Theories Of Meaning And Logical Truth: Edwards Versus Davidson

MIGUEL HOELTJE

In a recent paper[1] Jim Edwards is, among other things, concerned with the following passage from Davidson's programmatic essay *Truth and Meaning*:

> A truth definition does not distinguish between analytic sentences and others, except for sentences that owe their truth to the presence alone of the constants that give the theory its grip on structure: the theory entails not only that these sentences are true but that they will remain true under all significant rewritings of their non-logical parts. A notion of logical truth is thus given limited application, related notions of logical equivalence and entailment will tag along. (Davidson 1967, p. 33)

I will call the following two claims *Davidson's claims*: *(i)* A theory of truth which serves as a theory of meaning *TM* for a language *L* entails which sentences of *L* are logical truths (relative to the theory's choice of logical constants). *(ii) TM* also entails which sentences of *L* stand in a relation of logical consequence (relative to the theory's choice of logical constants). Edwards presents an argument which is supposed to show that Davidson's claims are *false* when 'entails' is taken to express the notion of logical consequence; that is, Edwards argues that, given a logical truth *LT* of the object-language L_0, it is *not* a logical consequence of *TM* that *LT* is a logical truth of L_0. I will only be concerned with this strong reading of Davidson's claims: *(i*)* It is a *logical consequence* of *TM* which sentences of *L* are logical truths. *(ii*)* It is also a *logical consequence* of *TM* which sentences of *L* stand in the relation of logical consequence.

In this paper I will show that Davidson is right. A Davidsonian theory of meaning like the one presented by Edwards provides the resources needed to derive in first-order predicate logic which sentences of the object-language are logical truths. In particular, this will cover all sentences usually considered to be logical truths of propositional or predicate logic. This — as Davidson says — also provides the means necessary for deriving which sentences of the object-language stand in the relation of logical consequence. Since we have soundness for first order logic, something derivable from a theory in first-order logic *is* a logical consequence of that theory, and thus Davidson's claims are vindicated.

---

[1] *Theories of Meaning and Logical Constants: Davidson versus Evans*, *Mind* 111, 2002.

I will proceed as follows. *(i)* I will first give an example of a derivation showing that it is a logical consequence of the theory of meaning employed by Edwards (*TM*) that a sentence of the form 'S v ¬S' is a logical truth of the object-language. The derivation will also make clear that and how this generalizes to logical truths of different forms. *(ii)* I will then compare my argument to the one Edwards gives and explain why the latter fails to prove its point.

1. Vindicating Davidson

In order to make a comparison between the argument Edwards gives and the one I am proposing as easy as possible, I will use the same apparatus as Edwards does. Thus, I will take the object-language L_0 to be a fragment of English lacking a truth predicate, augmented with standard symbols of first-order logic, and I will take the metalanguage L_1, in which *TM* is to be formulated, to be a fragment of English augmented with a truth predicate '$Tr_0(x)$ under *i*' ascribing truth in L_0 under an interpretation *i*, the concatenation function '^', and logical symbols. In particular, both L_0 and L_1 will contain a symbol for negation and one for disjunction; I will use '¬' and 'v' in the case of L_0 and 'not-' and 'or' in the case of L_1. I take *TM* to be a set containing several lexical and recursive axioms of which I will only list the two treating the object-language expressions 'v' and '¬' (cp. Edwards 2002, p. 251):

(A1) $\forall i \forall A \forall B$ ($Tr_0(A$^'v'^$B)$ under $i \leftrightarrow$ ($Tr_0(A)$ under i or $Tr_0(B)$ under i)).

(A2) $\forall i \forall A$ ($Tr_0($'¬'^$A)$ under $i \leftrightarrow$ not-($Tr_0(A)$ under i)).

The issue at hand is whether it is a logical consequence of *TM* that a sentence of L_0 like the following is a logical truth:

(S1) Snow is white v ¬ snow is white.

I will occasionally abbreviate S1 to 'S v ¬S', but we have to keep in mind that this is not a schema (since schemas are not *true*, neither logically nor otherwise). Now, the notion of a logical truth in this context is simply the notion of a sentence that is true under all permitted interpretations of the language.[2] Thus, the question whether it is a logical consequence of *TM* that S1 is a logical truth can be reformulated. It is the question whether it is a logical consequence of *TM* that S1 is true in L_0 under all permitted interpretations. Using standard logical notation, let us shorten 'S1 is true in L_0 under all permitted interpretations' to

---

[2] Edwards, at least, agrees that this is the relevant notion of logical truth; cp. Edwards 2002, p. 252.

(S2) $\forall i$ (Tr_0('S v ¬S') under i).

If it can be shown that S2 is a logical consequence of *TM*, Davidson's claim is vindicated with respect to the logical truth of 'S v ¬S'. I will first prove that S2 is in fact a logical consequence of *TM*, and then show how this generalises to object-language logical truths of different forms. The argument I am proposing is rather short:

1. S2 can be derived from the axioms of *TM*.
2. Everything that can be derived from the axioms of *TM* in first order logic *is* a logical consequence of *TM* (*soundness*).

Therefore:

C. S2 is a logical consequence of *TM*.

The second premise is a well established result and thus should not be questioned. The first premise can be shown to be correct by giving the derivation. This derivation makes use of the two axioms A1 and A2, an instance of the valid schema 'P or not-P', and a way to handle the concatenation function:[3]

1	(1)	$\forall i \forall A \forall B$ ($\mathrm{Tr}_0(A\char94\text{'}v\text{'}\char94B)$ under $i \leftrightarrow$ ($\mathrm{Tr}_0(A)$ under i or $\mathrm{Tr}_0(B)$ under i))	Assumption
2	(2)	$\forall i \forall A$ (Tr_0('¬'$\char94A$) under $i \leftrightarrow$ not-($\mathrm{Tr}_0(A)$ under i))	Assumption
1	(3)	Tr_0('S'$\char94$'v'$\char94$'¬'$\char94$'S') under $\alpha \leftrightarrow$ (Tr_0('S') under α or Tr_0('¬'$\char94$'S') under α)	1 $\forall$E
2	(4)	Tr_0('¬'$\char94$'S') under $\alpha \leftrightarrow$ not-(Tr_0('S') under α)	2 $\forall$E
	(5)	Tr_0('S') under α or not-(Tr_0('S') under α)	P or not-P
6*	(6)	Tr_0('S') under α	Assumption*
6*	(7)	Tr_0('S') under α or Tr_0('¬'$\char94$'S') under α	6 or-I
8*	(8)	not-(Tr_0('S') under α)	Assumption*
2, 8*	(9)	Tr_0('¬'$\char94$'S') under α	4, 8 $\leftrightarrow$E
2, 8*	(10)	Tr_0('S') under α or Tr_0('¬'$\char94$'S') under α	9 or-I
2	(11)	Tr_0('S') under α or Tr_0('¬'$\char94$'S') under α	5; 6, 7; 8, 10 or-E
1, 2	(12)	Tr_0('S'$\char94$'v'$\char94$'¬'$\char94$'S') under α	3, 11 $\leftrightarrow$E
1, 2	(13)	$\forall i$ Tr_0('S'$\char94$'v'$\char94$'¬'$\char94$'S') under i	12 $\forall$I
	(14)	'S'$\char94$'v'$\char94$'¬'$\char94$'S' = 'S v ¬S'	Def. '$\char94$'
1, 2	(15)	$\forall i$ (Tr_0('S v ¬S') under i)	14, 13 =E

The last line *solely* depends on the two axioms, no other assumptions are used. Thus the first premise of my argument should be beyond doubt. It is important to note that in proving this premise we rely on nothing that should be questioned in the context of Davidsonian semantics; the technical apparatus employed is simply first order logic plus a way to handle the sign for concatenation

---

[3] Let me add some remarks which should make the derivation more readable. In line 1 and 2 I introduce the two relevant axioms as assumptions. In line 3 and 4 I eliminate all universal quantifiers, substituting '"S"' for '*A*', '"¬"$\char94$"S"' for '*B*' and 'α' for '*i*'.

(which has to be available to the Davidsonian in any case). Since the argument given above is valid and has provably true premises, we should accept its conclusion. Since nothing more than the truth of the conclusion is needed for (the relevant instance of) Davidson's claim to be correct, we should accept (the relevant instance of) Davidson's claim. But once we accept this line of reasoning for a particular instance, we are in a position to see that the Davidsonian claim can be defended across the board.

In general, we will be able to provide a derivation of this kind for *any* truth of propositional and predicate logic as long as *TM* covers the standard connectives and quantifiers; since the axioms of *TM* provide biconditionals that allow us to move from, for example, an object-language disjunction/negation to its metalanguage counterpart and *vice versa*, and since we have the corresponding metalanguage derivations as a simple matter of logic, we can derive for an arbitrary interpretation α that object-language sentences corresponding to logical truths of the metalanguage are true under α. Since the axioms themselves do not mention any particular interpretation, we are free to quantify into the position of 'α' and thus arrive at the desired result. The same holds, *mutatis mutandis*, for quantified logical truths like '$\forall x$ (x is a man v $\neg$ x is a man)' (since the quantifier-axioms of *TM* are somewhat more complicated than those for the sentential connectives, I spare the reader the trouble to go through the respective derivations).

We thus have for every logical truth *LT* of L_0 a theorem of *TM* which says that *LT* is true in L_0 under all interpretations. As Davidson said, a notion of logical consequence will now 'tag along': S* is a logical consequence of S if and only if the corresponding instance of the following schema can be derived from *TM*:

$\forall i$ Tr_0('$\neg$(S & $\neg$S*)') under *i*.

I conclude that it is a logical consequence of *TM* which sentences of L_0 are logical truths and which sentences of L_0 stand in a relation of logical consequence; and this is precisely what Davidson claimed.

2. Edwards's argument

As we have seen, it can be derived from *TM* that S1 is a logical truth of L_0. Thus it is a logical consequence of *TM* that S1 is a logical truth of L_0. But Edwards claims to have *shown* that it cannot be a logical consequence of *TM* that 'S v $\neg$S' is a logical truth of L_0. Edwards and Davidson cannot

both be right. Since the above argument shows that Davidson is right, one might wonder where Edwards went wrong. It is this question that I shall be concerned with in this section.

Let me briefly review Edwards's argument. Using '$x \models_{Ln} y$' to say that y is a logical consequence of x in L_n and accordingly '$\models_{Ln} x$' to say that x is a logical truth of L_n, we can reformulate S2 — the claim that 'S v ¬S' is a logical truth of L_0 — as

(S3) $\models_{L0}$ 'S v ¬S'.

Since *TM* itself is formulated in L_1, the claim that S3 is a logical consequence of *TM* will have to be formulated in L_1's metalanguage L_2:

(S4) *TM* $\models_{L1}$ ' $\models_{L0}$ "S v ¬S" '.

S4 is *false* in L_2 if there is a permitted interpretation $i^\dagger$ of the constants of L_1 under which the axioms of *TM* remain true in L_1 while ' $\models_{L0}$ "S v ¬S" ' turns out to be false, and Edwards tries to provide just such an interpretation. Now, a *permitted* interpretation of L_1 is one that does not change the interpretation of L_1's logical constants; and the interpretation that Edwards suggests is one where *(i)* ' "v" ' — a *name* in L_1 of the logical constant 'v' of L_0 — is assigned the logical constant '&' of L_0 and *(ii)* the predicate '$Tr_0(x)$ under i' receives an extension that is equivalent to the extension of 'not-($Tr_0(x)$ under i)' under the intended interpretation.[4] Thus, while A1, the axiom for 'v', remains *true* under this interpretation, it seems that under this interpretation ' $\models_{L0}$ "S v ¬S" ' means what ' $\models_{L0}$ "S & ¬S" ' means under the intended interpretation. But since 'S & ¬S' is *not* a truth, let alone a *logical* truth of L_0, this seems to constitute a counterexample to S4. Thus, Edwards concludes, since S4 is false, it is *not* a logical consequence of *TM* that 'S v ¬S' is a logical truth of L_1.

This argument might be attacked at various stages. I will consider two ways of showing that it fails to prove its point. *(a)* First I shall try to show that given the interpretation of L_1 that Edwards proposes there is a reasonable reading under which S4 indeed turns out to be *true*, not false. This will rely on taking ' $\models_{Ln}$' to be what Tarski calls a *defined sign*.[5] *(b)* The second attack is of a more general nature and, in my opinion, more powerful. Here I will argue that given the *form* of the question at hand, the *kind* of argument Edwards proposes cannot be but insufficient. According to

---

[4] Where '&' of L_0 receives the obvious recursive axiom in *TM*. The additional features of the proposed interpretation $i^\dagger$ that Edwards mentions are of no concern to the present point.

[5] This way of putting the point was suggested by Jim Edwards.

this attack, even if Edwards succeeded in showing that S4 was false, this would not constitute a reason to abandon Davidson's claim.

(a) Let us consider the first attack. For the sake of the argument, let us simply grant that under the interpretation proposed by Edwards the axioms of *TM* indeed turn out to be true. Everything now hinges on the question whether ' $\models_{L0}$ "S v ¬S" ' is false under this interpretation; if it is, Edwards is right with respect to the falsity of S4. I want to argue that there is a natural reading of ' $\models_{L0}$ "S v ¬S"' under which — given *Edwards's* interpretation — it turns out to be *true* rather than false. This reading takes ' $\models_{L0}$' to be what Tarski calls a *defined sign*. We are familiar with the idea of thinking of technical expressions like ' $\models$' as mere shorthand for something complex, something we could spell out in more primitive vocabulary. Such defined signs must be replaced by the primitive defining expressions before we test for logical consequence or logical truth. But what would — in the current context — be the natural way of spelling ' $\models_{L0}$ "S v ¬S"' out? Edwards's own remarks suggest that we simply end up with my S2 from section 1 above:[6]

> (S2) $\forall i$ (Tr$_0$("S v ¬S") under i).

But now it is crucial to note that S2 is *true* under the interpretation Edwards proposes. For under this interpretation S2 means what ' $\forall i$ not-(Tr$_0$("S & ¬S") under i)' means under the intended interpretation, and this is a *truth* of L$_1$ (to the effect that for all interpretations of L$_0$ 'S & ¬S' is *not* true in L$_0$). Furthermore, not only does *Edwards's* proposed interpretation not falsify S2, *no* relevant interpretation of L$_1$ does. This is secured by the interplay between the truth predicate and the logical constants; in order to provide an interpretation of L$_1$ that re-assigns the names of the L$_0$-logical constants *and* lets the axioms treating those logical constants come out true, you have to reinterpret the truth predicate. But doing this will counter the effect of re-assigning the names of the logical constants; given the axioms are true, sentences like S2 will always come out true as well (even though they might differ in meaning from the meaning they have under the intended interpretation). In the light of the argument that I presented in section 1, this should not be surprising; since S2 is a logical consequence of *TM*, there is no permitted interpretation of L$_1$ under which the axioms of *TM* are true while S2 is false.

As I said, this reply relies on the assumption that ' $\models$' is to be treated as a defined term. Although this assumption may be rather natural, it does not seem impossible to simply deny it. This brings us

---

[6] See e.g. p.253 where Edwards says about a sentence of the form ' $\models$ "S"' that it 'states that there is no permitted interpretation of the sentence "S" of L$_0$ such that it is not true-in-L$_0$'.

to the second reply to Edwards's argument, according to which the alleged falsity of S4 is not relevant to the question at hand.

(b) According to the second attack, Edwards's argument is flawed at a very general level. The flaw can be uncovered by taking a closer look at the form of the question at hand. The question with which we started is the following:

(Q) Is it a logical consequence of *TM* that 'S v ¬S' is a logical truth of L_0?

Let us represent the form of this question by the following schema:

(QF) Is it a logical consequence of *SET* that *p*?

The first important thing to note is that while the placeholder 'SET' in QF stands in the position of a singular term referring to a set of sentences, '*p*' stands in the position of *a sentence that is used*, not mentioned. But the relation of logical consequence at issue is one that is supposed to hold between sets of sentences (the premises) and *a sentence* (the conclusion). Thus it might seem unclear how to understand Q in terms of the relevant notion of logical consequence. Edwards seems to agree that one would miss the point of instances of QF by taking them to amount simply to the question whether the sentence which replaces '*p*' is a logical consequence of SET – otherwise he would be concerned with the question whether the sentence

(S5) 'S v ¬S' is a logical truth of L_0

is a logical consequence of *TM*, not with the question whether his more technical (and closely related but not identical) sentence

(S6) $\models_{L_0}$ 'S v ¬S'

is a logical consequence of *TM*. What *is* the point of a question that takes the form QF? It seems to me that in order to understand questions of this form in terms of a relation of logical consequence that holds between linguistic items such as sentences, a natural way is to think of them along the following lines: a positive answer to such a question is substantiated given that *some* sentence which *is* a logical consequence of SET says that *p*. Thus we can restate QF:

(QF*) $\exists S$ (*S* is a logical consequence of *SET* & *S* means that *p*)?

According to this proposal one way of showing that a certain set of sentences has it as a logical consequence *that* Kaa is a snake is to show that the sentence 'Kaa is a serpent' (which happens to mean that Kaa is a snake) is a logical consequence of that set. But if this is the right way of understanding questions like Q, it immediately becomes apparent where Edwards's argument goes wrong. Although Edwards's S6 is relevant in the sense that it is one way of putting the claim that 'S v ¬S' is a logical truth of L_0, it is merely *one* out of many. Or, to put it somewhat differently, although showing that S6 is a logical consequence of *TM* would be a way of proving Davidson *right*, showing that S6 is not a logical consequence of *TM* is *not* sufficient for proving Davidson *wrong*. You cannot falsify an existential claim by giving one false instance; merely presenting *one* sentence which says that 'S v ¬S' is a logical truth of L_0 and which is *not* a logical consequence of *TM* falls a long way short of providing a good argument for a negative answer to Q – there might be *other* sentences which also say that 'S v ¬S' is a logical truth of L_0 and which *are* logical consequences of *TM*. I take S2 to be such a sentence.

Before I sum up, let me stress that although I made explicit use of semantic notions like that of synonymy in rebutting Edwards's argument, this need not worry the Davidsonian. In section 1 I showed that '$\forall i$ (Tr_0("S v ¬S") under i)' (which is just a shorter version of '"S v ¬S" is true in L_0 under every permitted interpretation') is a logical consequence of *TM*, and in doing so I did not use any questionable means whatsoever; first order logic, soundness and the concatenation function are all available to the Davidsonian. It is our *understanding* of this result which shows us that Davidson's claims are correct. It was only in explaining where Edwards went wrong that questions of synonymy arose. After all my preferred way of putting it:

(I) '$\forall i$ (Tr_0("S v ¬S") under i)' is a logical consequence of *TM*

has as good a claim of being faithful to Davidson's idea that the logical truth of 'S v ¬S' is a logical consequence of *TM* as Edwards's

(JE) $TM \models_{L1}$ ' $\models_{L0}$ "S v ¬S" '.

Since I can be shown to be correct without invoking any notions that are unavailable in the context of Davidsonian semantics, Davidsonians can and should prefer my way of putting the relevant claim – quite independently of the question whether they agree with my analysis of Edwards's error.

Thus, I take Davidson's claims to be vindicated. It is worth noting that in view of Edwards's overall project, he can in fact welcome my results: *if* he is right in claiming that Evans's criterions for a transcendent theory of meaning yield a theory in the Davidsonian style as presented by Edwards, *then* my results show that such a theory will also provide transcendent notions of logical truth and consequence. I do accept the conditional, but in this short paper I have neither denied nor asserted its antecedent.[7]

MIGUEL HOELTJE

PhD-Programme
Philosophisches Seminar
University of Hamburg

References

Davidson, Donald 1967: 'Truth and Meaning', in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press 1984, pp. 18–36. Originally published in *Synthese*.

Edwards, Jim 2002: 'Theories of Meaning and Logical Constants: Davidson versus Evans', *Mind*, 111, pp. 249–79.