form or as being guaranteed by the non-logical lexical axioms.

Evans was perhaps the first to note that this is a fundamental and potentially fairly general problem for the Logician Thesis in the Davidsonian framework:

Provided we have a scheme for the suitable translation of what are inferentially the weakest members of a certain grammatical category, we seem to be enjoined to run that scheme quite generally, treating the inferences which the inferentially stronger members sustain as being consequences of their analysis [i.e. as being captured by the lexical axioms].²³

²³ Evans 1985a, p. 58.

A classic illustration of Evans's point is Richard Montague's treatment of extensional transitive verbs like 'find' and intensional transitives like 'seek'. Verbs in these two categories differ in the structural entailments they license, e.g.:

- 20. a. John found a unicorn.
b. There is a unicorn John found.
- 21. a. John sought a unicorn.
b. There is a unicorn John sought.

While (20a) entails (20b), (21a) does not entail (21b). Despite this difference, Montague proposes a uniform strategy for mapping both sentences containing 'seek' and sentences containing 'find' into formulas of intensional logic. The entailment in (20) is then accommodated within the lexical semantics for 'find'.²⁴ Another example is provided by intensional adjectives like 'fake' and 'counterfeit', which do not occur in the pattern exemplified by (1):

- 22. a. Snowball is a fake cat.
b. Snowball is a cat.

Intensional adjectives are notoriously difficult to handle within a Davidsonian truth-conditional framework, yet if the framework is to have any chance of producing a comprehensive semantic theory there must be *some* way of developing an interpretive T-theory for sentences like (22a). With appropriate adjustments to the lexical axioms, such a theory is likely to be applicable to ordinary extensional adjectives like 'black' as well; on such a theory the validating conditional for (1) will follow from the lexical axioms and (1) won't be valid in virtue of logical form.

So the Logicist Thesis is implausible when understood in the context of Davidson's theory of logical form. Simply put, if our only guide in assigning logical forms is to find a way to fit natural language into an interpretive T-theory then, in many cases, it is left open whether or not to treat structural entailment patterns as valid in virtue of logical form.

In actual practice, however, this is usually not our only guide—most logical form theorists tend to use structural entailment patterns themselves as an indication of what logical forms to assign. For example, all else being equal, most logical form theorists would accept a theory, such as Parsons's, that treats the causative pattern as valid in virtue of logical form over one that does not. This is not a practice that can be condoned within Davidson's framework,

²⁴ Cf. Dowty, Wall and Peters 1981, ch. 7 for discussion.

however. To argue for a theory on the grounds that it captures structural entailment patterns in terms of logical form is to help oneself to a semantic notion, and, as we saw in section 2, that is off-limits according to Davidson. Davidson's hope is rather that the facts about structural entailment will emerge along with facts about the meanings of sentences from an independently grounded theory.

Nothing said here shows that we cannot ignore Davidson's scruples, however, and use structural entailment as a guide when assigning logical forms. To do so, however, is to undercut the explanatory value of the Logician Thesis. We can, if we like, require that structural entailments be treated as valid in virtue of logical form and then reject theories that say otherwise. But then the question remains, what is it about structural entailment that calls for this kind of distinguished treatment, and that gives it a privileged role in evaluating semantic theories? It will be necessary to look beyond logical form to answer this question.

4. Cognitivism and the Logician Thesis

4.1 Logical form in truth-conditional semantics II: Larson and Segal's view

Up to this point we have been requiring of assignments of logical form only that they yield interpretive T-sentences. These days, however, there is a growing consensus that an adequate semantic theory must do more than correctly specify meanings—it must also correctly describe the knowledge competent speakers have, and in virtue of which they count as competent. A version of this view is endorsed by Larson and Segal. Larson and Segal combine the Davidsonian truth-conditional approach with a view they label *cognitivism*. According to cognitivism, semantic theory is (part of) the theory of competent speakers' knowledge of language, and, as they say, "knowledge of language is knowledge of a set of rules and principles that assign representations and meanings to the physical forms of signs..."²⁵ Larson and Segal maintain that as competent speakers, people tacitly know the axioms of an interpretive T-theory. On their view the role of Davidson's translation manual is played by one component of the language faculty. The language faculty assigns representations to sentences, and the meaning assigned to a given sentence by the language faculty is determined by applying the speaker's T-theory to its corresponding representation in order to derive an interpretive T-sentence.

Larson and Segal's cognitivism provides the basis for a transcendent account of logical form. In a cognitivist framework we can treat claims about logical form as, at bottom, claims

²⁵ Larson and Segal 1995, p. 11.

about speakers' psychology—about what representations speakers assign to sentences for the purposes of semantic interpretation. On their view, to say that Parsons's translation scheme for causatives is correct, for example, is to say that his scheme is the one employed by the competent speaker's language faculty. That is, the representation the competent speaker assigns to 'Bart broke the window' is (11b) rather than (11b'), and she applies her knowledge of her T-theory to (11b) in order to determine the meaning of 'Bart broke the window'. If so then a theory that assigns the sentence (11b') is an incorrect description of her knowledge of language, whether or not it yields an interpretive T-sentence for (11a). So Larson and Segal's cognitivism provides a further criterion for choosing between theories that make different claims about logical form, even between theories that succeed in assigning the right meanings to sentences.

A consequence of this view is that it is an open question whether all, some or no structural entailments are valid in virtue of logical form; the final verdict is hostage to speakers' psychology. Still, one can argue that within the context of cognitivism speakers' judgments about structural entailments should count as evidence about what semantic representations they assign to sentences. Perhaps, that is, a speaker's judgment that the entailment from S_1 to S_2 is structural is indicative of the fact that she assigns semantic representations in such a way that S_2 is a logical consequence of S_1 . If this is so then we may be able to draw conclusions about logical form in advance of results in cognitive psychology and neuroscience. We'll return to this point below, but first we need to look at another important feature of Larson and Segal's view.

Recall from section 2 that Davidson maintains that a subject who knew an interpretive T-theory for L, and who knew that it was interpretive, would be able to understand the sentences of L. This is why, according to Davidson, an interpretive T-theory for L can be regarded as a semantic theory for L. It is easy to demonstrate, however, that knowledge of an interpretive T-theory does not suffice for understanding. The problem is that even an interpretive T-theory will have as theorems T-sentences that are not interpretive:

T1. 'Krusty is illiterate' is true iff Krusty is illiterate.

T2. 'Krusty is illiterate' is true iff Krusty is illiterate and (if Lisa is smart then Lisa is smart).

While only (T1) is interpretive, (T2) is a logical consequence of (T1) and so will be a theorem of any theory of which (T1) is a theorem. A speaker might know such a theory and yet not know whether (T1) or (T2) gives the meaning of 'Krusty is illiterate'. So a speaker might know an

interpretive T-theory for English, know that it is interpretive, and yet not understand English because she doesn't know which T-sentences of the theory are the interpretive ones. Larson and Segal label this the *extension problem*.²⁶

The extension problem is even more acute for Larson and Segal than for Davidson. On their view it is not merely that knowledge of an interpretive T-theory *would* suffice for understanding; such knowledge is what is supposed to explain our *actual* understanding. To solve the problem Larson and Segal must offer an account of how we, as competent speakers, distinguish the interpretive T-sentences of our T-theory from the non-interpretive ones. Their solution is to maintain that linguistic competence involves more than just knowledge of an interpretive T-theory. According to them the language faculty also contains a set of inference rules that the speaker uses to derive T-sentences from the T-theory she tacitly knows. The T-sentences that the speaker knows in virtue of her semantic competence are all and only those that can be derived from the T-theory using the rules built into the language faculty. Larson and Segal's hope is that, given the right rules, this set will comprise the interpretive T-sentences of the speaker's language.

Whether this way of addressing the extension problem succeeds depends on what the supposed rules of the language faculty are. Larson and Segal rely primarily on only two rules. The first is a rule of universal instantiation:

$$\text{UI. } \frac{\forall \alpha \dots \alpha \dots}{\dots \tau \dots}$$

(Here $\lceil \dots \alpha \dots \rceil$ is any meta-language sentence containing zero or more occurrences of variable α and $\lceil \dots \tau \dots \rceil$ is the result of replacing every occurrence of α with an occurrence of term τ .) The second rule is a rule allowing substitution of proven equivalents:

$$\text{SE. } \frac{\dots \Phi \dots \quad \Phi \text{ iff } \Psi}{\dots \Psi \dots}$$

(Here Φ and Ψ are any meta-language sentences, $\lceil \dots \Phi \dots \rceil$ is any meta-language sentence containing Φ and $\lceil \dots \Psi \dots \rceil$ is the result of replacing one or more occurrences of Φ with occurrences of Ψ .) While (T2) is a logical consequence of (T1), (UI) and (SE) are much too

²⁶ Larson and Segal 1995, pp. 34; cf. also Fodor and Lepore 1992, chapter 2.

weak to let us derive (T2) and therefore, on their view, (T2) is not a theorem that the speaker knows in virtue of her competence with English. The success of this strategy in general requires being able to specify a set of rules that yield all the interpretive T-sentences for the language without yielding un-interpretive ones as well. As Larson and Segal note, whether this can be done remains an open question.²⁷

4.2 A dilemma

While Larson and Segal's proposal may provide a solution to the extension problem, it presents a dilemma for the Logicist Thesis. On the original Davidsonian view an entailment is valid in virtue of logical form according to a theory just in case its validating conditional is a logical consequence of the theory's translation manual and the axioms for the logical constants. In the context of Larson and Segal's cognitivism, however, we are faced with a choice: to either keep the original Davidsonian conception or instead to classify as valid in virtue of logical form only those entailment patterns whose validating conditionals can be proven using the rules contained in the language faculty.

The latter choice would be disastrous for the Logicist Thesis. Given Larson and Segal's rules not even (1)—repeated below in (23)—is valid in virtue of logical form, even if it is assigned a conjunctive logical form:

23. Snowball is a black cat; so Snowball is a cat.

Proving the validating conditional corresponding to (23) requires a rule of conjunction elimination, and according to Larson and Segal, the language faculty has no such rule. For essentially the same reason, it will also turn out that Davidson's entailment patterns among action sentences—the ones that lead him to posit quantification over events—will also fail to be valid in virtue of logical form. Following this horn of the dilemma, in fact, yields the result that hardly any entailment pattern is valid in virtue of logical form.

Despite its disastrous results for the Logicist Thesis there is some pressure for cognitivists like Larson and Segal to take this horn of the dilemma. As mentioned above, within the context of cognitivism, competent speakers' judgments of structural entailment are regarded

²⁷ Larson and Segal also propose three highly constrained rules for meta-language conjunction, disjunction and negation (cf. Larson and Segal 1995, pp. 85-86). Even with these rules it will not turn out that (T2) is a theorem that the speaker knows in virtue of her competence with English.

as providing evidence about the logical forms they assign to sentences. When a competent speaker judges that ‘Brutus stabbed Caesar with a knife’ entails ‘Brutus stabbed Caesar’, this is because she judges, on the basis of her knowledge of language, that the entailment is valid in virtue of logical form. If this sort of reasoning is to work then it must be that the logic the competent speaker uses to arrive at her judgment is the logic she possesses in virtue of being a competent speaker.

Another way of making the same point is to say that if we suppose that speakers go beyond the logic they know in virtue of their competence when making judgments about structural entailments then it becomes unclear why speakers’ judgments should be evidence about logical form rather than evidence about their extra-linguistic reasoning capacities. For example, suppose that what the speaker knows in virtue of her linguistic competence yields only the homophonic T-sentences (T8a) and (T9a), rather than Davidson’s T-sentences that involve quantification over events:

T8a. ‘Brutus stabbed Caesar’ is true iff Brutus stabbed Caesar.

T9a. ‘Brutus stabbed Caesar quickly’ is true iff Brutus stabbed Caesar quickly.

If we are allowed to appeal to speakers’ extra-linguistic reasoning capacities when explaining her judgments, it will be quite straightforward to explain her judgment that ‘Brutus stabbed Caesar quickly’ entails ‘Brutus stabbed Caesar’. Given that she knows (T8a) and (T9a), she can use her knowledge that when one stabs quickly one thereby stabs—or more generally, that performing an action quickly is a way of performing the action—in order to conclude that if the (9a) is true then (8a) will also be true. There is no need to suppose that she assigns the sentences a logically complex structure in order to explain her judgment. Conversely, if a speaker fails to recognize the entailment pattern as valid then we have no grounds for concluding—as we are strongly inclined to do—that her failure is a failure of understanding rather than a confusion about the metaphysics of actions.

There is strong pressure, then, for the cognitivist to take the first horn of the dilemma and give up the Logicist Thesis.²⁸ Suppose we put this aside, however, and take the second horn—

²⁸ Larson and Segal do, as a matter of fact, give up the Logicist Thesis. According to them, most of our judgments about entailment relations among sentences are to be explained in terms of our extra-linguistic reasoning capacities. Nevertheless they do endorse a version of Davidson’s proposal about the logical forms of action sentences. In the light of the remarks of the last paragraph, it is unclear why this is so. They might have something like the following justification in mind: the structural entailment patterns

that an entailment is valid in virtue of logical form as long as its validating conditional is a logical consequence of the speaker's T-theory, whether or not it can be proven using the rules built into the language faculty. If so then—assuming that the speaker assigns the right kinds of semantic representations to sentences—all sorts of entailment patterns might turn out to be valid in virtue of logical form. Many of them, however, are patterns that should not count as structural. For example, the entailments from 'Marge is tall' to all of the following count as valid in virtue of logical form on this view:

24. Either Maggie is short or Maggie is not short.

25. If Maggie is short then Marge is tall.

26. Either everything is tall or there is something that isn't Marge.

All of these entailments are obviously valid, but they're very much unlike the examples of structural entailment with which we began. The entailment to (24), for example, intuitively depends on the meaning of 'either...or...'. If we were to replace that expression with another of the same syntactic category, say 'and', the entailments would no longer go through. Similarly, the inference to (26) depends, *inter alia*, on the meaning of 'everything'—if we were to replace it with a quantifier like 'a few things' the sentence would no longer follow from 'Marge is tall'. These entailments should be classified as lexical rather than structural, but on the present suggestion they'll all be valid in virtue of logical form. Notice that this last objection is a problem for Davidson's view as well. In fact, it points to a fundamental problem for the attempt to explain structural entailment in terms of logical form, and we'll return to this issue in the conclusion.

While Larson and Segal's view places more demands on assignments of logical form than merely that they yield interpretive T-sentences—and thereby has the potential to avoid the problem raised in section 3—the Logicist Thesis nevertheless fares badly. Perhaps, then, we should begin to doubt that the Logicist Thesis succeeds in getting at what is distinctive about structural entailment.

exhibited among action sentences (e.g. that (9a) entails (8a)) are extremely obvious to any competent speaker, and by giving the sentences logical forms like the ones Davidson suggests we can account for this obviousness by showing that speakers need only use their (extra-linguistic) logical capacities in order to detect the patterns. This justification doesn't seem persuasive. Surely it is no less obvious that one can't stab quickly without stabbing—given that the speaker understands the sentences, why isn't this enough to explain the obviousness of the entailment?

5. Conclusion

In discussing the Logicist Thesis we have focused exclusively on Davidsonian theories of logical form, and I have argued that in the context of these theories the Logicist Thesis can't plausibly be sustained. While I am certainly not in a position to be able to show that there is no other way of understanding logical form such that the Logicist Thesis turns out to be true, there is good reason to be pessimistic about the chances of finding a way to save it.

For one thing, while not all philosophers and semanticists share Davidson's view that the right approach to semantics is via a Tarski-style truth theory, many share a conception of logical form that has much in common with his. One frequently encounters the view that logical forms are given by mapping sentences onto formulas of some regimented language, with the rationale that mapping sentences onto formulas allows us to bring the well-understood semantic machinery developed for formal languages to bear in capturing the meanings of natural language sentences. Any such view will encounter the problem raised in section 3, whether the preferred semantic machinery is a Tarski-style truth definition, the semantics of intensional logic, a theory of structured propositions, or what have you. In many cases the meanings of sentences can be adequately captured without positing the underlying logical structure necessary to support the Logicist Thesis.

Theories that, like Larson and Segal's, expect more of assignments of logical form than merely that they yield correct meaning specifications will encounter the problem raised at the end of section 4. Any theory that can claim to be a theory of *logical* form will classify the familiar logical entailment patterns as valid in virtue of logical form. But, as I argued above, these are lexical entailment patterns: they depend for their validity on the meanings of the logical constants 'and', 'or', 'not' and the rest. Indeed, it is something of an orthodoxy that the entailment patterns the logical constants license (or perhaps a privileged subset of them) are what *constitute* or *determine* their meanings. Validity in virtue of logical form therefore cuts across the structural/lexical distinction, and so misses what is distinctive about structural entailment.

This is a very simple point, but I think it is often obscured by a tendency to think of logic as purely formal—as the study of the forms or structures of sentences and arguments, abstracted from their specific contents. When we do so it is easy to lose sight of the fact that when we study logic we are not abstracting from *all* aspects of the contents of sentences and arguments; if we did we would abstract away from the difference between $\lceil\Phi$ and $\Psi\rceil$ and $\lceil\Phi$ or $\Psi\rceil$, and between $\lceil\forall x\Phi\rceil$ and $\lceil\exists x\Phi\rceil$, and logic would become very boring. The tendency to lose sight of the fact

that logical entailment patterns are lexical is, I suspect, encouraged by the standard way of defining formulahood in formal systems. Logical constants are treated syncategorematically, as the syntactic glue that holds together complex formulas, and this can encourage us to think of the constants as mere markers for this or that kind of well-formed formula. Quine, for example, lumps logical constants together with parentheses and accent markers on variables as “particles” to be distinguished from the “lexicon”, and describes the former as mere devices for indicating the various sorts of syntactic constructions dealt with in logic.²⁹ Once we do this it becomes easy to think of the difference in meaning between $\lceil \Phi \text{ and } \Psi \rceil$ and $\lceil \Phi \text{ or } \Psi \rceil$ as due to a difference in syntactic structure, rather than as due to the difference in meaning between ‘and’ and ‘or’. This despite the fact that there is a clear sense in which ‘and’ and ‘or’ play the same syntactic role: both can be used to construct a sentence by conjoining two other sentences.³⁰ And since logical constants are treated as syncategorematic, once we turn to the semantics for formal systems the logical constants, like parentheses and accent markers, do not receive interpretations and their contribution to the meanings of sentences is regarded as purely a matter of indicating syntactic structure.

There is, of course, nothing wrong with treating logical constants as syncategorematic, and for distinguishing types of formulas in terms of what constants they contain. When we turn from formal logic to the study of structural entailment in natural language, however, it is crucially important not to think of logical constants as mere indicators of syntactic structure. The sentences ‘Snowball is a black cat’ and ‘Snowball is black and Snowball is a cat’ do not differ merely in that the latter contains a particle indicating the syntactic structure it shares with the former.³¹ The latter contains a word that has distinctive inferential properties, properties it doesn’t share with other words such as ‘or’ that play the same syntactic role. The former contains no such word. ‘Snowball is a black cat’ isn’t really just a conjunction, intro logic notwithstanding.

If we do give up the Logicist Thesis, as I have argued that we should, where does that leave us with respect to the distinction between structural and lexical entailment? One *could* take

²⁹ Quine 1970, pp. 26-27.

³⁰ This way of thinking about logic is given an especially clear statement by Bertrand Russell: “It is one of the marks of a proposition of logic that, given a suitable language, such a proposition can be asserted in such a language by a person who knows the syntax *without knowing a single word of the vocabulary.*” (Russell 1920, p. 201; italics added.)

³¹ It is interesting to note that there is nothing in the syntax of modification that corresponds to ‘Snowball is brown or Snowball is a cat’. If ‘Snowball is a brown cat’ is really just a conjunction, one wonders why there is no analogous sentence (in any natural language) that is really just a disjunction. An interesting explanation of this fact is suggested by Pietroski’s (2002) work on the semantic significance of syntactic concatenation.

the failure of the attempt to draw the distinction in terms of logic as indicating that there really is no deep distinction here, appearances to the contrary notwithstanding. I think this conclusion would be too hasty. Gareth Evans nicely summarizes the structural/lexical distinction as being the distinction between,

...inferences whose validity depends merely upon the *kind* of semantic elements out of which a sentence is constructed, and its manner of construction, on the one hand, and inferences whose validity depends upon the special variation a particular semantic element is playing upon the theme all expressions of its kind must play, on the other.³²

Evans sees words as being sorted into various semantic kinds or categories, just as they are sorted into various syntactic categories such as nouns, verbs, etc. And in the above passage Evans suggests that this way of thinking of words can provide insight into the nature of structural entailment: perhaps the structural entailments are those whose validity is guaranteed by the semantic kinds or categories of the words out of which they are composed, rather than by their specific meanings. On this way of thinking of things, words such as ‘black’, ‘furry’, ‘hungry’, and ‘young’ belong to a unified semantic kind, and the validity of the pattern in (1)/(23) is partly a consequence of the distinctive semantic properties that words of that kind share.

I think Evans’s suggestion is a promising one. Notice, for starters, that this way of thinking of things provides a diagnosis for the failure of Davidsonian logical form observed in section 3: differences in semantic kind between ‘black’ and ‘fake’ (or between ‘find’ and ‘seek’) need not be reflected as differences in Davidsonian logical form between sentences containing words of the two semantic kinds, and so we can’t expect differences in what entailments the two kinds of words license to show up as differences in which entailments are classified as valid in virtue of logical form. If this is right then the road to a more adequate understanding of structural entailment will proceed through an investigation into the nature of semantic kinds. This investigation, however, is work for another occasion.³³

³² Evans 1985a, p. 60–61. Italics in original.

³³ Earlier drafts of this paper were read at Cornell University and Lewis & Clark College, and I am grateful to audiences at both places for interesting discussion and helpful comments. I have also benefited enormously from many discussions on the topics of this paper with Eric Gilbertson, Harold Hodes, Zoltán Szábo and especially with Delia Graff. Many thanks are due to these patient and thoughtful interlocutors.

References

- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge: MIT Press.
- Davidson, Donald. 1984. *Inquiries Into Truth and Interpretation*. Oxford: Clarendon Press.
- . 1984a. "Semantics for Natural Languages" in Davidson 1984.
- . 1984b. "In Defence of Convention T" in Davidson 1984.
- . 1984c. "Radical Interpretation" in Davidson 1984.
- . 1984d. "Reply to Foster" in Davidson 1984.
- . 1980. *Essays on Actions & Events*. Oxford: Clarendon Press.
- . 1980a. "The Logical Form of Action Sentences" in Davidson 1980.
- . 1980b. "Criticism, Comment and Defence" in Davidson 1980.
- Dowty, David. 1979. *Word Meaning and Montague Grammar*. Boston: D. Reidel Publishing Co.
- Evans, Gareth. 1985. *Collected Papers*. Oxford: Clarendon Press.
- 1985a. "Semantic Structure and Logical Form" in Evans 1985.
- 1985b. "A Further Reflection on Semantic Structure and Logical Form" in Evans 1985.
- Evans, Gareth and John McDowell. 1976. *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.
- Fodor, Jerry. 1970. "Three Reasons not to Derive 'Kill' from 'Cause to die'." *Linguistic Inquiry* v. 1, pp. 429-38.
- Fodor, Jerry and Ernest Lepore. 1992. *Holism: A Shopper's Guide*. Cambridge: Blackwell Publishers.
- Foster, J.A. 1976. "Meaning and Truth Theory." in Evans and McDowell 1976.

- Harman, Gilbert. 1999. *Reasoning, Meaning, and Mind*. Oxford: Clarendon Press.
- 1999a. "The Death of Meaning" in Harman 1999.
- Higginbotham, James. 1992. "Truth and Understanding." *Philosophical Studies* v. 65, pp. 3-16.
- Larson, Richard and Gabriel Segal. 1995. *Knowledge of Meaning*. Cambridge: MIT Press.
- Linsky, Leonard. 1952. *Semantics and the Philosophy of Language*. Urbana: University of Illinois Press.
- May, Robert. 1985. *Logical Form: Its Structure and Derivation*. Cambridge: MIT Press.
- Parsons, Terence. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. Cambridge: MIT Press.
- Pietroski, Paul. 2002. "Function and Concatenation" in Preyer and Peter 2002.
- Preyer, Gerhard and Georg Peter. 2002. *Logical Form and Language*. New York: Oxford University Press.
- Quine, W.V. 1991. "Two Dogmas in Retrospect." *Canadian Journal of Philosophy* v. 21, pp. 265-274.
- . 1970. *Philosophy of Logic*. Englewood Cliffs: Prentice-Hall, Inc.
- Russell, Bertrand. 1920. *Introduction to Mathematical Philosophy*. New York: Dover Publications, Inc.
- Soames, Scott. 1992. "Truth, Meaning, and Understanding." *Philosophical Studies* v. 65, pp. 17-35.
- Soames, Scott. 1989. "Semantics and Semantic Competence." *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*, pp. 575-596.
- Stanley, Jason. 2000. "Context and Logical Form." *Linguistics and Philosophy* v. 23, pp. 391–434.
- Tarski, Alfred. 1952. "The Semantic Conception of Truth" in Linsky 1952.