Fostering Liars[1]

Paul M. Pietroski

Human children regularly acquire languages, spoken or signed, that connect meanings of some kind with pronunciations. These languages, let's call them Slangs, are distinctively human; and our capacity to acquire them is somehow interwoven with our distinctively human cognitive capacities. It isn't obvious what Slangs are, and this invites inquiry. But with regard to the meanings that Slangs connect with pronunciations, an orthodoxy has emerged.

Davidson (1967a) conjectured that for each language **L** of the relevant human kind, a suitably formulated Tarski-style theory of *truth* for **L** can "do duty" as a theory of *meaning* for **L**, at least if (i) the specification of which sentences are true-in-**L** can be plausibly extended to cover non-declarative sentences, and (ii) the language-relative notion of truth is relativized to speakers, times, things demonstrated, and perhaps other aspects of the contexts in which sentences can be used. Lots of good work has been framed in these terms; and over the last fifty years, it has become standard to assume that declarative sentences have compositionally determined truth conditions, and that understanding the expressions of a Slang—knowing what they mean—is fundamentally a matter of being suitably related to a truth theory. But the conjecture faces a pair of old objections that are, in my view, fatal when combined.

Foster (1976) noted that given any Tarski-style theory whose theorems instantiate schema (T) or some variant that is appropriately relativized to contexts,

(T) True-in-$\mathbf{L}(\Sigma) \equiv p$

there will be many equally true theories whose theorems pair endlessly many sentences of **L** with very different specifications of whether or not those sentences are true. And if **L** is a Slang whose lexicon includes 'true', then for reasons stressed by Tarski (1944), it's hard to see how any truth theory for **L** could be correct. Moreover, each of these concerns amplifies the other.

Section one offers reminders of these concerns and the relevance of "contingent liar sentences." Section two reviews some aspects of grammatical structure and Chomsky's (1957) notion of *derivation*, which helps sharpen "Foster's Problem," as discussed in section three. In sections four and five, I return to the connection with liar sentences and stress that the deep challenge for truth-theoretic conceptions of meaning has two prongs. Sections six and seven explore some implications for how we shouldn't describe the relations among truth, possibility, grammar, and communication. Pietroski (2017a) discusses some of these points, emphasizing issues regarding logic and linguistic competence. Here, I focus more on connecting Foster-style objections to Davidson with reasons for denying that Slang sentences have truth conditions.

For these purposes, I'll set aside other reasons. But let me mention polysemy. We can use 'window' to describe an opening in a wall, a pane of glass that fills such an opening, a display space behind glass at the front of a store, or a space in a counter at a bank. Cases of polysemy, which invite talk of sub-senses rather than distinct lexical entries, shade off into metaphor. But the clear cases suggest that a word can be used to access a family of concepts that are not extensionally equivalent. Pietroski (2018) argues that Slang expressions are, quite generally, conceptually equivocal in ways that make Davidson's conjecture untenable. The proposed alternative is that meanings are composable instructions for how to access and assemble concepts of a special sort. I won't assume this view, according to which the meaning of 'near a window' determines neither a concept nor an extension.[2] But let's also not assume that Slang expressions map contexts to extensions. Otherwise, some penetrating objections to truth-theoretic conceptions of meaning will seem like mere puzzles about how certain truisms could be true.

1. First Pass

Imagine a box that contains four slips of paper and is otherwise empty. On each slip, a sentence is inscribed. The first three sentences are '1 + 1 = 2', '2 + 2 = 5', and '3 + 3 = 7'. Now suppose that someone utters (1) and thereby claims that three of the sentences in the box are not true.

(1) Three of the sentences in the box are not true.

The utterer's claim is true if the fourth sentence is '4 + 4 = 9' or any other sentence that is not true. The claim is false if the fourth sentence is '4 + 4 = 8' or any other true sentence.

Of course, the fourth sentence is (1). This sentence isn't true, not even relative to the relevant time, box, and sense of 'in'. For if (1) is true, then two of the four sentences in the box are true. In which case, given what (1) means, it isn't true—not even if the meaning of (1) determines a truth condition according to which (1) is true, relative to the imagined context, if and only if three of the sentences in the box are not true. So if (1) is true, it's also not true; hence, (1) isn't true. Thus, three of the sentences in the box are not true. So the utterer's claim is true. Indeed, this claim is true in part *because* the sentence used to make the claim isn't true.

Since the issues here concern meanings, it doesn't matter if we take the claim to be an action of uttering (1) or some associated propositional content.[3] The point of the example is that if (1) has a truth condition, this condition is met. Yet (1) isn't true. So it seems that (1) doesn't have a truth condition, even though in many contexts, (1) can be used to make true claims.

That's not a paradox. It's an argument—borrowing from Parsons (1974), Kripke (1975), and other discussions of "contingent liar sentences"—against the hypothesis that (1) has a truth condition. Similarly, given that (2) is my favorite sentence,

(2) My favorite sentence isn't true.

it isn't true that (2) is true relative to a speaker if and only if that speaker's favorite sentence isn't true. If a theory implies that (2) has this truth condition, the theory falsely predicts that relative to me, (2) is true if and only if it isn't true. My utterances of (2) can be true, because (2) isn't true. But while I can use (2) to make true claims, (2) doesn't have a truth value that can be specified conditionally relative to contexts in which I am speaking. If your favorite sentence is '3 + 3 = 7', then your utterances of (2) are also true; though it wouldn't follow that (2) is true relative to you.

One might reply that good truth theories for Slangs will specify the conditions in which *utterances* of sentences are true (cp. Burge 1974), rather than specifying truth conditions for *sentences* relative to contexts *a la* Kaplan (1989). Given a theory according to which an utterance u of (1) is true if and only if (at the time of u) three of the sentences in the box are not true, one can grant that (1) doesn't have a truth value. Likewise, one can say that an utterance of (2) is true if and only if the utterer's favorite sentence is true. I'll return to a doubt about whether theories that imply instances of a schema like (T-utter) can serve as good theories of meaning for Slangs.

(T-utter) $\forall u$:UtteranceOf(u, Σ) & SentenceOf(Σ, $\mathbf{L}$)[True-in-$\mathbf{L}$(u) $\equiv \Phi(u)$]

But in any case, a slight change of example restores the objection.

Suppose that at a department party, four utterances are chosen for a study. The chosen four are utterances of '1 + 1 = 2', '2 + 2 = 5', '3 + 3 = 7', and (3).

(3) The utterances chosen for the study will include three that fail to be true.

It's not the case that for each utterance u of (3), u is true if and only if the utterances chosen for the study will (i.e., after u) include three that fail to be true. Consider the chosen utterance of (3). If it was true, it failed to be true. So it *wasn't* true, and hence, the utterances chosen for the study *did* include three that fail to be true.[4] Given such examples, one might try to offer theories that deliver instances of (T-utt), with 'TE' adding a restriction to utterances that are truth-evaluable.

(T-utt) $\forall u$:UtteranceOf(u, Σ) & SentenceOf(Σ, $\mathbf{L}$) & TE(u)[True-in-$\mathbf{L}$(u) $\equiv \Phi(u)$]

But as we'll see, such tricks are too clever to yield good theories of meaning for Slangs. And let me emphasize that my claims concern Slangs, not languages invented for scientific purposes.

I'm not offering any solution to the genuine puzzles concerning truth and truth-evaluable things, whatever they turn out to be. But our ignorance/confusion regarding truth isn't a license to say false things about Slangs, or to ignore arguments against truth-theoretic conceptions of meaning. The arguments presented here don't employ tendentious logical principles (e.g., excluded middle for propositions), or inferences known to be paradox-inducing (e.g., Sorites reasoning with vague predicates), in an attempt to prove that Davidson's conjecture is false. The claim is that a bold hypothesis about Slangs leads to contradictions given plausible ancillary assumptions. If we *knew* that sentence meanings map contexts to contents—or at least determine truth conditions for utterances—we might bracket arguments to the contrary as puzzles to be solved someday. But that's not our situation. If we want to *find out* how meanings are related to truth, we can't keep assuming an answer and setting obvious objections aside.

In reply to some of Russell's (1905) claims about 'the', Strawson (1950) suggested that sentence meanings are "directions for use," rather than logically possible conditions of the world. But he also noted that such directions can determine something like conditions of truth for uses of sentences in contexts that satisfy the "presuppositions" for being true or false. Strawson stressed that some sentences can be used to express, or at least voice, truth-evaluable thoughts that have subject-predicate form and subjects that are singular rather than quantificational. Such uses depend on successful acts of reference. So while (4) is as meaningful as (5),

(4) Vulcan affects the orbit of Mercury.

(5) Neptune affects the orbit of Uranus.

(4) cannot be used to voice a subject-predicate thought whose subject denotes a planet, given that our solar system is unfriendly to referential uses of 'Vulcan'. But since Neptune exists, (5) can be used to voice such a thought.[5] Contingent liar sentences can be viewed as illustrations of the broader point that an uncooperative world can undermine truth-evaluability in many, many ways.

Being truth-evaluable may be a property of certain thoughts, along with certain invented sentences that were designed to be truth-evaluable, but not the quirky sentences of Slangs. Grammaticality doesn't even ensure coherence, as Chomsky (1957) illustrated with (6).

(6) Colorless green ideas sleep furiously.

And sentence meanings reflect a kind of *derivational* equivalence that is independent of logical/conceptual equivalence; see section two.[6] So even if we temporarily ignore examples like (1-3), why be temporarily confident that declarative Slang sentences have truth conditions?

Of course, one can still speculate that sentences like (7) have truth conditions.

(7) Three of the tools in the box are not blue.

And if one restricts attention to sentences with no "semantic" words, then critics won't be able to provide liar-esque reductios of the claim that these sentences have truth conditions. But even if we can't prove that (7) doesn't have a truth condition, (7) seems very similar to (1).

(1) Three of the sentences in the box are not true.

So if (1) doesn't have a truth condition, why think (7) does? Examples like (1) suggest that Slang sentences fail to have truth conditions, and that such sentences are meaningful in another way.[7]

This is not a marginal point. It can also be seen as a special case of Foster's (1976) challenge, often illustrated with (8) and (9).

(8) Snow is white.

(9) Grass is green.

For simplicity, let's suppose that in fact, snow is white, and grass is green—ignoring unwhite snow on city streets, the many ungreen desert grasses, and any context sensitivity due to 'is'. Imagine a theory, Θ, whose theorems include (8T) and (9T);

(8T) True('Snow is white.') ≡ **SW**

(9T) True('Grass is green.') ≡ **GG**

where '**SW**' and '**GG**' abbreviate true sentences of the relevant metalanguage, and relativization to the object language has been suppressed for simplicity. However (8T) and (9T) are derived from Θ, there will be many equally true theories whose theorems include (8T*) and (9T*).

(8T*) True('Snow is white.') ≡ **GG**

(9T*) True('Grass is green.') ≡ **SW**

For example, one can add '**SW** ≡ **GG**' as an extra theorem if Θ respects the following principles: if '$\alpha \equiv \beta$' and '$\beta \equiv \gamma$' are theorems, so is '$\alpha \equiv \gamma$'; and if '$\alpha \equiv \beta$' is a theorem, so is '$\beta \equiv \alpha$'. In sections four and five, I'll say more about the connection with liar sentences. For now, just imagine a theory, Θ, whose theorems include an initially plausible instance of (1T).

(1T) True('Three of the sentences in the box are not true.') ≡ **P**

The worry is that many equally true theories, perhaps including Θ, will have (1T*) as a theorem.[8]

(1T*) True('Three of the sentences in the box are not true.') ≡ ~**P**

As shown in section three, effects of supplementary theorems can also be packed into modified axioms that license derivations of biconditionals like (8T*) and (9T*). This makes it hard to see how any theory of truth could be the core of a good theory of meaning for a Slang. How can providing a theory that pairs (8) with '**SW**' and (9) with '**GG**' be a way of specifying what (8) and (9) mean, if an equally correct theory pairs (8) with '**GG**' and (9) with '**SW**'?

That said, it remains tempting to say that while biconditionals like (10) and (11) are true,

(10) 'Snow is white.' is true (in English) if and only if grass is green.

(11) 'La neige est blanche.' is true (in French) if and only if grass is green.

they are "uninterpretive" or "non-translational," in contrast with "interpretive" or "translational" biconditionals like (12) and (13).

(12) 'Snow is white.' is true (in English) if and only if snow is white.

(13) 'La neige est blanche.' is true (in French) if and only if snow is white.

So one might hope that if a suitably *simple* truth theory, Θ, for a Slang, **L**, has theorems that include a translational T-sentence for each declarative sentence of **L**, Θ won't also have theorems that are not translational. But in my view, this hope is misguided.

2. Some Grammatical Background: Derivations and Homophony

For purposes of illustration, let's avoid complicated sentences like (8) and (9), which can be used to make generic claims about uncountable stuff. Instead, let's pretend that 'Ernie snores' and 'Bert yells' have the simplified grammatical structures displayed in (14) and (15).

(14) $[_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]]$

(15) $[_{S}\ [_{NP}\ [_{N}\ \textit{Bert}]][_{VP}\ [_{V}\ \textit{yells}]]]$

Let's also pretend that each proper noun is an atomic expression that is true of at most one thing, and each present tense verb is an atomic expression that can be true of zero or more things, as specified with lexical axioms like (L1-L4);

(L1) TrueOf(*Ernie*, x, **c**) ≡ Identical(x, Ernie)

(L2) TrueOf(*Bert*, x, **c**) ≡ Identical(x, Bert)

(L3) TrueOf(*snores*, x, **c**) ≡ Snores(x)

(L4) TrueOf(*yells*, x, **c**) ≡ Yells(x) where ‘x’ ranges over the entities in some domain that includes the two named individuals, ‘**c**’ ranges over contexts (whatever they are), and the variables are understood as bound by universal quantifiers.[9] This notation allows for axioms like (L5), which encodes a familiar idea about ‘I’.

(L5) $\text{TrueOf}(\textit{I}, \text{x}, \mathbf{c}) \equiv \text{Identical}(\text{x}, \text{SpeakerOf}[\mathbf{c}])$

But let’s suppose that the constituents of (14) and (15) are not semantically context sensitive. To derive theorems like (14T), we also need some schematic combinatorial axioms like (C1-C3).

(14T) $\text{True}([_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]], \mathbf{c}) \equiv$
$\exists\text{x}\{\text{Identical}(\text{x}, \text{Ernie}) \ \&\ \text{Snores}(\text{x})\}$

(C1) $\text{True}([_{S}\ [_{NP} \ldots][_{VP} \ldots]], \mathbf{c}) \equiv$
$\exists\text{x}\{\text{TrueOf}([_{NP} \ldots], \text{x}, \mathbf{c}) \ \&\ \text{TrueOf}([_{VP} \ldots], \text{x}, \mathbf{c})\}$

(C2) $\text{TrueOf}([_{NP}\ [_{N} \ldots]], \text{x}, \mathbf{c}) \equiv \text{TrueOf}([_{N} \ldots], \text{x}, \mathbf{c})$

(C3) $\text{TrueOf}([_{VP}\ [_{V} \ldots]], \text{x}, \mathbf{c}) \equiv \text{TrueOf}([_{V} \ldots], \text{x}, \mathbf{c})$

Though (C2) and (C3) might be viewed as special cases of a more general principle according to which a “phrase” consisting of a single word is semantically equivalent to that word.[10]

Initially, the axioms of a truth theory can seem odd. But their roles become clear given derivations of biconditionals like (14T). One such derivation is shown below.

1. $\text{True}([_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]], \mathbf{c}) \equiv$
$\exists\text{x}\{\text{TrueOf}([_{NP}\ [_{N}\ \textit{Ernie}]], \text{x}, \mathbf{c}) \ \&\ \text{TrueOf}([_{VP}\ [_{V}\ \textit{snores}]], \text{x}, \mathbf{c})\}$ [C1]
2. $\text{TrueOf}([_{NP}\ [_{N}\ \textit{Ernie}]], \text{x}, \mathbf{c}) \equiv \text{TrueOf}([_{N}\ \textit{Ernie}], \text{x}, \mathbf{c})$ [C2]
3. $\text{TrueOf}([_{N}\ \textit{Ernie}], \text{x}, \mathbf{c}) \equiv \text{Identical}(\text{x}, \text{Ernie})$ [L1]
4. $\text{TrueOf}([_{NP}\ [_{N}\ \textit{Ernie}]], \text{x}, \mathbf{c}) \equiv \text{Identical}(\text{x}, \text{Ernie})$ [2, 3]
5. $\text{TrueOf}([_{VP}\ [_{V}\ \textit{snores}]], \text{x}, \mathbf{c}) \equiv \text{TrueOf}([_{V}\ \textit{snores}], \text{x}, \mathbf{c})$ [C3]
6. $\text{TrueOf}([_{V}\ \textit{snores}], \text{x}, \mathbf{c}) \equiv \text{Snores}(\text{x})$ [L3]
7. $\text{TrueOf}([_{VP}\ [_{V}\ \textit{snores}]], \text{x}, \mathbf{c}) \equiv \text{Snores}(\text{x})$ [5, 6]
8. $\text{True}([_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]], \mathbf{c}) \equiv$
$\exists\text{x}\{\text{Identical}(\text{x}, \text{Ernie}) \ \&\ \text{TrueOf}([_{VP}\ [_{V}\ \textit{snores}]], \text{x}, \mathbf{c})\}$ [1, 4]
9. $\text{True}([_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]], \mathbf{c}) \equiv$
$\exists\text{x}\{\text{Identical}(\text{x}, \text{Ernie}) \ \&\ \text{Snores}(\text{x})\}$ [8, 7]

Each line in the derivation is, as indicated, an immediate consequence of an axiom or a pair of previous lines. Of course, drawing on two lines to license a third assumes a rule of inference. But at least for simple examples, simple rules of replacement suffice; see Larson and Segal (1995). The details depend on how the axioms and theorems are formulated; and if the theorems concern utterances, the details can get complicated. But for examples like the one above, the idea is simple and intuitive. Using bold quotes as corner quotes, matched instances of **‘**$\Phi(\text{x}) \equiv \Psi(\text{x})$**’** and **‘**$\Psi(\text{x}) \equiv \Omega(\text{x})$**’** license the corresponding instance of **‘**$\Phi(\text{x}) \equiv \Omega(\text{x})$**’**. So if **‘**$\Phi(\text{x}) \equiv \Psi(\text{x})$**’** and **‘**$\Psi(\text{x}) \equiv \Omega(\text{x})$**’** are derivable, **‘**$\Phi(\text{x})$**’** can be replaced with **‘**$\Omega(\text{x})$**’**.[11]

Let’s not fuss about appeals to such rules of inference/replacement. If they are logically trivial, and intuitively implicit in the combinatorial axioms, they are arguably like the “rewrite” rules employed in the generative procedures that Chomsky (1957) discussed as potential models for Slangs. Such procedures generate strings of words via derivations that begin with a start symbol that must be rewritten in accord with rules that include other symbols that must be rewritten and/or terminal symbols that cannot be rewritten; cp. Post (1943). For example, the rules on the left generate the two-word string *Ernie snores*, as shown on the right.

	1. S
S ➔ NP VP	2. NP VP
NP ➔ N	3. N VP
N ➔ *Ernie*	4. *Ernie* VP
VP ➔ V	5. *Ernie* V
V ➔ *snores*	6. *Ernie snores*

This analogy, which may run deep, is also a reminder that appealing to derivations may require a notion of derivational equivalence. The same rewrite rules can be applied, the same number of times, in different orders. Consider the derivations of *Ernie snores* shown below.

1. S	1. S
2. NP VP	2. NP VP
3. NP V	3. NP V
4. NP *snores*	4. N V
5. N *snores*	5. *Ernie* V
6. *Ernie snores*	6. *Ernie snores*

So as Chomsky stressed, while the rules impose a "syntactic structure" on each derivable string of terminal symbols, this structure corresponds to a class of equivalent derivations—each of which determines the "tree structure" indicated with brackets in [$_{S}$ [$_{NP}$ [$_{N}$ *Ernie*]][$_{VP}$ [$_{V}$ *snores*]]]. The rewrite rules generate the two-word string *Ernie snores*, but they generate this string in a certain way that can be represented with a class of equivalent derivations.

This matters in discussions of meaning, even if one initially thinks of languages as sets of strings (or string-interpretation pairs). For the meaning of a complex Slang expression reflects the syntactic/derivational structure of that expression. Ambiguity makes this vivid. The string of words in (16) can be understood in the two ways indicated with sentences (16a) and (16b).

(16) *it is drinking water*

(16a) It is ingesting some liquid water.

(16b) It is some liquid water that is fit for ingestion.

Neither of these sentences provides a good paraphrase of the other. Each has its own meaning. Likewise, while the two construals of (16) are related in ways that reflect shared constituents, it is intuitively clear that (16) supports two distinct sentence meanings. These meanings apparently correspond to different tree structures, along the lines of the simplified (16') and (16''),

(16') [$_{S}$ [$_{NP}$ [$_{N}$ *it*]][$_{VP}$ [$_{V}$ *is drinking*] [$_{NP}$ [$_{N}$ *water*]]]]

(16'') [$_{S}$ [$_{NP}$ [$_{N}$ *it*]][$_{VP}$ [$_{V}$ *is*] [$_{NP}$ [$_{A}$ *drinking*] [$_{N}$ *water*]]]]

each corresponding to its own equivalence class of derivations.

Chomsky (1957) described such examples as cases of "constructional homophony." Cases of lexical homophony—e.g., the various words pronounced /bɛr/, written as 'bear' or 'bare'—show that a Slang can connect a pronunciation with two or more lexical meanings. But the shared pronunciation of 'a name' and 'an aim' is homophonous, although the nouns are not. The same lexical items can also be combined in different ways, as illustrated with (17);

(17) *the duck is ready to eat*

(17a) The duck is prepared to dine.

(17b) The duck is fit for consumption.

'the duck' can be either the subject or the object of 'eat'. But (18) and (19) are unambiguous.[12]

(18) *the duck is eager to eat*

(18a) The duck is eager to dine.

(18b) #The duck is eager to be consumed.

(19) *the duck is easy to eat*

(19a) #It is easy for the duck to dine.

(19b) It is easy to consume the duck.

Other strings have two but not three meanings. In general, a string of terminals has *n* but not *n*+1 meanings, for some number *n*. So it seems that each syntactic structure—or better, each *way of generating* a string—reflects a single meaning, corresponding to one equivalence class of derivations. (I assume that lexical homophones are distinct words; the pronunciation of 'drew his gun near the bank' is shared by several distinct word strings, each with its own equivalence class of derivations.) If the structure that supports the actual meaning of (18) could also support the unattested meaning, then even if there is only one English way of generating (18), that wouldn't yet explain why (18) is unambiguous. Likewise, (19) admits one structure and one meaning.

Similar remarks apply to endlessly many other examples, including (20).

(20) *a doctor texted a lawyer from Boston*

(20a) A doctor texted a lawyer, and the lawyer was from Boston.

(20b) A doctor texted a lawyer, and the texting was from Boston.

(20c) #A doctor texted a lawyer, and the doctor was from Boston.

The attested readings correspond to two ways of generating *texted a lawyer from Boston* as a verb phrase: one in which *lawyer* combines with *from Boston* to form a noun phrase; and one in which *texted* combines with *a lawyer* to form a verb phrase that combines with *from Boston*. But neither of these structures also supports the unattested reading indicated with (20c). In particular, $[[_{VP}$ *texted a lawyer*$][_{PP}$ *from Boston*$]]$ cannot be understood as applying to an individual, x, if and only if x texted a lawyer and x is from Boston; see Pietroski (2018) for discussion.[13]

As these cases illustrate, there are non-obvious constraints on how Slangs connect meanings with pronunciations. But the important point here is that if each way of generating a generable string corresponds to exactly one meaning for that string, then while a proposed truth theory for a Slang can license distinct though equivalent derivations of a theorem like (14T),

(14T) True($[_S [_{NP} [_N$ *Ernie*$]][_{VP} [_V$ *snores*$]]]$, **c**) ≡
∃x{Identical(x, Ernie) & Snores(x)}

each "syntactic" way of generating a sentence of the object language should correspond to only one equivalence class of derivations of a T-theorem.[14] So if a truth theory is offered as the core component of a theory of meaning for a Slang, it shouldn't have (14T) *and* (14T*) as theorems.

(14T*) True($[_S [_{NP} [_N$ *Ernie*$]][_{VP} [_V$ *snores*$]]]$, **c**) ≡ Yells(Bert)

Such a theory would wrongly predict that (14) is ambiguous,

(14) $[_S [_{NP} [_N$ *Ernie*$]][_{VP} [_V$ *snores*$]]]$

at least if 'Yells(Bert)' is not a good translation of '∃x{Identical(x, Ernie) & Snores(x)}'.

As discussed below, extending ordinary notions of translation to Tarskian sentences is dubious business. But if 'Yells(Bert)' translates '∃x{Identical(x, Ernie) & Snores(x)}', it seems that such translation requires only sameness of truth value. And we don't want theories which imply that *if* Bingley is both eager to please and eager to be pleased, then the meaning of (21) can be specified with a formal analog that can also be used to specify the meaning of (22).

(21) Bingley is eager to please.

(22) Bingley is eager to be pleased.

The two Slang sentences differ in meaning, regardless of Bingley's psychology. Correlatively, if a theory has (14T) as a theorem, this doesn't show that the theory provides a good specification of what (14) means. At a minimum, the theory must also fail to have theorems like (14T).[15]

3. Fostering Skepticism

One might try to evade this concern by noting that adding (23) to a theory whose theorems include (14T) doesn't *guarantee* (14T*) as a theorem.

(23) ∃x{Identical(x, Ernie) & Snores(x)} ≡ Yells(Bert)

(14T) True([$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]], **c**) ≡
∃x{Identical(x, Ernie) & Snores(x)}

(14T*) True([$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]], **c**) ≡ Yells(Bert)

One can adopt *very* weak inference/replacement rules which don't ensure that **'**True([...], **c**) ≡ Q**'** is a theorem if **'**True([...], **c**) ≡ P**'** and **'**P ≡ Q**'** are theorems. But before saying why this won't help Davidsonians, let me stress that Tarski was not envisioning such logically anemic theories.

Tarski was thinking about *supplementing* some theory, Θ, of domain entities—e.g., a theory of arithmetic—with a theory of truth for the language in which Θ is formulated. He imposed no special restrictions on which principles of logic (or arithmetic, or physics) could be employed in deducing consequences from the extended theory. Relatedly, Tarski was not offering a theory of meaning for a Slang whose expressions have meanings about which one can theorize. He was stipulating semantic properties for invented expressions, in order to show how a language-relative notion of truth can be characterized in terms of a precise notion of satisfaction, as opposed to a more nebulous notion of meaning. So it's no problem for Tarski if a supplemented theory has theorems like (24), (24T), (25), *and* (24T*).

(24) 3 = 2 + 1

(24T) True('3 = 2 + 1') ≡ Identical[3, Plus(2, 1)]

(25) Identical[3, Plus(2, 1)] ≡ Identical[112, Plus(67, 45)]

(24T*) True('3 = 2 + 1') ≡ Identical[112, Plus(67, 45)]

On the contrary, one of the goals was to make it possible to evaluate both an initial theory and an associated theory of truth by considering their joint implications. This is surely one thing logic is for: highlighting consequences of combining premises. Like it or not, (24T*) follows from the conjunction of (24T) and (25). Similarly, (14T*) follows from the conjunction of (14T) and (23), regardless of whether or not some proposed theory of a Slang licenses this inference.[16]

Still, this leaves room for the idea that Slang meanings can be characterized in terms of truth and a suitably weak notion of derivation. So let's assume "minimal rewrite" rules that are just strong enough to license derivations of (14T) from axioms that include (L1), (L3), and (C1).

(L1) TrueOf(*Ernie*, x, **c**) ≡ Identical(x, Ernie)

(L3) TrueOf(*snores*, x, **c**) ≡ Snores(x)

(C1) True([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) ≡
∃x{TrueOf([$_{NP}$ …], x, **c**) & TrueOf([$_{VP}$ …], x, **c**)}

Foster's point remains. For (L1*), (L3*), and (C1*) are equally true if 'Yells(Bert)' is true.

(L1*) TrueOf(*Ernie*, x, **c**) ≡ Identical(x, Ernie) & Yells(Bert)

(L3*) TrueOf(*snores*, x, **c**) ≡ Snores(x) & Yells(Bert)

(C1*) True([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) ≡
∃x{TrueOf([$_{NP}$ …], x, **c**) & TrueOf([$_{VP}$ …], x, **c**)} & Yells(Bert)

Replacing (L1) or (L3) or (C1) with its asterisked counterpart, or simply adding one of the alternative axioms, yields a theory that delivers (14T**) as a theorem.

(14T**) True([$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]], **c**) ≡
∃x{Identical(x, Ernie) & Snores(x)} & Yells(Bert)

But presumably, (14T**) is not translational. Similarly, given a truth theory whose theorems include (14T), endlessly many equally true theories will have (14T***) as a theorem.

(14T***) True([$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]], **c**) ≡
∃x{Identical(x, Ernie) & Snores(x)} & $e^{i\pi} + 1 = 0$

One can also define ‘Grewsum’ via (G) and replace (C1) with (C1**).

(G) Grewsum([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) $\equiv_{df}$
TrueOf([$_{NP}$ …], Ernie, **c**) & ∀x[TrueOf([$_{VP}$ …], x, **c**) ≡ Snores(x)]

(C1**) True([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) ≡
{Grewsum([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) & Yells(Bert)} v
{~Grewsum([$_S$ [$_{NP}$ …][$_{VP}$ …]], **c**) &
∃x{TrueOf([$_{NP}$ …], x, **c**) & TrueOf([$_{VP}$ …], x, **c**)}}

For illustration, suppose that (26) is a sentence of the object language.

(26) [$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ [$_V$ *sings*]]]

Given a theory that has (26T) as a theorem,

(26T) True([$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ [$_V$ *sings*]]], **c**) ≡
∃x{Identical(x, Kermit) & Sings(x)}

an equally true theory will have (26T**) as a theorem.

(26T**) True([$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ [$_V$ *sings*]]], **c**) ≡
{Grewsum([$_S$ [$_{NP}$ *Kermit*][$_{VP}$ [$_V$ *sings*]], **c**) & Yells(Bert)} v
{~Grewsum([$_S$ [$_{NP}$ *Kermit*][$_{VP}$ [$_V$ *sings*]], **c**) &
∃x{Identical(x, Kermit) & Sings(x)}}

So if a truth theory with theorems like (26T**) doesn't provide a good specification of what (26) means, one wants to know how a theory whose theorems include (26T) and no *obviously* uninterpretive biconditionals like (26T**) can be a *better* specification of what (26) means.

The asterisked axioms are, to be sure, artificial. But the Davidsonian hypothesis is not that meaning-specifying theorems can be derived from meaning-specifying axioms via rules of inference/replacement that are meaning-preserving. The idea was that words have their semantic properties by virtue of their roles in determining the semantic properties of sentences. Moreover, one can change examples so that the counterparts of (14T**) and (14T***) are simple and lawlike according to whatever independent standards are offered. But the point would remain the same: derivable T-sentences can be as uninterpretive as (10) and (11).

(10) ‘Snow is white.’ is true (in English) if and only if grass is green.

(11) ‘La neige est blanche.’ is true (in French) if and only if grass is green.

At this point in the dialectic, advocates of Davidson's conjecture can't just reassert that (26T) is translational but (26T**) isn't. For the question is whether the right side of *any* derivable T-theorem is a good translation of the Slang sentence described on the left. And here, it matters that derivable T-sentences won't be “disquotational” like (14D) or (14D').

(14D) True(‘Ernie snores’) ≡ Ernie snores

(14D') True([$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]], **c**) ≡ Ernie snores

One can speculate that ordinary speakers tacitly know bridging biconditionals like (27);

(27) ∃x{Identical(x, Ernie) & Snores(x)} ≡ Ernie snores

cp. Larson and Segal (1995). But it's hard to see how such biconditionals could be derived, via minimal rewrite rules, from a truth theory of the desired sort. More importantly, the left side of (27) is an invented sentence that has a Tarskian satisfaction condition, but no further meaning. So why think that (14T) is any more translational than (26T**)?

$$\text{(14T)} \quad \text{True}([_S\, [_{NP}\, [_N\, \textit{Ernie}]][_{VP}\, [_V\, \textit{snores}]]], \mathbf{c}) \equiv \exists x\{\text{Identical}(x, \text{Ernie})\ \&\ \text{Snores}(x)\}$$

If it isn't, then it won't help to derive (14D') from (14T) and (27).

If (27) is an instance of **'**P ≡ Q**'**, the expressions on either side of '≡' are sentences of the Davidsonian metalanguage. So it would be question-begging for skeptics to insist that *pace* Davidson, Slang meanings differ in kind from the semantic properties stipulatively assigned to Tarskian expressions. But one can believe this without insisting on it. And skeptics can ask *what it is* for a biconditional like (14T) or (27) to be more translational than a biconditional like (26**) *if* Slang sentences have Tarskian satisfaction conditions but no further meanings.

I'll return to this question in section four. But note that the response of assigning more semantic properties to expressions of the metalanguage, and hypothesizing that Slang expressions also have these properties, is precisely what Davidson hoped to avoid—in part because he hoped to avoid the need for more sophisticated rules of inference/replacement. One can try to provide axioms and rules that deliver theorems like ($14T^+$);

$$(14T^+) \quad \text{True}^+([_S\, [_{NP}\, [_N\, \textit{Ernie}]][_{VP}\, [_V\, \textit{snores}]]], \mathbf{c}) \longleftrightarrow \underline{\exists x\{\text{Identical}(x, \text{Ernie})\ \&\ \text{Snores}(x)\}}$$

where 'True^+' and underlining indicate some intended supra-Tarskian character of the object language and metalanguage, and '<—>' can be correspondingly more demanding than '≡'. Then even if (28) is true, ($14T^{+}$*) might be neither true nor derivable from ($14T^+$) and (28).

$$(28) \quad \text{Identical}(\text{Ernie}, \text{Batman})\ \&\ \forall x[\text{Snores}(x) \equiv \text{AnnoysRobin}(x)]$$

$$(14T^{+}*) \quad \text{True}^+([_S\, [_{NP}\, [_N\, \textit{Ernie}]][_{VP}\, [_V\, \textit{snores}]]], \mathbf{c}) \longleftrightarrow \underline{\exists x\{\text{Identical}(x, \text{Batman})\ \&\ \text{AnnoysRobin}(x)\}}$$

But unless ($14T^+$) is a misleading notational variant of (14M),

$$\text{(14M)} \quad \text{MeaningOf}([_S\, [_{NP}\, [_N\, \textit{Ernie}]][_{VP}\, [_V\, \textit{snores}]]], \mathbf{c}) = \underline{\exists x\{\text{Identical}(x, \text{Ernie})\ \&\ \text{Snores}(x)\}}$$

which might be used to stipulate an interpretation for the formula on the right side of '=', there will presumably be instances of (29) that are as true as ($14T^+$) but not translational.

$$(29) \quad \text{True}^+([_S\, [_{NP}\, [_N\, \textit{Ernie}]][_{VP}\, [_V\, \textit{snores}]]], \mathbf{c}) \longleftrightarrow \underline{\ldots}$$

One is free to speculate that a theory with theorems like ($14T^+$), and none like ($14T^{+}$*), can "do duty" as a theory of meaning for a Slang. But skeptics will ask what it is for a biconditional like ($14T^+$) to be more translational than other equally true biconditionals that might be derived from other equally true theories. I'll return to this point, and the irrelevance of invoking possible worlds, in section seven. But before moving on, a recap may be useful.

Given any proposed truth theory for a Slang, **L**, there will be endlessly many equally true truth theories for **L** that don't have all and only the same theorems, even holding the derivational rules fixed.[17] If all these theories are equally good as theories of meaning, they are also equally bad; and endlessly many of the equally true truth theories are not remotely plausible as theories of meaning. But it's hard to see how Davidsonians can plausibly maintain that a select few of the equally true truth theories are semantically special, given their conception of what it is for Slang expressions to be meaningful. I take this to be the moral of Foster's discussion.

Tarski used ordinary expressions in special contexts to specify satisfaction conditions for certain invented formulae. Davidson hoped to reverse this procedure by using Tarskian sentences that have satisfaction conditions, and no further semantic properties, to specify what our ordinary expressions mean. But at least initially, this seems like putting a rabbit in a hat and declaring that the hat has vanished. Slang expressions—of English, Polish, etc.—have meanings that we can use for many purposes, including the special purpose of inventing formulae that *don't* have meanings of the same kind. But we shouldn't then conclude that our Slang expressions have meanings of the same kind as the invented expressions.

In rarified contexts, we can say things like (30),

(30) The result of concatenating 'P' with a variable is an open sentence that is satisfied by a sequence of domain entities if and only if the entity that the sequence assigns to the variable is a prime number.

thereby stipulating satisfaction conditions for certain invented sentences (e.g., 'Px') that are designed to have no other semantic properties. We can also stipulate that for each such sentence Σ, Σ is a true sentence of the invented language if and only if Σ is satisfied by every sequence of domain entities. But then we shouldn't be surprised if Σ doesn't specify the meaning of *any* Slang expression, not even one that a proposed truth theory happens to pair with Σ. If a Slang expression is satisfied by and only by certain sequences, those sequences might be specifiable with Σ. But if other Slang expressions are "satisfactionally equivalent" despite having different meanings, then pairing all these expressions with Σ isn't a way of specifying what the Slang expressions mean.

Davidson maintained that nonetheless, Slang expressions have satisfaction conditions and no further meanings (or at least none that are more finely individuated ways of presenting or determining the satisfaction conditions). He thought it helped to also say that Slang expressions have their semantic properties holistically: a theorem of a truth theory, Θ, correctly specifies what a sentence of the object language, **L**, means only if for each sentence Σ of **L**, Θ has a theorem that correctly specifies what Σ means; so no theorem of Θ counts as translational unless Θ delivers a translational theorem for every sentence Σ of **L**. In Quinean terms, the idea was that the units of translation are languages, not individual sentences. But whatever one thinks of this holism, Foster noted that it won't turn a truth theory into a good theory of meaning.[18]

4. Deepening the Skepticism

One can grant that there are important senses in which truth is holistic, yet balk at the idea that for each lexical item of a Slang, its meaning can be correctly specified *only* via some theory that correctly specifies the meaning of every expression of that Slang. Perhaps truth is, at bottom, a property exhibited by (contextualized) sentences of languages that have certain global features—e.g., having unboundedly many expressions that are systematically related in logically relevant ways. This wouldn't imply that the meaning of my word 'avocado' can be correctly specified only via some theory that also correctly specifies the meaning of 'correctly specifies'. Though if 'avocado' vividly illustrated some problem for a certain theory of meaning, one wouldn't be impressed if advocates of the theory responded by setting aside expressions that include 'avocado' or any other word that illustrates the same problem—especially if they replied to other objections by insisting on holism for 'avocado'-free fragments of English. So one might think that Davidsonians are especially *un*entitled to set examples like (1) aside for special treatment.

(1) Three of the sentences in the box are not true.

Prima facie, expressions containing 'true' tell against the idea that Slang expressions have recursively specifiable satisfaction conditions. Slang sentences can be used to make true claims. But it doesn't follow that the sentences used to make the claims are true. On the contrary, in certain contexts, (1) can be used to make a claim that is true in part *because* (1) isn't true. So it seems that Slang meanings are related to truth—and the things that our truth-evaluable claims are about—in ways that are more complicated and/or less direct than one might have thought if one focused on examples like (14) without considering the possibility of Fosterized truth theories.

(14) $[_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]]$

Of course, one need not reflect on 'true' and truth to suspect that however Slang meanings are related to the things we talk about, it's complicated. One has to be a practiced partisan to say with a straight face that the count noun 'democracy' is true of all and only the democracies, and that a corresponding mass noun is true of all and only the samples of democracy. It is often said that words like 'democracy' and 'duplicity' are especially hard cases that we shouldn't focus on in the "early days" of theorizing, as if these *words* differ from 'tiger' and 'triangle' along some relevant dimension. But absent independent reasons for thinking that some word meanings are somehow less theoretically tractable than others, the worry is that we're being asked to just accept whatever specifications Davidsonians are led to adopt (having already been asked to ignore the phenomenon of lexical polysemy).

Even this might be acceptable if it were plausible that a correctly formulated truth theory for a Slang, **L**, could be the core of a correct theory of meaning for **L**. But to make this plausible, absent independent evidence that Slang sentences are indeed things that have truth conditions, one needs plausible responses to both examples like (1) and the fact that truth theories can be Fosterized. For these are two especially vivid ways of sharpening the thought that Slang expressions have meanings that can't be characterized in terms of a truth-theoretic relation—or any mind-independent relation—that the expressions allegedly bear to the things we talk about. Principles concerning truth and derivations figure importantly in both ways of pressing the anti-Davidsonian thought; and while examples like (1) invite responses that appeal to sophisticated notions of truth and valid inference, the possibility of Fosterizing truth theories invites restriction to unsophisticated rules of inference/replacement. So the two objections should be seen as a pincer, not independent concerns, each of which can be bracketed when responding to the other.

Derivations for (1) will be complicated in ways that are not relevant here. So let's suppose that (31) and (32) have the indicated grammatical structure,

(31) [$_{S}$ [$_{NP}$ [$_{N}$ *Kermit*]][$_{VP}$ *isn't* [$_{A}$ *blue*]]]

(32) [$_{S}$ [$_{NP}$ [$_{N}$ *Linus*]][$_{VP}$ *isn't* [$_{A}$ *true*]]]

with the verb phrases consisting of an adjective combined with a tensed negative copula, and that 'Linus' has somehow been introduced as a name whose bearer is sentence (32). But these simplifying assumptions are dispensable. Neither *self*-reference nor copular negation is required for the argument that sentences like (1) and those in trilogue (33) don't have truth conditions.

(33) Ernie: Everything that Bert will say is true.
Bert: Everything that Kermit will say is true.
Kermit: Nothing that Ernie said is true.

More substantively, I assume that replacing 'blue' in (31) with 'true' preserves grammaticality. Sentence (34) is grammatical, even if it sounds odd, given that Kermit isn't truth-evaluable.

(34) [$_{S}$ [$_{NP}$ [$_{N}$ *Kermit*]][$_{VP}$ *isn't* [$_{A}$ *true*]]]

Likewise, (32) is grammatical whether or not it is truth-evaluable.

Whatever we think about Davidsonian holism, if a proposal about 'blue' (and the tensed negative copula) cannot be extended in a plausible way to 'true', that's some reason for thinking that the proposal was wrong—even if it initially seemed plausible for 'blue'. Ditto for 'Kermit' and 'Linus'. So if a proposal about (31) cannot be extended in a plausible way to (32), that's some reason for thinking that the proposal was wrong.

We can extend our toy truth theory from section two with axioms (C4), (L6) and (L7).

(C4) TrueOf([$_{VP}$ *isn't* [$_{A}$ …]], x, **c**) ≡ ~TrueOf([$_{A}$ …], x, **c**)

(L6) TrueOf(*Kermit*, x, **c**) ≡ Identical(x, Kermit)

(L7) TrueOf(*blue*, x, **c**) ≡ Blue(x)

Then (31T) is a theorem of the extended theory.[19]

(31T) True([S [NP [N *Kermit*]][VP *isn't* [A *blue*]]], **c**) ≡
∃x{Identical(x, Kermit) & ~Blue(x)}

Likewise, if we add (L8), then (34T) is a theorem of the extended theory.

(L8) TrueOf(*true*, x, **c**) ≡ True(x)

(34T) True([S [NP [N *Kermit*]][VP *isn't* [A *true*]]], **c**) ≡
∃x{Identical(x, Kermit) & ~True(x)}

This might seem fine if Kermit is a frog, assuming that frogs are not truth-evaluable; cp. note 7. But we should doubt (34T), since (L9) seems as plausible as (L6), yet adding (L9) leads to grief.

(L9) TrueOf(*Linus*, x, **c**) ≡ Identical(x, Linus)

The grief is (32T).

(32T) True([S [NP [N *Linus*]][VP *isn't* [A *true*]]], **c**) ≡
∃x{Identical(x, Linus) & ~True(x)}

Unsurprisingly, sentence (32)—a.k.a. Linus—isn't true. If (32) is true, then it is true if and only if Linus—a.k.a. (32)—isn't true; and nothing, not even (32), is true if and only if it isn't true. So the left side of biconditional (32T) isn't true, but the right side of this material biconditional is true. So (32T) isn't true, and neither is any theory that implies (32T).

Without (L9), the implausibility of Davidson's conjecture couldn't be revealed in this vivid way. But in my view, the falsity was there all along, even with axioms like (C4), (L6), and (L7). Examples like (1) and (2) make the point in ways that rely on special contexts.

(1) Three of the sentences in the box are not true.

(2) My favorite sentence isn't true.

But if a theory leads to contradictions when true ancillary assumptions about a particular context are added, the theory is false, even if the context is rare or constructed as part of an experiment.

5. Potential Replies and Spreading the Skepticism

Given (C4) and (L8), (35) follows, given rules of the sort of we've been imagining.

(C4) TrueOf([VP *isn't* [A …]], x, **c**) ≡ ~TrueOf([A …], x, **c**)

(L8) TrueOf(*true*, x, **c**) ≡ True(x)

(35) TrueOf([VP *isn't* [A *true*]], x, **c**) ≡ ~True(x)

But after reading Kripke (1975), one might think that (35) is tolerable, even if 'x' ranges over a domain that includes Slang sentences like (32) and (34).

(32) [S [NP [N *Linus*]][VP *isn't* [A *true*]]]

(34) [S [NP [N *Kermit*]][VP *isn't* [A *true*]]]

For perhaps there are three truth values that a Slang sentence can have. In which case, (32) can have a truth value (unlike Kermit), yet (like Kermit) fail to be true and fail to be false.

Trivalent models for standard predicate calculi, with no semantic vocabulary, are familiar from Kleene (1950). Kripke proved that given a "strong Kleene" model, **M**, for such a calculus,[20] one can always add a truth predicate and extend **M** in a way that preserves all instances of (36);

(36) True(Σ) ≡ Σ

where Σ is any sentence of the language, even if Σ implies and is implied by **'**~True(Σ)**'**. Indeed, there may be many such ways of adding a truth predicate with a "transparent" interpretation, thereby ensuring that (in referentially transparent contexts) instances of **'**True(Σ)**'** and **'**Σ**'** are inter-substitutable *salva veritate*. Kripke provided a recipe that is, as he showed, sure to deliver at least one such interpretation for 'True': start with any interpretation that respects a certain intuitive condition on extending **M** for sentences that have one of the two classical truth values; then recursively tweak that interpretation in a clever way, so that at each stage, the interpretation for 'True' preserves any classical values assigned to sentences at the previous stage. Kripke's procedure will eventually settle on an interpretation for 'True' that extends **M** in a way that preserves (36) and assigns the nonclassical third truth value to any analogs of Linus.[21]

That's nice, especially if respecting (36) is an essential feature of a genuine *truth* predicate. But like Tarski, Kripke was talking about truth, not the meanings of Slang expressions. (Kripke's invented truth predicate doesn't have the meaning of 'true', or any other Slang meaning.) Positing three truth values is part of a helpful and perhaps correct way of thinking about truth. But this doesn't make (35) any less implausible.

We can introduce a Slang adjective 'klee', on a par with 'true' and 'false', and say that each truth-evaluable thing is true or klee or false; where klee things, if there are any, are the bearers of the non-classical third truth value. But even if (32) is klee, and not false, (32) isn't true. So even if (32) is klee, the right side of (32T) is true, and (32T) is true only if (32) is true.

(32T) True([$_{\text{S}}$ [$_{\text{NP}}$ [$_{\text{N}}$ *Linus*]][$_{\text{VP}}$ *isn't* [$_{\text{A}}$ *true*]]], **c**) ≡
∃x{Identical(x, Linus) & ~True(x)}

So if (32) is klee, (32T) is true only if (32) is both klee and true. But nothing, not even (32), is both klee and true. So saying that (32) is klee doesn't make (32T) an acceptable consequence of a hypothesis about Slangs. This shouldn't be surprising, regardless of what one thinks about trivalent models of truth-in-**L** that preserve (36), since the objection to truth-theoretic semantics for Slangs *wasn't* that sentences like (37) are contradictory and hence false.

(37) The sentence 'Three of the sentences in the box are not true.' is true,
in English, if and only if three of the sentences in the box are not true.

On the contrary, the skeptical objector suspects that (1) doesn't have a truth condition,

(1) Three of the sentences in the box are not true.

The skeptic suspects that (1) is among the many things that are not truth-evaluable, and likewise for more complicated Slang sentences like (38) and (39),

(38) The sentence 'Snow is white.' is true, in English, if and only if snow is white.

(39) The sentence 'Snow is white.' is not truth-evaluable,
but it can be used to say that snow is white.

even if (39) can be used to make a true claim. So it's no reply to say that (1) is klee rather than false, as if the only debate was whether Davidsonians should embrace a bivalent conception of truth. Put another way, Davidsonians can't just point to a model of how a formal analog of (37) could be true. The challenge was that given what (1) actually means, there seems to be a good argument to the conclusion that (1) is not a truth-evaluable thing. And the argument is not that it's impossible to provide formal models of truth-in-**L** that preserve formal analogs of (37).[22]

Given a Slang, **L**, with lexical items like 'true', a more radical and direct application of Kripke's technique would be to propose (i) a truth theory, Θ, for a "smaller" language, $\mathbf{L}^{-}$, that is shorn of any semantic lexical items, and (ii) a way of extending Θ to obtain a coherent truth theory for **L** in some clever way that does not merely add an axiom like (L8).

(L8) TrueOf(*true*, x, **c**) ≡ True(x)

But if 'true' is treated differently than 'blue', the obvious question is whether the resulting theorems for sentences containing 'true' are translational in the relevant sense. And if not, why we should think that the theorems for other sentences of **L** are any more translational?

Lycan (2012) explores a related line of thought. His idea is that what I called $\mathbf{L}^{-}$ is really a base language, **B**, to which the predicate **'**True-in-**B**(…)**'** can be added. The extended language, $\mathbf{B}^{+}$, can be supplemented with **'**True-in-$\mathbf{B}^{+}$(…)**'** to form the language $\mathbf{B}^{++}$; etc. For any sentence Σ of **B**, **'**True-in-**B**(Σ)**'** is not a sentence of **B**; for any sentence Σ of $\mathbf{B}^{+}$, **'**True-in-$\mathbf{B}^{+}$(Σ)**'** is not a sentence of $\mathbf{B}^{++}$; etc. This approach will still face difficulties with trilogues like (33) above. Though as Lycan shows, given a truth theory for **B**, one can specify truth theories for the more inclusive languages; and if one describes Slangs as sets of sentences, then **B** is very different than its superset $\mathbf{B}^{+}$, which is very different than its superset $\mathbf{B}^{++}$, etc. But if Slangs are generative procedures, as Chomsky (1986) argues, then **B** is simply the Slang in question *minus* some words that reveal some problems for truth-theoretic conceptions of meaning. In my view, there is no interesting sense in which a speaker of a Slang with words like 'true' is also a speaker of a Slang that lacks these words, and this undercuts the initial attractions of Lycan's proposal; see Pietroski (2017a, 2018). Moreover, the Foster worry remains. Why think that theorems formulated in terms of notions like 'True-in-$\mathbf{B}^{\cdots}$(Σ)'—where '$\mathbf{B}^{\cdots}$' ranges over an alleged families of Slangs, rooted in an alleged base Slang with no semantic vocabulary—will be translational?

This makes it tempting to restrict the rules of inference/replacement so that they don't apply to sentences that contain 'true'. But there are many ways to create liar-esque sentences given factive expressions like 'knows' ('regrets', 'found out', etc.) and the possibility of using deictic expressions to refer to sentences/claims in contexts. Moreover, there's nothing inherently problematic about sentences that contain 'true'. In endlessly many contexts, (1) can be used to make true claims; likewise for and (2) and (40).

(2) My favorite sentence isn't true.

(40) This sentence isn't true.

We understand these sentences perfectly well, and there is no independent evidence that we understand 'true' in a special way. On the contrary, we seem to understand sentences like (41)

(41) Three of my favorite sentences in the box with the most true sentences
are neither true nor examples of arithmetic falsehoods.

in the same human way that we understand grammatically parallel sentences in which 'true' is replaced with 'blue'. One wants to know how cognitively limited speakers can understand so many expressions. Bracketing words that reveal trouble for a proposal leaves this central explanandum for theories of meaning in place, absent reasons for thinking that the apparent counterexamples are understood in a special way. (Holists should agree emphatically.)

A related option is to revise the axioms so that the theorems are restricted to utterances that are truth-evaluable, along the lines of (T-utt).

(T-utt) $\forall u$:UtteranceOf(u, Σ) & SentenceOf(Σ, **L**) & TE(u)[True-in-**L**(u) ≡ Φ(u)]

Then the theory would be silent regarding any other utterances. But even if theorems like (31T-utt) are derivable from a theory that is otherwise suitably constrained,

(31T-utt) $\forall u$:UtteranceOf(u, [$_{S}$ [$_{NP}$ [$_{N}$ *Kermit*]][$_{VP}$ *isn't* [$_{A}$ *blue*]]]) & TE(u)
[True(u) ≡ ∃x{Identical(x, Kermit) & ~Blue(x)}]

we have to ask whether or not (31T-utt) is suitably translational. My own view is that quantifying over utterances is already conflating meaning with use in way that makes such theorems non-translational. I can use (31) in thought without ever pronouncing it.

(31) [$_{S}$ [$_{NP}$ [$_{N}$ *Kermit*]][$_{VP}$ *isn't* [$_{A}$ *blue*]]]

And prima facie, such uses imply nothing about utterances.[23] One can replace 'utterance' with a technical term that includes internal uses. But even if meaning can in principle be specified in terms of a suitably broad notion of use, invoking technical terms invites the worry that the derivable theorems are no more translational than (14T).

(14T*) True([$_{S}$ [$_{NP}$ [$_{N}$ *Ernie*]][$_{VP}$ [$_{V}$ *snores*]]], **c**) ≡ Yells(Bert)

This worry, already in the house, is greatly amplified if the derivable theorems include a technical notion that excludes any utterances that are not truth-evaluable.

One might think that this kind of argument proves too much, since it might be used to challenge the idea that biconditionals like (42) are translational.

(42) True([S [NP [N *I*]][VP [V *snored*]]], **c**) ≡
$\exists$x{Identical(x, SpeakerOf[**c**]) & $\exists$e[SnoringBy(e, x) & Before(e, TimeOf[**c**])}

But while I wouldn't *start* an argument this way, the challenge is apt. Prima facie, the invented sentence on the right side of (42) isn't a good translation of the Slang sentence mentioned on the left, even if this is the best one can do given Davidsonian constraints; cp. (43).

(43) J'ai ronflé.

If Ernie is speaking French and utters (43) as a way of saying why Bert yelled, the right side of (42) is a poor choice for translating his utterance, unless the audience will treat this formalism *as if* it has the meaning of the English sentence mentioned on the left side of (42).

The issue here is quite general. What translational theorem for a sentence with quantifiers, like (44), is a truth theory supposed to deliver?

(44) Every odd number precedes an even number.

Given a first-order predicate calculus whose sentences include (45),

(45) $\forall$x:Ox[$\exists$x':Ex'(Pxx')]

we can stipulate a truth theory in terms of sequences of domain entities. Let's say that a sentence of the calculus is true if and only if every sequence satisfies the sentence, and for any sequence σ: σ satisfies (45) if and only if each sequence, σ*, that associates the first variable, 'x', with an odd number and is just like σ with regard to other variables is such that σ* satisfies (46);

(46) $\exists$x':Ey(Pxx')

where a sequence, σ*, satisfies (46) if and only if some sequence, σ**, that associates the second variable, 'x'', with an even number and is just like σ* with regard to other variables is such that the entity that σ** associates with 'x' precedes the entity that σ** associates with 'x''. To provide such stipulations, we need suitable Slang words and some technical devices. If we employ subscripts to disambiguate anaphoric uses of 'it' as in (47)—cp. Boolos (1998)—

(47) For any sequence$_{\sigma}$, every sequence$_{\sigma*}$ that associates the first variable with an odd number and is just like it$_{\sigma}$ with regard to other variables is such that some sequence$_{\sigma**}$ that associates the second variable with an even number and is just like it$_{\sigma*}$ with regard to other variables is such that the entity that it$_{\sigma**}$ associates with the first variable precedes the entity that it$_{\sigma**}$ associates with the second variable.

then we can offer an almost-English paraphrase of the intended satisfaction condition for (45). But neither (45) nor (47) seem like good translations of (44) or (48).

(48) Chaque nombre impair précède un nombre pair.

Sentence (47) is, to understate, technical in ways that (44) and (48)—which say nothing about sequences or variables—are not. Formula (45), an invented expression in which '$\forall$x' and '$\exists$y' are syncategorematic sentence-*prefixes*, has a stipulated satisfaction condition. It doesn't have any other semantic property, unless one adds further stipulations—e.g, about an intended sense that determines the already stipulated satisfaction condition.[24] Likewise, 'Ox' has a satisfaction condition, but not the Slang meaning of 'is an odd number'. So (45) doesn't have the meaning of (47). And even if (45) somehow inherited the meaning of (47), that won't help if (47) and (48) have different meanings. One can say that (45), with the satisfaction condition specified by something like (47), is a regimentation of (48). But a regimentation can be a poor translation.

And the issue is not about Davidson's preference for a first-order metalanguage. It won't help to replace (45) and (47) with a second-order formula like (49) and an English construal.

(49) $\exists X \exists Y[\forall x(Xx \equiv Ox) \;\&\forall x(Yx \equiv \exists x'[Pxx' \;\&\; Ex'])\;\&\; Includes(Y, X)$

6. True and Translational, or Cognitively Correct

In my view, 'translation' is polysemous. So my concern is not that there is some sense in which Davidsonian T-theorems are not translational. My worry is that there is no interesting sense in which such theorems are translational. I suspect that if there is a "core" concept of translation, it is a concept of a context-sensitive relation exhibited by utterances of meaningful expressions. Given an utterance *u* of a sentence Σ of a Slang **L**, and some standards for translation, we can ask if an utterance *u*' of a sentence Σ' of a Slang **L'** is a good translation of *u* (or a better translation than some utterance *u*'' of a sentence Σ'' of a Slang **L''**). The meanings of Σ and Σ' will be among the relevant factors. But there will be many other factors, including the length of the sentences and the importance of brevity in the context. A simultaneous translator, trying to implement a real-time conversation between people who do not share a language, faces a different task than someone translating a historical text with a book that can include many explanatory footnotes. (And of course, Davidsonians are not entitled to assume that T-theorems are translational in the sense of preserving meaning or Fregean sense.)

I don't deny that Slang expressions can be translated, as in English-French dictionaries. Given an expression Σ of one Slang, we can ask which expression (or class of expressions) of some other Slang provides the best bet for a relatively good translation for uses of Σ across "typical" contexts. From this perspective, the notion of translating expressions is derivative, but not because languages are the units of translation. On the contrary, the practice of translating Slangs is even farther removed from the practice of translating (or paraphrasing) one utterance with another. The idea of translating Slang sentences with invented Tarskian sentences—or Slangs with Tarskian languages—may be an odd residue of the Quinean idea that Slang sentences can be "regimented" even though they don't really have meanings of their own.[25]

Davidsonians are free to develop technical senses in which T-theorems can be translational; see, e.g., Lepore and Ludwig (2007). We might even agree that *if* Slang sentences have truth conditions, then a correct truth theory for a Slang can serve as a theory of meaning for that language, so long as (all of) its theorems are translational in the relevant technical sense. But it would still need showing that Slang sentences have truth conditions. If truth theories for Slangs had theorems like (13), which bilingual speakers regard as decently translational,

(13) 'La neige est blanche.' is true (in French) if and only if snow is white.

that might be an argument that the object language sentences really do have truth conditions, and that the phenomenon of Slang expressions having meanings is a matter of these expressions having the properties specified by the truth theories. But if the theorems are more like (14T),

(14T) $\text{True}([_{S}\,[_{NP}\,[_{N}\,\textit{Ernie}]][_{VP}\,[_{V}\,\textit{snores}]]], \mathbf{c}) \equiv$
$\exists x\{\text{Identical}(x, \text{Ernie}) \;\&\; \text{Snores}(x)\}$

and they are translational only in some technical sense, then one can't just assume that the theorems—or biconditionals like (13)—are true.

This point holds even if the technical notion is spelled out in terms of the hypothesis that speakers of Slangs unconsciously encode truth theories in a mental language that is used as a natural metalanguage. I have nothing against the idea that humans and other animals enjoy languages of thought; and perhaps a psychologized version of Davidson's conjecture, as in Larson and Segal (1995), offers the best hope of defending a version of the conjecture.[26] But one can't just assume that some version of the conjecture is correct.

Using small capitals to indicate expressions of “mentalese,” perhaps something like (14TM) can be derived from internalized analogs of the axioms from which (14T) was derived.

$$\text{(14TM) } \textsc{True}([_{S}\, [_{NP}\, [_{N}\, \textit{\textsc{Ernie}}]][_{VP}\, [_{V}\, \textit{\textsc{snores}}]]], \mathbf{\textsc{c}}) \equiv \exists \textsc{x}\{\textsc{Identical}(\textsc{x}, \textsc{Ernie}) \ \&\ \textsc{Snores}(\textsc{x})\}$$

This wouldn’t show that the mental sentence $\exists \textsc{x}\{\textsc{Identical}(\textsc{x}, \textsc{Ernie}) \ \&\ \textsc{Snores}(\textsc{x})\}$ translates the Slang sentence $[_{S}\, [_{NP}\, [_{N}\, \textit{Ernie}]][_{VP}\, [_{V}\, \textit{snores}]]]$.[27] But one can abandon the idea that a suitably formulated truth theory for a Slang, **L**, will have translational theorems, in favor of the hypothesis that such a theory can be “cognitively correct” in the following sense: each speaker of **L** has, as part of the linguistic competence she employs to understand expressions **L**, an isomorphic theory that assigns the same truth-theoretic properties to expressions of **L**. On this view, what makes a truth theory a theory of meaning is that one or more speakers have and use it in the right way. And one can hypothesize that speakers of **L** have pretty similar theories, none of which are outrageously Fosterized versions of others.

Of course, if Slang expressions have truth-theoretic properties by virtue of being suitably related to mentally represented truth theories, there is a sense in which Slang expressions do have meanings that go beyond (but determine) their truth-theoretic properties. This un-Davidsonian idea invites un-Davidsonian questions. Do symbols of the relevant mental metalanguage have contents that are individuated more finely than truth values? If so, are these contents sets of (Kripkean or Lewisian) possible worlds, Russellian propositions, Fregean senses, or abstracta of some other kind? In virtue of what do the relevant mental symbols have whatever semantic properties they have? What kinds of variation can be exhibited by the hypothesized mental encodings of truth theories for a particular Slang? Perhaps such questions can be left open. But one large question cannot be bracketed: why think that speakers mentally encode *truth* theories, especially given the implications for Liar Sentences?

If psychologized truth theories for Slangs have theorems like (14TM),

$$\text{(14TM) } \textsc{True}([_{S}\, [_{NP}\, [_{N}\, \textit{\textsc{Ernie}}]][_{VP}\, [_{V}\, \textit{\textsc{snores}}]]], \mathbf{\textsc{c}}) \equiv \exists \textsc{x}\{\textsc{Identical}(\textsc{x}, \textsc{Ernie}) \ \&\ \textsc{Snores}(\textsc{x})\}$$

then perhaps Slang sentences really do have truth conditions, and the phenomenon of Slang expressions having meanings is a matter of these expressions having the properties specified by the psychologized truth theories. Though even given the antecedent of the conditional, the psychologized theories might be wrong; and while one can try to maintain that false theories can be correct theories of meaning—see Eklund (2002), Heck (2004, 2007), Lepore and Ludwig (2007), but cp. Collins (2015)—this undercuts the idea that biconditionals like (13) are true.

(13) ‘La neige est blanche.’ is true (in French) if and only if snow is white.

But if such biconditionals are not true, then theories of meaning need not and should not explain why such biconditionals are not true.[28] In which case, we need to start over and ask whether speakers mentally encode *truth* theories, without assuming that theories of *meaning* for Slangs are (at their core) suitably formulated truth theories.

Once psychologistic hypotheses are (back) on the table, one wants to know why it’s better to say that speakers of English have mental analogs of axioms like (L1), (L3), and (C1),

$$\text{(L1) } \mathrm{TrueOf}(\textit{Ernie}, x, \mathbf{c}) \equiv \mathrm{Identical}(x, \mathrm{Ernie})$$

$$\text{(L3) } \mathrm{TrueOf}(\textit{snores}, x, \mathbf{c}) \equiv \mathrm{Snores}(x)$$

$$\text{(C1) } \mathrm{True}([_{S}\, [_{NP} \ldots][_{VP} \ldots]], \mathbf{c}) \equiv \exists x\{\mathrm{TrueOf}([_{NP} \ldots], x, \mathbf{c}) \ \&\ \mathrm{TrueOf}([_{VP} \ldots], x, \mathbf{c})\}$$

as opposed to saying that meanings are concepts, or classes of extensionally equivalent concepts. If understanding $[_{S}\, [_{NP}\, [_{N}\, \textit{Ernie}]][_{VP}\, [_{V}\, \textit{snores}]]]$ requires suitable mental axioms—and this requires concepts of Ernie, the snorers, and some logical operations—why not say, without also invoking truth, that speakers can understand the Slang sentence because they can use it as a guide for how to build a thought of a certain kind by combining concepts accessed via the lexical items in certain ways specified the structure of the sentence? (Cp. Katz and Fodor [1963].)

Put another way, one wants to know what work is left for appeals to truth in theories of meaning for Slangs, once one assumes (*pace* Lewis [1975]) that speakers understand Slang expressions by pairing them with mental representations. Moreover, given this assumption, it seems especially odd to bracket the phenomenon of polysemy. Prima facie, competent speakers don't represent 'window' (or 'true') as having an extension. Even for a single speaker, the word can be used in a single sentence to access various mental representations that *are not* extensionally equivalent. So at a minimum, any psychologized version of Davidson's conjecture should be compared with overtly psychologistic conceptions of meaning that reject the conjecture and don't ignore polysemy; see, e.g., Pietroski (2005, 2018). But the more important point is more general. One can't just assume that Davidson's conjecture is correct if (i) some psychologized version is sustainable, and (ii) all the unpsychologized versions face Foster's Problem. We need to consider the possibility that the conjecture was just wrong.

7. Worlds Won't Help: Meanings vs. Communicative Contents

In section three, I said that associating sentences with sets of possible worlds (as opposed to truth values) won't help in addressing Foster's Problem. Let me conclude by being explicit about this.

Maybe there are worlds where Ernie snores and Bert doesn't yell, or Ernie doesn't snore and Bert yells. But maybe it's an essential truth of muppets that Ernie snores if and only if Bert yells. I don't know the modal facts. Still, I'm sure that (14) and (15) have different meanings.

(14) $[_{S}\ [_{NP}\ [_{N}\ \textit{Ernie}]][_{VP}\ [_{V}\ \textit{snores}]]$

(15) $[_{S}\ [_{NP}\ [_{N}\ \textit{Bert}]][_{VP}\ [_{V}\ \textit{yells}]]]$

So I doubt that these sentences differ in meaning because there are possible worlds at which exactly one of the sentences is true. If others are confident that there are such worlds, I suspect they are somehow characterizing worlds in terms of meanings, as opposed to characterizing meanings in terms of an independently plausible conception of possible worlds.

Similarly, there may be possible worlds at which snow is white and grass isn't green, or snow isn't white but grass is green. My own judgments, regarding the modal status of generic color ascriptions, are neither clear nor confident. Yet I know that (8) and (9) differ in meaning.

(8) Snow is white.

(9) Grass is green.

While these epistemic points are surely not decisive, they are suggestive. Consider (50) and (51).

(50) All snow, if there is any, is snow.

(51) All grass, if there is any, is grass.

Since these sentences differ in meaning, pairing both with the set of all possible worlds is not a way of pairing these sentences with their meanings. Yet every possible world is such that any snow there is snow, and any grass there is grass. So even if there are worlds at which snow is white and grass isn't green, I don't think we should conclude that (52) is *true* at such worlds.

(52) Snow is white and grass isn't green.

There are reasons for doubting that Slang sentences have truth conditions. In principle, these reasons might be outweighed by a good theory that specifies meaning in terms of truth; but given Foster's Problem, such outweighing seems unlikely. One might have hoped that appealing to possible worlds would help in responding to Foster, and making it plausible that meaning can be specified in terms of truth after all. But even if truth conditions are specified relative to worlds, truth theories will—and should—fail to distinguish endlessly many pairs of sentences that differ in meaning but not with regard to truth. So a long road leads back to where we started. Even if (52) isn't a clear counterexample to the hypothesis that sentence meanings are sets of worlds, this doesn't make it plausible that the meaning of (52) is a set of worlds. And there are many examples like (50-51).

Following Kripke (1980), I don't think there are any possible worlds at which Hesperus isn't Phosphorus, unicorns and centaurs exist, or woodchucks fail to be groundhogs. Nonetheless, (53) and (54) differ in meaning, as do (55) and (56).

(53) A woodchuck flew to Hesperus on a unicorn.
(54) A groundhog flew to Phosphorus on a centaur.
(55) Unicorns despise centaurs.
(56) Centaurs despise unicorns.

But even if one insists that there *are* possible worlds with unicorns and such, the point remains given endlessly many sentences like (57) and (58), which are as grammatical as (59).

(57) Three colorless green ideas slept furiously.
(58) Three colorless red ideas slept peacefully.
(59) Three friendly brown dogs slept quietly.

Presumably, there are no possible worlds at which any colorless yet colored ideas sleep furiously or peacefully. In which case, for every possible world *w*, three colorless green ideas slept furiously at *w* if and only if three colorless red ideas slept peacefully at *w*.

Likewise, (60) and (61) differ in meaning.

(60) There is an odd number that precedes every prime number.
(61) Every prime number is preceded by some odd number.

But every possible world is such that at that world, there is an odd number—viz., 1—that precedes every prime number, each of which is preceded by at least one odd number. And at this point, one can't bite the bullet by saying that (60) and (61) have the same meaning for theoretical purposes. String (62) is ambiguous, with the distinct meanings indicated by (62a) and (62b).

(62) *some odd number precedes every prime number*
(62a) Some odd number is such that it precedes every prime number.
(62b) Every prime number is such that some odd number precedes it.

It may be that (62a) has the same meaning as (60), and (62b) has the same meaning as (61). But the two meanings of (62) are not the same. The "surface scope" meaning, indicated with (62a), is intuitively more demanding than the "inverted scope" meaning indicated with (62b); cp (63).[29]

(63) *some odd duck called every phone number*
(63a) Some odd duck is such that it called every phone number.
(63b) Every phone number is such that some odd duck called it.

If meanings are sets of possible worlds, then a Slang, **L**, pairs a sentential pronunciation π with n (but not $n + 1$) meanings if and only if **L** pairs π with n (but not $n + 1$) sets of possible worlds. But consider the pronunciation of (64) and its many disambiguations,

(64) *a colorless green idea with a name was ready to eat near the bank*

remembering that 'a name' and 'an aim' are homophonous. It just isn't plausible that there is a distinct set of possible worlds for each reading. On the contrary, it seems that the readings correspond to equivalence classes of derivations; see section two. And derivational distinctions can mirror distinctions in expression meanings without corresponding to any difference regarding the sets of possibilities that correspond to the various expressions.

In my view, this is the crucial point illustrated by the examples, which are not offered merely as apparent counterexamples to the hypothesis that meanings are sets of possible worlds. Identifying meanings with sets of possible worlds is, fundamentally, as implausible as identifying meanings with truth values or any other "objective contents" that efface the structural/derivational distinctions that underlie the phenomenon of constrained structural homophony that Chomsky (1957, 1964, 1965) stressed and began to explain.

Given a language-independent notion of possible world, we can characterize a technical notion of propositional content in terms of sets of worlds and hypothesize that certain uses of Slang sentences have such contents; see Stalnaker (1978, 1984). We can also invent languages whose sentences have, by stipulation, semantic properties that map contexts of some kind onto sets of worlds. Such languages can even include a transparent truth predicate. We can also imagine a community of Slang users who adopt, by explicit convention, a language whose sentences map contexts onto the defined objective contents. But it doesn't follow that Slang sentences perform this magic for us. There are reasons for doubting that declarative Slang sentences have truth conditions, relative to contexts, even in the spare Davidsonian sense of having truth values that can be specified conditionally. So there are reasons for doubting that these natural sentences have truth conditions in more robust senses, even if there are reasons for thinking that many acts of assertion are bearers of objective contents that are individuated more finely than truth values but more coarsely than Fregean Gedanken.

As Chomsky also stressed, one can and should ask if distinct expressions—corresponding to distinct equivalences classes of derivations—can have the same meaning. Animating examples include (60) and (62a), (61) and 62b), (65) and (66), (67) and (68), etc.

(65) Romeo loves Juliet
(66) Juliet is loved by Romeo
(67) Bingley is easy to please
(68) It is easy to please Bingley

One suggestion, which Chomsky began to explore, was that a derivation D^+ might extend a derivation D by employing one or more transformations of a generable expression.[30] And one can explore other ways of characterizing "semantic equivalence classes" of the (syntactic) equivalence classes that Chomsky described. But even if distinct expressions can have the same meaning, we shouldn't conclude that meanings are independent of derivational structure, if only because of the endlessly many examples of structural homophony. One can reasonably hypothesize that meanings are individuated more coarsely than labeled phrase markers, or "trees," and that some grammatical distinctions do not mark semantic distinctions. But identifying meanings with (functions from contexts to) sets of possibilities—or ways of updating contexts, or any other objective contents that *efface* derivational structure—goes far too far.[31]

Again, sets of possibilities may provide a useful notion of publicly communicated language-independent contents. Effacing representational distinctions, in order to get at that which various thinkers represent in various ways, is basically what objective contents are for. And if the goal is to describe *what* speakers communicate in episodes of using meaningful Slang expressions to make assertions, then it can be useful to describe speakers as thinkers who represent (*inter alia*) ways the world could be. This can make it tempting to oversimplify and say that the meanings of Slang sentences *are* (functions from contexts to) sets of possibilities. But examples like (50-64) are, along with contingent Liar Sentences, reminders that this is indeed an oversimplification. However meanings are related to objective contents, it's complicated.

There is a tendency to bracket examples like (50-64), along with contingent Liar Sentences, by saying that non-contingent claims and claims about truth are special cases. My suspicion is that these cases are special in being vivid illustrations of problems for truth-theoretic conceptions of meaning. But suppose that on a particular afternoon, Bingley is both eager to please and eager to be pleased, even though he is sometimes neither eager to please nor eager to be pleased. On that afternoon, both (21) and (22) could be used to express contingent truths.

(21) Bingley is eager to please.

(22) Bingley is eager to be pleased.

If the meaning if (21) is the set of possible worlds at which Bingley is eager to please, then since (21) and (22) differ in meaning, the meaning of (22) is some other set of worlds. But the meanings of (21) and (22) are compatible with it being essential to Bingley—and true in every possible world—that he is always eager to please if and only if he is eager to be pleased.

Fosterizing can be modalized. And *pace* Lewis (1986), the logical possibility of Bingley being eager to please without being eager to be pleased doesn't show that there are worlds at which this logical possibility obtains, much less that Slang sentence meanings—as opposed to stipulated contents for invented sentences—can be characterized as sets of the alleged worlds. Necessary truths that are not "truths of logic" don't have to be described as restricted necessities that are false at "distant" Lewisian worlds. And if insisting on such worlds is part of the best reply to Foster-ish objections to truth-theoretic conceptions of Slang meanings, that is telling.

To his credit, Davidson (1967a) didn't try to defend his hypothesis about Slangs by positing controversial possibilia as objective contents. Like Tarski, he didn't even insist on truth values as opposed to sequences of actual domain entities.[32] Davidson saw that the real challenge was to squeeze meaning out of truth without merely redescribing meanings in terms of dubious ontology. But as Foster noted, invoking holism and derivations of T-sentences won't do the trick; and Tarski had already indicated a serious difficulty regarding sentences that include 'true'.

More than fifty years later, much has been learned. Yet the real challenge remains. If the goal is to explain how speakers understand Slang expressions, one wants to know how invoking objective contents—truth values, sets of worlds, or whatevers—will help. One can hypothesize that Slang expressions *have* objective contents of some kind, and that by *specifying* these contents in the right way, one can explain how the expressions are understood. But both aspects of this conjecture still face old, powerful, and related objections.

References

Beall, J., Glanzberg, M., and Ripley, D. (2018). *Formal Theories of Truth*. Oxford: Oxford University Press.

Boolos, G. (1998). *Logic, Logic, and Logic*. Cambridge, MA: Harvard University Press.

Burge, T. (1973): Reference and Proper Names. *Journal of Philosophy* 70:425-39.

—(1974). Demonstrative constructions, reference, and truth. *Journal of Philosophy* 71:205-223.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Mouton.

—(1959). On Certain Formal Properties of Grammars. *Information and Control* 9: 137-167.

—(1964). *Current Issues in Linguistic Theory.* The Hague: Mouton.

—(1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

—(1977). *Essays on Form and Interpretation*. New York: North Holland.

—(1986). *Knowledge of Language*. New York: Praeger.

—(1995). Language and Nature. *Mind* 104:1-61.

Chomsky, N. and Berwick, R. (2017). *Why Only Us: Language and Evolution.* Cambridge, MA: MIT Press.

Collins, J. (2015). Truth and Language, Natural and Formal. In T. Achourioti *et.al*., eds., *Unifying the Philosophy of Truth.* Dordrecht: Springer.
Davidson, D. (1967a). Truth and Meaning. Reprinted in Davidson 1984.
—(1967b). The Logical Form of Action Sentences. Reprinted in Davidson 1980.
—(1973). Radical Interpretation. reprinted in Davidson 1984.
—(1976). Reply to Foster. In Evans and McDowell 1986.
—(1980). *Essays on Actions and Events*. Oxford: Oxford University Press.
—(1984). *Essays on Truth and Interpretati*on. Oxford: Oxford University Press.
—(1986). A Nice Derangement of Epitaphs. In Lepore (1986).
Davies, M. (1987). Tacit knowledge and semantic theory: Can a five per cent difference matter? *Mind* 96: 441–462.
Eklund, M. (2002). Inconsistent Languages. *Philosophy and Phenomenological Research* 64:251-275.
Evans, G. (1981). Semantic theory and tacit knowledge. In S. Holtzman and C. Leich, eds., *Wittgenstein: To Follow a Rule.* London: Routledge and Kegan Paul.
—(1982). *Varieties of Reference*. Oxford: Oxford University Press.
Evans, G. and McDowell, J., eds. (1976). *Truth and Meaning*. Oxford: Oxford University Press.
Foster, J. (1976). In Evans and McDowell (1976).
Glanzberg, M. (2004). A Contextual-Hierarchical Approach to Truth and the Liar Paradox. *Journal of Philosophical Logic* 33: 27-88.
—(2015). Logical Consequence and Natural Language. In C. Caret and O. Hjortland, eds., *Foundations of Logical Consequence*, 71-120. Oxford: Oxford University Press.
Heck, R. (2004). Truth and Disquotation. *Synthese* 142: 317–352.
—R. (2007). Meaning and Truth-conditions. In D. Greimann and G. Siegwart, eds., *Truth and Speech Acts: Studies in the Philosophy of Language*. New York: Routledge.
Higginbotham, J. (1991). Truth and understanding. *Iyyun* 40:271-288.
Kaplan, D. (1989). 'Demonstratives', in J. Almog, J. Perry, and H. Wettstein, eds., *Themes from Kaplan*. New York: Oxford University Press.
Katz and Fodor (1963). The Structure of a Semantic Theory. *Language* 39:170–210
King, J. (2007). *The Nature and Structure of Content*. Oxford: Oxford University Press.
King, J., Soames, S., and Speaks, J. (2014). *New Thinking about Propositions*. Oxford: Oxford University Press.
Kleene, S. (1950). *Introduction to Metamathematics*. Princeton: D. Van Nostrand.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy* 72: 690–716.
—(1980). *Naming and Necessity.* Cambridge, MA: Harvard University Press.
Larson, R. and Segal, G. (1995). *Knowledge of Meaning*. Cambridge, MA: MIT Press.
Lepore, E., ed. (1986). *Truth and Interpretation*. Oxford: Blackwell.
Lepore, E. and Ludwig, K. (2007). *Donald Davidson's Truth-Theoretic Semantics*. Oxford: Oxford University Press.
Lewis, D. (1975). Languages and Language. In K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, Volume 7, Minneapolis: University of Minnesota Press.
—(1986). *On the Plurality of Worlds*. Oxford: Blackwell.
Lycan, W. (2012). A Truth Predicate in the Object Language. In G. Preyer, ed., *Donald Davidson on Truth, Meaning and the Mental*. Oxford: Oxford University Press.
May, R. (1985). *Logical Form: Its Structure and Derivation*. Cambridge, MA: MIT Press.
Parsons, C. (1974). The Liar paradox. *Journal of Philosophical Logic* 3: 381–412.

Peacocke, C. (1986). Explanation in computational psychology: Language, perception and level 1.5. *Mind and Language* 1:101–123.
Pietroski, P. (2005). Meaning Before Truth. In G. Preyer and G. Peters, eds., *Contextualism in Philosophy*. Oxford: Oxford University Press.
—(2017a). I-Languages and T-sentences. In B. Armour-Garb, ed., *Reflections on the Liar*. Oxford: University Press.
—(2017b). Semantic Internalism. In J. McGilvray, ed., *The Cambridge Companion to Chomsky*. Cambridge: Cambridge University Press.
—(2017c). Meanings via Syntactic Structures. In N. Hornstein et. al., eds., *Syntactic Structures after 60 Years.* De Gruyter: Mouton.
—(2018). *Conjoining Meanings: Semantics Without Truth Values*. Oxford: Oxford University Press.
Post, E. (1943). Formal Reductions of the General Combinatorial Decision Problem. *American Journal of Mathematics* 65: 197-215.
Priest, G. (1979). Logic of Paradox. *Journal of Philosophical Logic* 8: 219–241.
—(2006). *In Contradiction: A Study of the Transconsis*tent, second edition. Oxford: Oxford University Press.
Russell, B. (1905). On denoting. *Mind* 14:479–93.
Soames, S. (1984). Linguistics and Psychology. *Linguistics and Philosophy* 7: 155-179.
Stalnaker, R. (1978). Assertion. In P. Cole, ed., *Pragmatics* (vol. 9) New York: New York Academic Press,
—(1984). *Inquiry*. Cambridge, MA: MIT Press.
Strawson, P. (1950). On referring. *Mind* 59:320–344.
Tarski, A. (1944). The Semantic Conception of Truth. *Philosophy and Phenomenological Research* 4: 341-75.

Endnotes

[1] This is the written version of an intentionally polemical talk presented at the Topoi conference in Torino. My thanks to the participants, and especially the organizers, for helpful comments and patience. I am also grateful to the anonymous referees, who provided valuable constructive remarks on the penultimate draft. Chapter four of Pietroski (2018) sets these issues in a somewhat different context, with still inadequate discussion of Lepore and Ludwig (2007). A fuller discussion would need to engage with Glanzberg (2004, 2015), along with other relevant literature on truth and the possibility of psychologized truth theories.
[2] Drawing on many authors, Pietroski (2018) also stresses (i) the sundry ways that truth can depend on context, (ii) some implausible implications of combining empirically justified event analyses, inspired by Davidson (1967b), with truth-theoretic conceptions of meaning, and (iii) the relative ease of recasting, in internalist terms, many proposals framed in truth-theoretic terms.
[3] The word 'claim' is polysemous, along with 'assertion', 'thought', etc. And even if we allow for the logical possibility of contradictory propositions—see Priest (1979, 2006)—this doesn't warrant the radical hypothesis that (1) is true and not true, relative to the imagined context, as opposed to simply not true (because it fails to be truth-evaluable). If a theory of meaning implies that (1) is true given that it isn't true, that tells against the theory. Moreover, if Davidson's conjecture is combined with the idea that a Slang sentence can be true and not true relative to the same context, this heightens Foster's concern. It's hard enough to see how a theory can specify what a sentence Σ *means* by implying that Σ has a certain truth-theoretic property, even without adding the assumption that Σ can have this property if and only if it doesn't.
[4] Other utterances of (3) may have been true. But if a theory predicts that the chosen utterance of (3) was true, that tells against the theory, whatever logic one adopts; see note 3.
[5] Even if the speaker is not "acquainted" with Neptune in Russell's sense; see Evans (1982).
[6] Correlatively, being explicit about notions of derivation—as Tarski and Chomsky were—reveals difficulties for Davidsonians. I read Foster in this light. Chomsky (1957, 1977, 1995) urged a conception of meaning like Strawson's, in an overtly mentalistic idiom, while granting that internalistic meanings can provide "truth indications," see Pietroski (2005, 2017b, 2017c).
[7] I can't prove that my favorite hammer doesn't have a truth condition. But if sentences are tools that can be used to make claims, it isn't surprising that a claim made by using a sentence Σ can be true in part because Σ has no truth condition. Suppose I put a hammer in a box after stating that *if* I ever put a hammer in a box, I will thereby be claiming that once the hammer is in there, three things in the box fail to be true. Then my odd claim, oddly made, is true if the box also contains a wrench and a screwdriver—or a wrench and a false sentence, or a wrench and (1). But the hammer isn't true or false, not even if putting it in the box is a way of making a true claim.
[8] See note 3. An anonymous referee suggested that perhaps relative to some contexts, (1) has a truth condition that cannot be realized. But if we allow for unrealized truth conditions, which would presumably differ from the realized truth conditions of boringly contradictory sentences like 'A dog barked, and no dog barked', the question is whether a theory that specifies (perhaps unrealized) truth conditions for Slang sentences would be a plausible theory of meaning for a Slang. To answer, we would need to know under what conditions distinct sentences are unrealizedly truth-conditionally equivalent.
[9] It's probably better to replace 'Identical(x, Ernie)' with something like 'IsAnErnie(x) & **c**:*Ernie*(x)', in which a predicate that applies to the many Ernies is contextually restricted by (uses of) an index associated with the noun; cp. Burge (1973). But this detail won't matter.

[10] The more interesting cases include $[_{S}\,[_{NP}\,\textit{a}\,[_{N}\,\textit{cat}]][_{VP}\,[_{V}\,\textit{ate}\,[_{NP}\,\textit{a}\,[_{N}\,\textit{fish}]]]]]$ and $[_{S}\,[_{NP}\,\textit{sm}\,[_{N}\,\textit{cats}]][_{VP}\,[_{V}\,\textit{ate}\,[_{NP}\,\textit{sm}\,[_{N}\,\textit{fish}]]]]]$; where '*sm*' indicates the (phonologically reduced) form of 'some' that is like the indefinite article for nouns that are not singular count nouns. The verb phrase $[_{VP}\,[_{V}\,\textit{ate}\,[_{NP}\,\textit{a}\,[_{N}\,\textit{fish}]]]]$ invites axioms like (C4) and (C5).

(C4) $\mathrm{TrueOf}([_{VP}\,[_{V}\,\ldots][_{NP}\,\ldots]], x, \mathbf{c}) \equiv$
$\exists x'\{\mathrm{TrueOf}([_{V}\,\ldots], \langle x, x'\rangle, \mathbf{c})\ \&\ \mathrm{TrueOf}([_{NP}\,\ldots], x', \mathbf{c})\}$

(C5) $\mathrm{TrueOf}([_{NP}\,\textit{a}\,[_{N}\,\ldots]], x', \mathbf{c}) \equiv \mathrm{TrueOf}([_{N}\,\ldots], x', \mathbf{c})$

But plural nouns and "mass nouns" reveal familiar complexities.

[11] One can adopt a special rule for instances of '$\mathrm{True}([_{S}\,\ldots], \mathbf{c}) \equiv P$' or rewrite them, perhaps as instances of '$\forall x[\mathrm{TrueOf}([_{S}\,\ldots], x, \mathbf{c})] \equiv P$', treating sentences as predicates true of everything or nothing. One can also adopt Lewis-Montague style theories with "function application" as the main combinatorial principle and correspondingly categorized lexical axioms that pair expressions with entities or functions that exhibit types from the now familiar Frege-Church hierarchy: basic things of type $\langle e\rangle$, truth values of type $\langle \mathbf{t}\rangle$, and functions of type $\langle X, Y\rangle$—i.e., functions from things of type $\langle X\rangle$ to things of type $\langle Y\rangle$, where $\langle X\rangle$ and $\langle Y\rangle$ are themselves types of the hierarchy. Then axioms like those below

(L1') $\|\textit{Ernie}_{\langle e\rangle}\|^{\mathbf{c}}) = \text{Ernie}$

(L3') $\|\textit{snores}_{\langle e, \mathbf{t}\rangle}\|^{\mathbf{c}}) \equiv \lambda x.\mathbf{Snores}(x)$

(C1') $\|[_{\langle Y\rangle}\,[\ldots_{\langle X\rangle}][\ldots_{\langle X, Y\rangle}]]\|^{\mathbf{c}} = \|[\ldots_{\langle X, Y\rangle}]\|^{\mathbf{c}}(\|[\ldots_{\langle X\rangle}]\|^{\mathbf{c}})$

(C2') $\|[_{\langle X\rangle}\,[_{\langle X\rangle}\,\ldots]]\|^{\mathbf{c}} = \|[_{\langle X\rangle}\,\ldots]\|^{\mathbf{c}} =$

yield parallel derivations like the one below.

1. $\|[_{\langle \mathbf{t}\rangle}\,[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]][_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]]\|^{\mathbf{c}} =$
   $\|[_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]\|^{\mathbf{c}}(\|[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]]\|^{\mathbf{c}})$ [C1']
2. $\|[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]]\|^{\mathbf{c}} = \|[\textit{Ernie}_{\langle e\rangle}]\|^{\mathbf{c}}$ [C2']
3. $\|\textit{Ernie}\|^{\mathbf{c}} = \text{Ernie}$ [L1']
4. $\|[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]]\|^{\mathbf{c}} = \text{Ernie}$ [2, 3]
5. $\|[_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]\|^{\mathbf{c}} = \|[\textit{snores}_{\langle e, \mathbf{t}\rangle}]\|^{\mathbf{c}}$ [C3]
6. $\|[\textit{snores}_{\langle e, \mathbf{t}\rangle}]\|^{\mathbf{c}} = \lambda x.\mathbf{Snores}(x)$ [L3]
7. $\|[_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]\|^{\mathbf{c}} = \lambda x.\mathbf{Snores}(x)$ [5, 6]
8. $\|[_{\langle \mathbf{t}\rangle}\,[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]][_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]]\|^{\mathbf{c}} =$
   $\|[_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]\|^{\mathbf{c}}(\text{Ernie})$ [1, 4]
9. $\|[_{\langle \mathbf{t}\rangle}\,[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]][_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]]\|^{\mathbf{c}} =$
   $\lambda x.\mathbf{Snores}(x)(\text{Ernie})$ [8, 6]

But one still needs a rule that licenses replacement of **'**α**'** with **'**γ**'** given instances of **'**$\alpha = \beta$**'** and **'**$\beta = \gamma$**'** . One can also add a rule of lambda-calculation that licenses an additional step.

10. $\|[_{\langle \mathbf{t}\rangle}\,[_{\langle e\rangle}\,[\textit{Ernie}_{\langle e\rangle}]][_{\langle e, \mathbf{t}\rangle}\,[\textit{snores}_{\langle e, \mathbf{t}\rangle}]]]\|^{\mathbf{c}} = \mathbf{Snores}(\text{Ernie})$ [9, λ]

But this rule of inference, which has nothing to do with the object language, is different in kind.

[12] See Chomsky (1964). Here, '#' indicates that the string cannot be understood as having the indicated meaning. Note that 'eager' and 'easy' can be replaced with 'reluctant' and 'hard'. The implausibility of (18b) is irrelevant.

[13] Theorists can profess to be unconcerned about overdetermination. But if the task is merely to provide an algorithm that pairs each pronunciation with its attested meanings, without concern about whether the algorithm also pairs pronunciations with unattested meanings, then the task is easily accomplished. Specify some class of possible interpretations that include all attested meanings, and then offer the "universal" procedure that pairs each pronunciation with each of the

possible interpretations—or each of the possible interpretations that could be constructed from the relevant lexical items; cp. Chomsky (1957).

[14] A complication: deriving an instance of **'**True([$_S$ ...], **c**) ≡ ∃x{Φx}**'** will involve deriving intermediate theorems in which **'**TrueOf', along with a description of some constituent of [$_S$ ...], appear in **'**Φ**'**. But the intermediate theorems are not specifications of subtly different meanings. So let's grant that such theorems—still partly metalinguistic on the right, and in this sense, not "fully discharged" specifications of truth conditions—don't count for purposes of evaluating how many meanings the theory assigns to an expression of the object language.

[15] See note 13. Davidson (1967a) thought it wouldn't be a problem if a theory of the sort he imagined had theorems like (14T*) in addition to (14T); cp. the discussion of Tarski in section three below. In later work, Davidson proposed additional restrictions on using truth theories as theories of meaning, sometimes in the form of stipulations about the kinds of theories and evidence that "radical interpreters" would consider; see note 18. Along with much of the literature, I don't think these further restrictions help; see Pietroski (2005), and for extended discussion, Lepore & Ludwig (2007). So I focus here on Davidson's idea that given the right notion of derivation, derivable T-theorems will be suitably translational.

[16] Tarski's idea was not that (24T) is more "semantically legitimate" than (24T*). If being true is a matter of being satisfied by all sequences, then (24T) and (24T*) are truth-theoretically alike. Imagine a supplemented theory whose theorems include '(c = b + a) ≡ (0 = $e^{i\pi}$ + 1)' and 'True('c = b + a') ≡ Identical[3, Plus(2, 1)]' and 'True('c = b + a') ≡ Identical[0, Plus($e^{i\pi}$, 1)]'. Someone who saw only the third theorem, after it was derived from the first two, might think that 'b' has a *meaning* akin to that of 'the result of raising the number whose natural logarithm is one to the power of the square root of negative one times the ratio of a circle's circumference to its diameter'. But this would be a mistake. In my view, it's bizarre to segregate theories of truth from theories of the entities that verify sentences that have truth conditions. But in any case, one can't do this and then say—as Davidson (1986, p. 446) does—that "[T]here is no boundary between knowing a language and knowing our way around in the world." If uninterpretive T-sentences can be distinguished from theorems of meaning theories, then *pace* Quine (1951), there is a theoretically interesting analytic/synthetic distinction.

[17] Likewise for any proposed truth$^+$ theory for **L**.

[18] According to Davidson (1973, p. 134, his quote marks)

> an acceptable theory of truth must entail, for every sentence s of the object language, a sentence of the form: s is true if and only if p, where "p" is replaced by any sentence that is true if and only if s is. Given this formulation, the theory is tested by the evidence that T-sentences are simply true; we have given up the idea that we must also tell whether what replaces 'p' translates s.

But even if we can get evidence that T-sentences are "simply true," without ignoring the possibility that Slang sentences don't have truth conditions, the proposed necessary condition doesn't exclude Fosterized theories. So the issues about translation remain.

[19] The derivation below parallels, line by line, the earlier one for [$_S$ [$_{NP}$ [$_N$ *Ernie*]][$_{VP}$ [$_V$ *snores*]]].

1. True([$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ *isn't* [$_A$ *blue*]]], **c**) ≡
   ∃x{TrueOf([$_{NP}$ [$_N$ *Kermit*]], x, **c**) & TrueOf([$_{VP}$ *isn't* [$_A$ *blue*]], x, **c**)} [C1]
2. TrueOf([$_{NP}$ [$_N$ *Kermit*]], x, **c**) ≡ TrueOf([$_N$ *Kermit*], x, **c**) [C2]
3. TrueOf([$_N$ *Kermit*], x, **c**) ≡ Identical(x, Kermit) [L6]
4. TrueOf([$_{NP}$ [$_N$ *Kermit*]], x, **c**) ≡ Identical(x, Kermit) [2, 3]
5. TrueOf([$_{VP}$ *isn't* [$_A$ *blue*]], x, **c**) ≡ ~TrueOf([$_A$ *blue*], x, **c**) [C4]

6. TrueOf([$_A$ *blue*], x, **c**) ≡ Blue(x) [L7]
7. TrueOf([$_{VP}$ *isn't* [$_A$ *blue*]], x, **c**) ≡ ~Blue(x) [5, 6]
8. True([$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ *isn't* [$_A$ *blue*]]], **c**) ≡
   ∃x{Identical(x, Kermit) & TrueOf([$_{VP}$ *isn't* [$_A$ *blue*]], x, **c**)} [1, 4]
9. True([$_S$ [$_{NP}$ [$_N$ *Kermit*]][$_{VP}$ *isn't* [$_A$ *blue*]]], **c**) ≡
   ∃x{Identical(x, Kermit) & ~Blue(x)} [8, 7]

[20] If we represent the three model-values for sentences with the numbers 1, ½, and 0, then given a Strong Kleene model: the values of **'**P & Q**'** and **'**P v Q**'** are, respectively, the minimum and maximum values of the conjuncts or disjuncts; **'**∀xΦx**'** and **'**∃xΦx**'** are treated similarly as conjunctions/disjunctions of their instances; and the value of **'**~P**'** is 1 minus the value of **'**~P**'**; so if 'P**'** the value ½, so does **'**~P**'**. Classical bivalent models can be viewed as special cases, with no sentences having the value ½.

[21] See *Formal Theories of Truth* (2018) for a helpful accessible discussion and guide to the more technical literature.

[22] If there is a truth-evaluable *proposition* that Linus isn't true, maybe this proposition is klee, and the principle of excluded middle (for thoughts) is false. That still leaves difficulties, including Curry's paradox regarding conditionals and "revenge" puzzles concerning interpretations; see note 21. But these difficulties for certain proposals about truth do not justify implausible claims about Slangs.

[23] Indeed, I suspect that pronunciation is evolutionarily more recent than the core system that generates meaningful combinations of lexical items that correspond to families of concepts; see Chomsky and Berwick (2017).

[24] And even if one stipulates that (37) has the meaning of (36), this doesn't guarantee that the Slang meaning determines the Tarskian satisfaction condition, much less that this meaning is compositionally determined by the components and syntax of (37).

[25] For purposes of a class in elementary logic, I can evaluate some regimentations of a Slang sentence as better than others. Though for these purposes, I find myself assuming that words like 'some' and 'all' have meanings that are not mere extensions, while forgetting that Tarskian squiggles like '∃' and '∀' are syncategorematic.

[26] Cp. Evans (1981), Peacocke (1986), Davies (1987), Higginbotham (1991).

[27] Indeed, if Slang expressions have their meanings because of how their (alleged) truth-theoretic properties are mentally specified, then it seems unlikely that the mental specifications also have meanings; though perhaps they have contents that go beyond Tarskian satisfaction conditions.

[28] Larson and Segal (1995) assume that such biconditionals are true, and that psychologized theories of meaning should help explain this.

[29] Chomsky (1957) noted that *everyone in the room speaks two languages* can be understood in two ways: (a) everyone in the room is bilingual; or (b) two languages are such that everyone in the room speaks them. This invites the hypothesis, developed and defended by May (1985), that the string is structurally homophonous. Given a room, there may be worlds in which (a) or (b) obtains, but not both. Though both sentence meanings can be used to make correct reports, even if as a matter of necessity, every in the relevant room speaks the same two languages.

[30] This was one reason for hoping that Slangs could be characterized in terms of a "base" component (analogous to a context-free grammar) and transformations, at least some of which preserve meaning, with recursion restricted to the base component—as opposed to more powerful (context-sensitive) grammars whose rewrite rules can license inversions, say from 'BA'

to ‘AB’, and hence from ‘ABABABCCC’ to ‘AAABBBCCC’. Such rules can mimic the effects of transforming a “Deep Structure.” But as Chomsky (1959) notes,

[31] By contrast, King’s (2007) notion of structured contents is designed to mirror meaningful aspects of grammatical structure. See also King, Soames, and Speaks (2014).

[32] That seems right to me. The study of Slangs can lead one to posit various *linguistic* entities. But if it leads one to posit universes that physicists can do without—or abstracta that only certain semanticists want—something has gone wrong.