Logical Form

Forthcoming in: Ernie Lepore & Kirk Ludwig (eds.), *Blackwell Companions to Philosophy - Companion to Donald Davidson*, Wiley-Blackwell. This is the penultimate version, please refer to the published version.
Comments are welcome: miguel.hoeltje[a-t]uni-due[d-o-t]de.

Donald Davidson contributed to the discussion of logical form in two ways. On the one hand, he made several influential suggestions on how to give the logical forms of certain constructions of natural language. His account of adverbial modification and so called action-sentences is nowadays, in some form or other, widely employed in linguistics (Harman (*forthcoming*) calls it "the standard view"). Davidson's approaches to indirect discourse and quotation, while not as influential, also still attract attention today. On the other hand, Davidson provided a general account of what logical form *is*. This paper is concerned with this general account. Its foremost aim is to give a faithful and detailed picture of what, according to Davidson, it means to give the logical form of a sentence. Discussions of Davidson's proposals concerning action-sentences and indirect discourse can be found in chapters 6 and 12 of this collection. For two excellent recent overview articles on logical form that are not confined to Davidson's understanding of the concept, see Pietroski (2009) and Ludwig (*forthcoming*).

The structure of the paper is as follows. (1) I will first informally introduce a notion of logical form as the form that matters in certain kinds of entailments, and indicate why philosophers have taken an interest in such a notion. (2) The second section develops constraints that we should arguably abide by in giving an account of logical form. (3) I then turn to Davidson's view of what is involved in giving such an account. To this end, I will try to reconstruct Davidson's view of the connection between an assignment of logical forms, a truth theory and a meaning theory. (4) Finally, I will briefly discuss possible problems of Davidson's account as developed in this paper.

The Idea of Logical Form and its Philosophical Significance

Whenever you believe something, what you believe can only be true if many other things are true as well. It is thus not hard to fathom why we should be interested in an account of the entailment relation. Such an account would further the avoidance of mistaken beliefs, by pointing to unacceptable consequences, as well as the discovery of new insights, by following wherever unshakeable assumptions may lead. But for such an account to be possible, the realm of entailments must be structured. After all, there are infinitely many things that are linked by the relation of entailment. Thus, if there were no general principles that imposed a significant order, all we could hope for would be an endless list that registered every entailment in its own right. But such a list could neither be compiled nor be used by creatures like us.

Fortunately, there does seem to be structure of the right kind. First, some terminology: An *argument* consists of a set of sentences (the premises) and a sentence (the conclusion); an argument is *valid* iff the premises entail the conclusion, i.e., iff, necessarily, if all the premises are true, the conclusion is also true. Since sentences are structured, arguments themselves have significant structure. The most fundamental observation in our quest for an account of the entailment relation is that valid arguments seem to come in two varieties. Some valid arguments are such that any argument of the same form is also valid; we may call these *formally valid*. The simplest example would be an argument that consists of a single premise and repeats this premise as its conclusion. In contrast, some valid arguments are *merely valid*: they are valid even though some arguments of the same form fail to be. Compare the argument that moves from “Peter knows that snow is white” to “Snow is white” with one that moves from “Peter believes that snow is white” to the same conclusion.

This notion of formal validity is rather shallow: to be formally valid simply means having a form such that all equiform arguments are valid. It is instructive to contrast this notion with a more ambitious one. According to a time honored distinction, while some entailments *depend* on the content of the sentences involved, other entailments hold solely *in virtue of* the form of premises and conclusion. This notion, too, has some claim to the title *formal validity*. This is not the place to provide a thorough discussion of this deeper notion (for some brief remarks, see the final section). Suffice it to say that the two notions differ conceptually and may turn out to differ extensionally. They differ conceptually for in contrast to the shallow notion, the deep notion contains an *explanatory* component expressed by “in virtue of”. And they may differ extensionally for it could turn out that the deep notion is stronger than the shallow one: the former entails the latter, while the reverse claim can reasonably be doubted. The more ambitious notion may thus be considered to be more interesting. Note, however, that the goal set out in the opening remarks will already be served by an account of the more modest conception. Where nothing hinges on the difference, I will talk indiscriminately about *formal validity*.

The project of describing the formally valid arguments of a language is only worthwhile if there are significant classes of such arguments. In pursuing this project one will thus be tempted to claim that the surface forms of some sentences are misleading. For often, on discovering certain regularities, there will be troublemakers that appear to break the rules. The sentences “This is a black horse”, “This is a beautiful horse”, “This is a strong horse” (and many more like them) all entail “This is a horse”. On these grounds, we may be attracted to classifying (A) as *formally* valid, while also noticing that (B) is not even valid:

(A) This is a black horse.
Therefore: This is a horse.

(B) This is a fake horse.
Therefore: This is a horse.

We have two choices: either we deny that (A) is formally valid, or we deny that the premise of (A) has the same form as that of (B).

Considerations like these have encouraged talk of *merely superficial* surface forms as opposed to *real*, *underlying* forms. Among philosophers, the term *logical form* has come to be used to mark this distinction. This idea has had a significant impact on philosophy, and it is easy to see why. Given a serious distinction between surface form and logical form, an account of formal validity should operate at the level of logical form. Furthermore, the notion of misleading surface form would be of interest to other philosophical inquiries, e.g. in descriptive metaphysics. Sentences whose real form differs from their surface form are bound to lead us astray. What looks like, say, reference and quantification may turn out to *merely* look like reference and quantification. Hence, in order to draw ontological conclusions from the assumed truth of certain sentences, we should be able to defend the claim that the ontologically relevant devices seemingly present are to be taken at face value. (On the connection between logical form and ontology see Davidson 1967c, 162; 1969, 166; 1970b, 181f.; 1977a, 210f.).

The distinction between mere appearance and the real thing raises a question. As Varzi (2006: 150) puts it: "The big lesson from the early days of analytic philosophy is that what you see is not what you get. But who decides what lies beneath the surface?" Consider two philosophers quarrelling about the existence of abstract objects. Both agree that "2 is a prime number" or "I have a favored color" express truths. But while one takes the apparent reference to and quantification over abstract objects at face value, the other rejects the appearances in favor of assigning logical forms in which the apparent reference and quantification is no longer present. How should we settle such a dispute?

Constraints on Accounts of Logical Form

In this section, I will do two things: (1) I will introduce the notion of an *LF-theory for a language L*. (2) I will call attention to a number of *constraints* that we should aim to meet in constructing such theories. Together these two parts yield, I hope, a fair picture of what Davidson takes to be the proper framework for evaluating claims about logical form.

(1) What is an account of logical form an account of? We must distinguish two different answers. First, one might be interested in an analysis of the *concept* of logical form. Second, we may attempt to specify, for a certain range of items, what the logical form of each of these items is. An account of the first kind will not necessarily yield one of the second. Compare the case of sentential meaning: an analysis of this concept will leave you in the dark as to what "snow is white" means. And an account of the second kind does not necessarily yield one of the first: knowing what "snow is white" means does not by itself put you into a position to say what sentential meaning is. I will call an account of the first kind a *theory of logical form* and one of the second an *LF-theory*.

Varzi's question can now be slightly rephrased: by which criteria are we to choose between competing LF-theories? Before I turn to Davidson's answer, I must address one further fundamental issue. What are the kinds of things that an LF-theory should assign logical forms to? Surely *sentences*; for sentences can be true in virtue of their logical form. But we also talk of *arguments* as being valid in virtue of their form. It may seem that we do not need to bother with arguments in addition to sentences, for once all the relevant sentences have been covered, the logical forms of the arguments will already have been settled. This is not so. If any sentences have the same logical form, these do:

(C) Socrates is wise.
(D) Plato is wise.

But if the logical form of an argument was fully determined by the logical forms of its premises and its conclusion, then the argument that has (C) as both premise and conclusion would have the same logical form as the argument that moves from (C) to (D). Yet only the first is valid in virtue of form. Hence, the logical form of an argument is not fully determined by the logical forms of its premises and conclusion. We should thus cover both sentences and arguments. When I sometimes talk as though an LF-theory was only concerned with sentences, I would ask the reader to keep this caveat in mind. (Logical form is sometimes also attributed to non-linguistic items such as *facts* or *proposition*. I will confine myself to accounts of logical form that target the linguistic realm, since Davidson's conception clearly falls into this category. Again, for surveys with a less restricted focus see Pietroski (2009) and Ludwig (*forthcoming*).)

Suppose someone suggested an LF-theory for a language-fragment *F*. What Davidson urges us to observe is that certain features of the logical forms of the sentences in *F* may only become apparent once we consider constructions *outside* of *F*. Suppose that *F* contained "Earwicker slept" and "Shem kicked Shaun" and the conjunctions that can be built from these by using "and". If we consider this fragment in isolation, it may seem that nothing would keep us from identifying the relevant primitive expressions as two predicates (one monadic, the other dyadic), three terms, and a two placed connective. But now consider

(E) Earwicker slept before Shem kicked Shaun.

(E) entails both its constituent sentences. Following Frege, we may attribute a quantificational-conjunctive structure to (E); it says, roughly, that Earwicker slept at some time, that Shem kicked Shaun at some time, and that the first time came before the second. But note that according to this account, there must be an argument place for times in "Earwicker slept" at the level of logical form. Thus, even though (E) is not in *F*, accounting for certain properties of (E) can have consequences for the assignment of forms to the sentences in *F*. Since there is no general way to predict which constructions may call for further revisions, we cannot be sure that we have covered all the bases until we have given an account for the whole language (Davidson 1970c, 138f.; 1977a, 211).

In principle, then, we should be concerned with an LF-theory that covers the whole language. What features should such a theory have to qualify as *adequate*? First we should note that, according to Davidson, an LF-theory is *only* acceptable if it is *finite*. As indicated in the introduction, an account that was essentially infinite would not be of much use to us. But as I understand Davidson, he would point to a further reason. Competent speakers are finite beings, and yet, Davidson (1967b, 123) assumes, they have some intuitive grasp of the logical forms of the sentences of their language. An LF-theory should illustrate how this is possible, and it can only do so by being finite. It is worth noting, however, that the first reason suffices to justify the desire for finite theories. From now on, I will understand “LF-theory” to mean “*finite* LF-theory”.

(2) In theory-building, we should try to meet demands like the following:

(2.1) Account for as much data as possible.

(2.2) Do so in a simple, elegant and parsimonious way.

(2.3) Other things being equal, take appearances seriously.

In the remainder of this section, I will try to spell out Davidson’s view of what these constraints entail with respect to LF-theories.

(2.1) In a sense, the first principle is obvious. But what exactly is to count as *data*? And what is it to *account for* the data? There have been conflicting views on these points and it will pay off to be precise. It is generally assumed that the relevant data has something to do with entailment relations. But arguably not every case of entailment should matter. Given that “Norma Jeane Baker = Marilyn Monroe” is necessarily true, it is entailed by “Socrates is wise”. However, most people would not consider this to be the business of an LF-theory.

Gilbert Harman (1972) has stated that an LF-theory must account for all “obvious implications”. This may rule out the case above, but it still includes more than Davidson would want to. “3 is larger than 2” obviously entails “2 is smaller than 3”, but Davidson (1967d, 125) denies that this is a matter of logical form (cp. Davidson 1970c, 143). In Davidson’s view, the data is not constituted simply by obvious entailments. As competent speakers, Davidson assumes, not only can we make intuitive judgments about the *validity* of certain arguments, but indeed about their *formal* validity (Davidson 1967a, 33). Furthermore, Davidson has the *explanatory* notion of formal validity in mind (Davidson 1967b, 125; 1969, 166; 1970c, 139, 144.). According to Davidson, these judgments constitute the primary data (this is particularly palpable in Davidson 1970c, 144; 1985, 295.). If, for example, you believe, like Davidson did, that (F) is not merely valid, but valid in virtue of its form:

(F) Sebastian strolled through the streets of Bologna at 2 a.m.
Therefore: Sebastian strolled through the streets of Bologna.

then, *ceteris paribus,* you would want an LF-theory to respect this belief. Correspondingly, if you think that a certain argument is *not* formally valid, then, *ceteris paribus*, you would want an LF-theory to respect *this* belief.

But what would an LF-theory have to do to respect these beliefs? We can distinguish two requirements, a weak and a strong one. The weak requirement states that an LF-theory should not assign the same logical form to two arguments if only one of them is intuitively formally valid. The strong requirement states that, given that an argument is intuitively formally valid, an LF-theory should assign a logical form on the basis of which we can *explain* the validity of the argument. I will go into more detail in the next section.

Let us take a brief look at the remaining two constraints. (2.2) We should not merely strive to account for the data, but we should try to do so in an *efficient* and *elegant* way. Thus, if two LF-theories accounted for the same amount of data while one used fewer and simpler principles, this constitutes a reason to prefer it to its more convoluted competitor.

(2.3) I believe Mark Sainsbury is correct when he writes: "[M]ost adherents of the methodology of logical form, including Davidson, feel happier when logical forms are as close as possible to surface forms" (Sainsbury 2005, 41). Although I do not think that Davidson stated this point explicitly, it seems rather natural. It simply states that, other things being equal, we should take appearances seriously. If, for example, there doesn't seem to be a quantifier in "Peter drank his beer" then we will not assign this sentence a quantified logical form *unless we have a good reason*. Thus, an LF-theory is better off than a competitor if it minimizes departures from surface form.

This list of constraints is not exhaustive. But it contains entries that have been, and arguably should be, central in the discussion of logical form. It may be worth pointing out that these constraints will often be in conflict. Some LF-theories may account for more data at the price of employing more complicated principles; some may gain in efficiency at the expense of assigning more deviant logical forms, etc. But here the situation is not different from other areas of scientific inquiry.

Davidson's Account of Logical Form

We have not yet touched upon the question of how we can actually go about in constructing LF-theories. I will now give a reconstruction of what I take to be Davidson's answer.

It will be helpful to distinguish three projects that we may pursue with respect to some infinite language (i.e. a language that contains infinitely many sentences): giving an LF-theory; giving a finite truth theory; giving a meaning theory. These projects are intimately intertwined in Davidson's philosophy and to appreciate Davidson's account of logical form, we need to understand his view of the connection between them. First we should note that all of the three projects necessitate the assignment of some kind of structure to the sentences of the target language. This arises from the fact that in each case we want a finite theory which covers infinitely many sentences. Central to Davidson's approach to logical form is the emphasis of a strong link between these projects. Roughly, I believe Davidson would accept the following identity claim (*modulo* certain refinements and disambiguations to be introduced later):

LF-1 The logical form of a sentence is the form assigned to it by a meaning theory, which in turn is the form assigned to it by a certain kind of truth theory.

(1) To understand LF-1, we need a grasp of what truth theories and meaning theories are. I will first briefly introduce the pertinent notions. (2) I will then sketch an argument for LF-1 in the spirit of Davidson. (3) For LF-1 to be of any real use, we need to be clear about how a truth theory assigns a form to a sentence. I will try to give an account of what this amounts to and illustrate how, with this account in hand, we can use certain truth theories to obtain LF-theories.

(1) Let me explicitly introduce some terminology. A *theory* consists of a set of *axioms* and a set of *rules*. A *theorem* of a theory is a sentence that can be derived from the axioms by the rules. (It should be noted that the rules need not constitute a *logic* in any robust sense. In particular, they are not intended to fix the *logical consequences* of the theory; more on this later.) A *T-sentence* is a sentence of the form "*S* is true iff *p*", where "*p*" does not contain semantic vocabulary. An *interpretive* T-sentence is one where "*p*" translates *S*. A *truth theory* for *L* is a theory which has, for every *L*-sentence, an interpretive T-sentence among its theorems. In the following, I will only be concerned with finite truth theories. The defining task of a *meaning theory* for an infinite language, in Davidson's view, is that of explaining how a finite being can have knowledge of such a language. In order to do so, the theory must itself be finite and it must be such that knowledge of the theory would suffice for knowledge of the language (Davidson 1973b; 1976, 171; 1977b, 215; 1990, 312).

One of the central tenets of Davidson's philosophy of language is the view that there is a close connection between a truth theory and a meaning theory. Whether this is plausible and if so, what exactly this connection could amount to, is still a matter of debate (Foster 1976; Davidson 1976; Soames 1992; 2008; Kölbel 2001; Ludwig 2002; Lepore & Ludwig 2007; Hoeltje *forthcoming*). For the purposes of this paper, it will suffice to note that a truth theory will only be able to play a central role within a meaning theory by identifying T-sentences that could serve as the basis for explicit meaning specifications. But a truth theory which employs the full deductive apparatus of, say, classical logic, will have infinitely many non-interpretive T-sentences among its theorems; e.g. it will not only yield "'Sokrates ist weise' is true iff Socrates is wise", but also "'Sokrates ist weise' is true iff (Socrates is wise and Plato = Plato)". However, only in the first case is the move to a genuine meaning-ascription - i.e. to a sentence like "'Sokrates ist weise' means that Socrates is wise" - truth preserving.

The standard way of addressing this problem makes use of the following observation: At least for simple languages, and given an adequate set of semantic axioms, it is possible to identify an *interpretive* T-sentence for each sentence of the objectlanguage in a purely syntactic way. This may be done either by specifying a set of rules that allow the derivation of a T-sentence only if it is interpretive (Larson & Segal 1995; Lepore & Ludwig 2007); or we may specify certain restrictions on how the deductive apparatus of e.g. classical logic is to be applied (Davies 1981; Kölbel 2001). Ambiguities aside, we will typically end up with a theory that has, for every objectlanguage sentence *S*, exactly one T-sentence among its theorems, and every such T-sentence will be interpretive. I will call such truth theories *canonical*. It seems clear that Davidson is thinking of truth theories along roughly these lines (Davidson 1970c, 139, 145).

(2) What speaks in favor of LF-1? Natural languages are infinite. But LF-theories must be finite. In constructing an LF-theory we will thus have to identify a finite stock of expressions and characterize these in a way that enables us to deduce a specification of a logical form for each sentence. Furthermore, if we want to account for the data in a strong sense, these specifications should give us some insight into why certain validities hold; minimally, it should be clear from an LF-theory why, given that certain sentences are true, other sentences will be true as well. Thus, an LF-theory should take as its starting point the contributions that lexical items make to a sentence's truth-conditions and issue in characterizations of sentences from which their logical profile can be read off. But, as Davidson notes, in a sense this is precisely what a standard truth theory does (Davidson 1967a, 33; 1970a, 61; 1973a, 71). First, a finite truth theory gives an account of the truth-conditionally relevant features of primitive expressions. Second, a truth theory does go some way towards specifying the logical status of objectlanguage sentences. Let me use "*x* logically entails *y*" to say that a sentence or theory *x* has some sentence *y* as a logical consequence in some standard sense (for the purposes of this paper, we may think of the relevant relation of logical consequence as being captured by classical predicate logic; but not much hangs on this). Take a theory *T* which logically entails (I) and (J) as interpretive T-sentences for (G) and (H):

(G) Sokrates ist weise und Plato ist weise.
(H) Sokrates ist weise.
(I) (G) is true iff (Socrates is wise and Plato is wise).
(J) (H) is true iff Socrates is wise.

Since the right hand side of (I) logically entails that of (J), *T* logically entails

(K) If (G) is true, then (H) is true.

Also note that, on a standard treatment of "und", there is no need to invoke (I) and (J); apart from syntax and semantic machinery, (K) will depend only on axioms for logical constants (Evans 1976 calls (K) the *validating conditional* for the inference from (G) to (H)). Furthermore, given the standard treatment of "und", *T* will logically entail that this holds for *all* sentences which stand in the relevant formal relationship; i.e., *T* will logically entail

(L) $\forall S \forall S^*$ (if $S^\frown$"und"$^\frown S^*$ is true, then S is true).

On some conceptions of logical consequence, (L) already comes fairly close to an explicit statement of the fact that (H) is a logical consequence of (G) (Quine 1936; 1954). Stronger results would issue if we used a modally strengthened truth theory or one that quantified over permitted interpretations.

In probing the informational potential of a truth theory with regard to the logical status of objectlanguage sentences, I have appealed to the *logical consequences* of such a theory. As indicated above, the class of logical consequences need not (and in the case of canonical truth theories: does not) coincide with the class of the theory's theorems in my sense. Thus, a canonical truth theory may well fail to yield (K) or (L) as theorems. Does this show that such a theory is silent on the logical connection between (G) and (H) (as Szabó (*forthcoming*) seems to hold)? Of course not; the axioms of the theory logically entail (K) and (L), and these express substantial truths about the connection in question. Anyone for whom the axioms express knowledge, and who can draw logical conclusions from her knowledge, is in a position to know that what is stated by (K) and (L). But neither the axioms nor (K) and (L) are logical truths. (Relatedly, given that we are concerned with a relation of logical consequence holding between linguistic items, Szabó seems to be wrong when he claims that "any theory logically entails the validity of all logically valid inferences". That a given logical validity is valid is not logically true. Try to logically prove something like (L) from the theory of evolution.)

There are interesting issues involved in spelling out the connection between the logical status of objectlanguage sentences and the way in which they are characterized by a truth theory (see e.g. Hoeltje 2007, Edwards 2007, Hoeltje 2012). But we need not go into the details here. What has been said so far should be sufficient to motivate the claim that in our search for LF-theories, it is natural to look to finite truth theories. (For considerations in favor of LF-1 along the lines developed here, see in particular Davidson 1968, 94f.)

LF-1 does not simply claim a connection between truth theories and logical form. So how do *meaning theories* enter into the picture? The relevance of meaning theories becomes apparent once we ask ourselves how we could get from a truth theory to an explicit LF-theory. Consider the following passage:

> [A]ll I mean by saying that 'Jones buttered the toast' has the logical form of an existentially quantified sentence, and that 'buttered' is a three-piece predicate, is that a theory of truth meeting Tarski's criteria would entail that this sentence is true if and only if there exists ... etc. (Davidson 1970c, 143).

First a small point. This passage could be read as providing a *stipulation* of the term "logical form". In this case, there would be no need to *argue* for a close connection between logical form and truth theories. However, as I have indicated above, Davidson takes the connection asserted by LF-1 to be more substantial. So let consider the passage more closely. A truth theory would entail that S is true iff there exists ... etc. by having a theorem of the form "S is true iff $\exists x$ ...". Thus, the following thought suggests itself: with respect to a truth theory T, the logical form of a sentence S is given by the right hand side of T's T-sentence for S. But, in the classical case, there is no such thing as *the* T-sentence for S. We have, of course, already noted a way around this problem: in order to do duty in a meaning theory, a truth theory should be *canonical*. We are thus led to LF-1: the logical form of a sentence *S* is given by a truth theory which can do duty in a meaning theory. More specifically, it is given by the right hand side of the canonical T-sentence for *S*.

(3) This idea needs to be both clarified and modified. First, let us state the claim at issue more precisely. I will call the right hand side of the T-sentence of a canonical truth theory *T* for a sentence *S T's counterpart for S*, and denote it by "$C_T(S)$". Given a canonical truth theory *T*, the following schema yields, for every sentence *S* in the relevant language, a claim of the form "*x* gives the logical form of *y*":

LF-2 $C_T(S)$ gives the logical form of *S*.

If we knew how to move from "*x* gives the logical form of *y*" to something of the form "*z* is the logical form of *y*", we could obtain explicit LF-theories.

The languages in which we give truth theories are usually held to rather strict standards. Typically they are assumed to be context-insensitive, non-ambiguous, with a clear syntax and semantics. Minimally, I will assume, they are languages with respect to which we can have an unproblematic notion of the *surface form* (SF) of a sentence, where the SF may be represented by a schema that results from systematically replacing all primitive expressions by placeholders of the appropriate syntactical type (note that I do not *identify* the SF of a sentence with a schema; I will briefly return to this in the final section). Thus, pretending that (M) and (N) belonged to such a language, their SF may be represented by (O):

(M) Socrates is wise and Plato is wise.
(N) Socrates is wise or Plato is wise.
(O) aF C bF.

This notion of SF accounts for the fact that, in one sense, (M) and (N) clearly have the same form. We may distinguish this notion from one that, for lack of a better term, I will call *logical* surface form (LSF). The LSF of a sentence can be represented by a schema that results from systematically replacing *non-logical* vocabulary. On a common choice of logical vocabulary, (M) and (N) do not share their LSF; only that of (M) is represented by (P):

(P) aF and bF.

We can use this notion of LSF to formulate the necessary bridge-principle: If *x* gives the logical form of *y*, then the logical form of *y* is the logical surface form of *x*. Applying this to LF-2 yields:

LF-3 The logical form of *S* is the LSF of $C_T(S)$.

We can obtain an explicit LF-theory by adding LF-3 to a canonical truth theory *T* and an account of LSF for *T*'s language. Two issues remain to be settled: What is an account of LSF for a language? Does this approach make logical form in some sense *relative* or *indeterminate*?

In justifying LF-1, I have appealed to a notion of logical consequence that can be applied to the language of a truth theory. If we are absolutists about logic, there is no need to specify "a logic" in order for this to be legitimate. However, we would want an LF-theory to be explicit. It should be clear from an LF-theory according to which the logical form of *S* is given by $C_T(S)$ what the logical properties of $C_T(S)$ are. Hence, an LF-theory must specify a logical consequence relation for the metalanguage. I will call this component of an LF-theory *a logic*. Among other things, it settles the LSFs of metalanguage-expressions.

Suppose that we could give non-equivalent canonical truth theories T1 and T2 which, together with particular logics for their languages, assigned different forms to a sentence *S*. Combining T1 and T2 with the principle LF-3 yields two non-equivalent LF-theories. Indeed, if we take the definite article at the beginning of claims of the form LF-3 seriously, these LF-theories will not only be non-equivalent, but *incompatible*. Two options suggest themselves. We could resolve the incompatibility and say that sentences have logical forms only *relative* to an LF-theory. Alternatively, we could try to decide in favor of one of the contenders by appealing to the constraints mentioned in the preceding section and say that the logical form of a sentence is the form assigned to it by the *best* LF-theory. Note, however, that this opens the possibility of logical form being *indeterminate*, for there could be several non-equivalent LF-theories that cannot be improved upon.

It has often been pointed out that, according to Davidson, logical form is *relative* - relative to the choice of a truth theory and a logic. Davidson has indeed made statements to this effect (e.g. Davidson 1970c, 139). But I think we must be careful how to understand this. It is not that "*x* is the logical form of *y*" is relative to a theory in the same way that "*x* is to the left of *y*" is to a perspective. In the latter case, we only need to specify the relevant perspective; there is no point in arguing over which perspectives are better than others. Not so in the case of logical form. Consider the following passage:

> That the two sentences ["Shem kicked Shaun" and "$\exists x$ Kicking(x, Shem, Shaun)"] have very different syntactical structures is evident; that is why the claim that the logical form is the same is interesting and, if correct, revealing. (Davidson 1970c: 145)

Davidson cannot be concerned here simply with a claim about what, *relative to a given LF-theory*, the logical form of "Shem kicked Shaun" is. For, given the theory, there are no interesting questions of this sort left open. Rather, as the surrounding text makes clear, what would be revealing would be the fact that the theory issuing in such assignments could be seen to be the best on offer. A choice of a truth theory and a logic *give rise to* claims about logical form; claims that, in light of the data, we can use to evaluate our choices.

It is a familiar phenomenon that the choice of a theory may be underdetermined by the available data. Thus, given this account of logical form, it is natural to expect logical form to be indeterminate to a certain extent. And indeed, while Davidson (1973b, 136) conjectured that no indeterminacies of logical form will arise, he later explicitly stated (1979, 228): "logical form may be indeterminate: two satisfactory theories may differ in what they count as singular terms or quantifiers or predicates, or even with respect to the underlying logic itself." We could of course also put this in terms of relativity: logical form is relative to *the best LF-theories*. This kind of relativity does not lead to triviality, since it is a non-trivial matter what the best theories are.

Above, I mentioned the need to specify not only the logical forms of sentences, but also of arguments. This is easily done on the current account. A theory *T*'s counterpart $C_T(A)$ for an argument *A* consists of *T*'s counterparts of *A*'s premises and conclusion. Given a logic, the logical form of *A* according to *T* is the LSF of $C_T(A)$, where the LSF is represented by systematically replacing non-logical vocabulary in all elements of *A*. Thus, even though "Socrates is wise" and "Plato is wise" have identical LSFs, the LSF of an argument that has the first sentence as both premise and conclusion differs from one that moves from the first sentence to the second.

Figure 1 provides an overview of how the components of this account interact. A canonical truth theory *T* will map every argument of the objectlanguage *L* onto an argument in the metalanguage *ML*. This gives us the set of *T*'s counterparts for *L*-arguments, C_T(Arguments of *L*). Let *IFV-L* be the set of (in Davidson's sense) intuitively formally valid arguments of *L*. This gives us $C_T(IFV\text{-}L)$, the corresponding set of counterparts. Let *LOG-V* be the set of elements of C_T(Arguments of *L*) that can be shown to be valid by a logic *LOG* and *IFV-ML* the set of elements of C_T(Arguments of *L*) that are intuitively formally valid. Ideally, we want the following to hold: $C_T(IFV\text{-}L) = LOG\text{-}V = IFV\text{-}ML$. The first identity ensures that *T*, in combination with *LOG*, yields all and only validating conditionals for arguments considered to be formally valid (but let's not quibble about arguments with infinitely many premises). The second identity ensures that the logic is maximally acceptable.

Criticisms

(1) Davidson explicitly appeals to the notion of *validity solely in virtue of logical form*. This is a close relative of the notion of *validity solely in virtue of meaning*. However, the latter notion has received substantial criticism (Quine 1951, Boghossian 1997, Williamson 2007). It is thus not surprising that the weapons that have been used to assault analyticity would also be turned against the explanatory notion of formal validity. Szabó (*forthcoming*) argues that there is no robust sense in which natural languages contain inferences that hold solely in virtue of logical form. I want to briefly indicate where I see room for maneuver.

Szabó raises two *triviality objections*. Here is the first: No inference is valid solely in virtue of its logical form, because matters of fact are always also relevant. This is witnessed by the absurdity of saying (cp. Williamson 2007, 59):

(Q) No matter whether every father in fact is a father or a mother, the inference "Alex is a father; therefore Alex is a father or a mother" is valid simply because of its form.

Let us distinguish two proposals about what it means to say that *A* is valid solely in virtue of its logical form: (i) There is an acceptable explanation of *A*'s validity that draws only on *A*'s logical form. (ii) Every acceptable explanation of *A*'s validity draws only on *A*'s logical form. The argument, if it works at all, works against (ii); but proponents of logical form can be content with (i).

Pretend you do not understand a word of German. The following does *not* sound absurd (or, at least, it does not sound absurd for reasons that the proponent of validity in virtue of form needs to be troubled by):

(R) Never mind whether every father in fact is a father or a mother; "Alex ist ein Vater" entails "Alex ist ein Vater oder Alex ist eine Mutter" because, *first* the conclusion is a sentence formed by "oder" that contains the premise as a constituent sentence, and *second* a sentence formed by "oder" is true if any of its constituent sentences are.

Where does the absurdity of (Q) come from? Given that you understand English, you know that whether

(S) every father is a father or a mother

is relevant to the validity-claim in (Q). But on the proposed understanding of "solely in virtue of form", the relevance of (S) need not be disputed. Almost no acceptable explanation cites everything that is relevant. We are bound to slide into absurdity if, as in (Q), we preface an explanation by calling attention to a fact *known* to be relevant while at the same time insisting that we shouldn't worry about whether it is actually the case. But this does not show that the explanation wasn't acceptable to begin with. The explanation embedded in (R) is acceptable and it draws only on logico-formal features. It is exactly the kind of explanation that is made available by a truth theory. Moreover, in many other cases of valid inferences, this kind of explanation is not available. The corresponding class of arguments is thus neither empty nor trivial.

Szabó's second objection states that even inferences involving "or"-introduction are not valid solely in virtue of form, for they rely on the definition of "or". I think this is correct as far as it goes: the explanation in (R) would cease to be acceptable if we did not specify the particular contribution of “or”. However, “or” is a logical constant, and the claim is that the inference is valid in virtue of its *logical* form; i.e. that there is an acceptable explanation of its validity that mentions only formal features *and the function of logical constants*. This relies on an antecedent characterization of logical constants. But it does not require a very strong or absolute notion. Even a simple list, if carefully compiled, would yield an interesting notion.

(2) On the account developed above, the logical form of a sentence is identified with the logical surface form of another sentence. This idea is rather old fashioned. (While e.g. the account proposed by Kalish (1952) is similar in certain respects, there are also important differences.) According to Ludwig (*forthcoming*), proposals of this kind run the risk of making logical form language-relative to an undesirable degree. Depending on whether we use infix, prefix or postfix notation, we might end up with the following counterparts for (T):

(T)	Socrates is wise and Plato is wise.
C_{T1}(T)	Wise(Socrates)&Wise(Plato).
C_{T2}(T)	&Wise(Socrates)Wise(Plato).
C_{T3}(T)	Wise(Socrates)Wise(Plato)&.

On some natural notion of form, C_{T1}(T) - C_{T3}(T) have *different* forms. But then, if the logical form of a sentence is the form of its counterpart, we would have to say that theories assigned different logical forms simply on the grounds that one uses infix, the other prefix notation, and surely that shouldn’t make a difference. I agree that it shouldn’t. But there is, I think, a straightforward notion of form available on which C_{T1}(T) - C_{T3}(T) all exhibit the same form. All of them are composed out of two sentences governed in the same way by the same kind of connective. While it is correct that we arrive at different *schemas* if we try to give the LSF of C_{T1}(T) - C_{T3}(T) this does not mean that their LSFs differ. A schema *represents* an LSF, and different schemas may represent the same LSF. There is no way to simply write down the LSF itself, because a form is not the kind of thing that can be written down.

Lepore & Ludwig (2002) prefer to do away with talk of logical forms as objects in their own rights and are content with an equivalence relation of *sameness of logical form*. According to their proposal, S1 and S2 have the same logical form if some interpretive truth theories have corresponding canonical proofs for their T-sentences. For reasons of space, I cannot go into the details of their account here. Let me just point out that I believe they must strengthen their notion of corresponding proofs in order to differentiate between sentences like “Socrates=Socrates or Plato=Plato” and “Socrates=Plato or Plato=Socrates”. A natural way of doing so would require a canonical proof of a T-sentence to mirror the structure of the T-sentence’s right hand side. Such an amended approach may turn out to be equivalent to the one developed here. Since I see no reason to doubt the existence of LSFs, I see no harm in appealing to them directly.

(3) Truth theories play a central role within Davidson’s account of logical form. However, according to the way I have set things up, not all truth theories are on par when it comes to revealing logical form. Rather, each pair of a truth theory and a logic gives rise to certain

claims about logical form and we can bring additional constraints to the table to evaluate these claims. But we should note two things: First, if we did not impose significant constraints on LF-theories, this would probably lead to a high degree of indeterminacy. Second, we may wonder whether, in light of Davidson's larger philosophical enterprise, it is legitimate for him to appeal to these further constraints. Let me briefly elaborate the relevant worry and then indicate how it might be disarmed.

Suppose we could give two canonical truth theories T1 and T2 which produced the following counterparts for (U) and (V) (I use "Boiling$_t$" and "Boiling$_i$" as event-predicates corresponding to transitive and intransitive uses of "boiled" and I will pretend that boiling something simply consists in causing it to boil):

(U) Chloe boiled the lobster.
(V) The lobster boiled.
C_{T1}(U) $\exists$e Boiling$_t$(e, the lobster, Chloe)
C_{T1}(V) $\exists$e Boiling$_i$(e, the lobster)
C_{T2}(U) $\exists$e (Agent(e, Chloe) & $\exists$e* (Boiling(e*) & Patient(e*, the lobster) & Cause(e, e*))).
C_{T2}(V) $\exists$e* (Boiling(e*) & Patient(e*, the lobster)).

Which theory gives the logical form of (U)? Given that we can appeal to the constraints developed in section 2, we have a certain grip on this question. In combination with standard logic, T2, but not T1, yields a validating conditional for the inference from (U) to (V). Thus, according to the LF-theory associated with T2, but not according to that associated with T1, the inference is a matter of logical form. If you consider the move from (U) to (V) to be valid in virtue of form, this speaks in favor of T2. If you do not, you may prefer T1, since surely C_{T2}(U) constitutes an even further departure from surface form than C_{T1}(U).

There is ample textual evidence for the claim that Davidson in fact does appeal to the constraints I have made explicit. In particular, it is clear that Davidson takes judgments about formal entailment relations to be central. But this, we might worry, involves a semantic notion as applied to the objectlanguage and we may think that this is off limits for Davidson (cp. Jackson 2007, 364f.). As Jackson points out, without appealing to substantial constraints, we have no basis to decide between T1 and T2. Thus, given the restrictions imposed by Davidson's larger philosophical enterprise, logical form may turn out to be highly indeterminate - perhaps much more so than Davidson was willing to countenance.

I believe this worry is unsubstantiated. The information we can appeal to in order to choose between T1 and T2 need not be semantic information about the objectlanguage. Rather, what may count in favor of one theory over the other is that it complies better with *speaker-judgments* about which entailment relations hold as a matter of form. Thus, what counts as data is not the (assumed) fact that e.g. (U) formally entails (V), but rather that speakers *believe* or *accept* that (U) formally entails (V). However, as Pagin (2006) points out, this latter fact is *not*, strictly speaking, a semantic fact about the objectlanguage, but a *psychological* fact about the speakers of the language.

Let me end with some brief concluding remarks. I have tried to give a precise explication of Davidson's conception of logical form. I am, however, sympathetic to those who feel the need for further clarifications. First, since the conception relies on some intuitive notion of validity solely in virtue of logical form, it is reasonable to demand a thorough account of what this amounts to. I have briefly indicated what I believe would be involved in such an account, but I am aware that much more would need to be said to satisfy the skeptic. Second, even granted a viable account of formal validity e.g. in terms of explanations that draw only on logico-formal material, we may have doubts about the assumption that ordinary speakers apply such a notion to the arguments of their language. Third, even granted that speakers do apply such a notion, it may not be entirely clear how a radical interpreter should be able to find out when they do - in particular how she could distinguish speaker's accepting that a certain argument is *formally* valid from their accepting that it is (merely) valid.

Although these are serious challenges, I am optimistic that they can be met (for lack of space, however, I will not attempt to show this here). In any case, I feel that these remaining issues should not keep us from appreciating the fact that Davidson has drawn attention to a theoretical framework in which "the concept of logical form can be clarified and thus defended" (Davidson 1970c, 138). In particular, Davidson's emphasis on framing questions about logical form in terms of truth theories for natural languages that are backed by speaker judgments about formal validities seems to me to have had a lasting influence, and rightly so. And while Davidson was surely not alone in developing an account of logical form along roughly these lines, one would be hard pressed to find someone who rivals him in having made such systematic and important contributions in this connection.[1]

References

Boghossian, Paul. 1997. "Analyticity." In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright, 331–368. Oxford: Blackwell.

Davidson, Donald. 1967a. "Truth and Meaning." In Davidson 2001b.

- 1967b. "The Logical Form of Action Sentences." In Davidson 2001a, 105–122.
- 1967c. "Causal Relations." In Davidson 2001a, 149–162.
- 1967d. "Reply to Castaneda on Agent and Patient." In Davidson 2001a, 125–127.
- 1968. "On Saying That." In Davidson 2001b, 93–108.
- 1969. "The Individuation of Events." In Davidson 2001a, 163–180.
- 1970a. "Semantics for Natural Languages." In Davidson 2001b, 55–64.
- 1970b. "Events as Particulars." In Davidson 2001a, 181–188.
- 1970c. "Reply to Cargile." In Davidson 2001a, 137–146.
- 1973a. "In Defence of Convention T." In Davidson 2001b, 65–75.
- 1973b. "Radical Interpretation." In Davidson 2001b, 125–139.

[1] For helpful discussions of a previous draft of this paper I am very grateful to Nick Haverkamp, Benjamin Schnieder, Moritz Schulz, Thomas Spitzley, Alex Steinberg and Raphael van Riel.

- 1976. “Reply to Foster.” In Davidson 2001b, 171–179.

- 1977a. “The Method of Truth in Metaphysics.” In Davidson 2001b, 199–214.

- 1977b. “Reality Without Reference.” In Davidson 2001b, 215–225.

- 1979. “The Inscrutability of Reference.” In Davidson 2001b, 227–241.

- 1985. “Adverbs of Action.” In Davidson 2001a, 293–304.

- 1990. “The Structure and Content of Truth.” *The Journal of Philosophy*, 87: 279–328.

- 2001a. *Essays on Actions and Events*. New York: Clarendon Press.

- 2001b. *Inquiries into Truth and Interpretation*. New York: Clarendon Press.

Davies, Martin. 1981. *Meaning, Quantification, Necessity*. London: Routledge & Kegan Paul.

Edwards, Jim. 2007. “Response to Hoeltje: Davidson Vindicated?” *Mind*, 116: 131–141

Foster, John. 1976. “Meaning and Truth Theory.” In *Truth and Meaning*. edited by Gareth Evans and John McDowell, 1976, 1–32. Oxford: Oxford University Press.

Harman, Gilbert. 1972. “Logical Form.” *Foundations of Language*, 9: 38-65.

- Forthcoming. “Davidson’s Contribution to the Philosophy of Language.” In *Donald Davidson on Truth, Meaning, and the Mental*, edited by Gerhard Preyer. Oxford: Oxford University Press.

Hoeltje, Miguel. 2007. “Theories of Meaning and Logical Truth: Edwards versus Davidson.” *Mind*, 116: 121–129.

- Forthcoming. “Lepore and Ludwig on ‘Explicit Meaning Theories’.” Forthcoming in *Philosophical Studies*.

- 2012. *Wahrheit, Bedeutung und Form. Eine Auseinandersetzung mit dem Davidson’schen Programm*. Münster: mentis.

Jackson, Brendan. 2007. “Beyond Logical Form.” *Philosophical Studies*, 132: 347–380, DOI 10.1007/s11098-005-2466-3.

Kalish, Donald. 1952. “Logical Form.” *Mind* 61: 57–71.

Kölbel, Max. 2001. “Two Dogmas of Davidsonian Semantics.” *Journal of Philosophy*, 98: 613–635.

Larson, Richard and Gary Segal. 1995. *Knowledge of Meaning*. Cambridge: MIT.

Lepore, Ernest and Kirk Ludwig. 2007. *Donald Davidson’s Truth-Theoretic Semantics*. Oxford: Oxford University Press.

- 2002. “What is Logical Form?” In *Logical Form and Language*, edited by Gerhard Preyer and Georg Peter, 54–90. Oxford: Oxford University Press.

Ludwig, Kirk. 2002. “What is the Role of a Truth Theory in a Meaning Theory?” In *Meaning and Truth: Investigations in Philosophical Semantics*, edited by Joseph Keim Campbell, Michael O’Rourke, and David Shier, 142–163. New York: Seven Bridges Press.

- Forthcoming. “Logical Form.” In *The Routledge Companion to the Philosophy of Language*, edited by Gillian Russell and Delia Graff.

Pagin, Peter. 2006. “Meaning Holism.” In *The Oxford Handbook of Philosophy of Language*, edited by Ernest Lepore and Barry Smith, 213–232. Oxford: Oxford University Press.

Pietroski, Paul. 2009. “Logical Form.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, URL = <http://plato.stanford.edu/archives/fall2009/entries/logical-form/>.

Quine, Willard Van Orman. 1936. “Truth by Convention.” In *The Ways of Paradox and Other Essays*, 70–99. New York: Random House.

- 1951. “Two Dogmas of Empricism.” *The Philosophical Review*, 60: 20–43.

- 1954. “Carnap and Logical Truth.” In *The Ways of Paradox and Other Essays*, 100–125. New York: Random House.

Sainsbury, Mark. 2005. *Reference Without Referents*. Oxford: Oxford University Press.

Soames, Scott. 1992. "Truth, Meaning, and Understanding." *Philosophical Studies*, 65: 17–35.

- 2008. "Truth and Meaning in Perspective." *Midwest Studies in Philosophy*, 32: 1–19.

Szabó, Zoltan Gendler. *Forthcoming*. "Against Logical Form." In *Donald Davidson on Truth, Meaning, and the Mental*, edited by Gerhard Preyer. Oxford: Oxford University Press.

Varzi, Achille. 2006. "Review of: From an Ontological Point of View." *Philosophical Books*, 47: 503–538.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.