Chapter Four: Truth or Understanding

In chapter two, I discussed a model language that has a Tarskian semantics, and how such a language might be used to introduce a Frege-Church language whose expressions denote entities and truth values. But in chapter zero, I also sketched some initial reasons for doubting that there are theories of truth for the spoken/signed languages that children naturally acquire, much less truth theories that can serve as adequate theories of meaning for human I-languages. In chapter three, I focused on one cluster of reasons concerning "event" variables. In this chapter, I discuss a cluster of more general reasons for rejecting thesis (D).

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

1. Overview

Some reasons for rejecting (D) are independent of any particular conception of what human languages *are*. Though given thesis (C), discussed in chapter zero,

(C) Each human language is an I-language in Chomsky's sense.

(D) implies the very bold conjecture (B).

(B) Each human language is a biologically implementable generative procedure **H**
such that some theory of truth for **H**
is the core of a correct theory of meaning for **H**.

In my view, (B) is very implausible, and this tells against (D). While truth-theoretic models of human I-languages can be valuable, the *interpretations* that such models connect with bracketed strings of lexical items—e.g., [see [a [man [with [a telescope]]]]]—seem to have very different properties than the *meanings* that child-acquirable procedures connect with pronunciations in certain human ways. This first section is an outline of the argument developed below.

1.1 Liars

In section two, I discuss "Liar Sentences" like (1);

(1) Lari is not true.

where 'Lari' is, by stipulation, a name for the second numbered sentence in this chapter.[1]

(2) The second numbered sentence in chapter four of
Conjoining Meanings: Semantics Without Truth Values is not true.

I think that (1), (2), and their kin vividly illustrate a general property of human language sentences: they are neither true nor false—and not because they have other truth values, or have truth values relative to contexts. In my view, sentences of a human I-language are like most things in *not* having truth conditions. In this respect, (1-3) are like galaxies, horses, and numbers.

(3) Snow is white.

Such things are neither true nor false because they are not truth-evaluable. Likewise, in my view, neither (4) nor (5) has a truth condition; and hence, neither of these I-language sentences are true.

(4) The sentence 'Lari is not true.' is true if and only if Lari is not true.

(5) The sentence 'Snow is white.' is true if and only if snow is white.

I readily grant that there are truth-evaluable things, including many acts of making judgments or assertions *by using* sentences like (3). And it may be that Liar Sentences cannot be so used. But this doesn't show that sentences like (1) or (2) have defective *meanings*. The corresponding thoughts may well be non-ideal in various respects. So for various purposes, we

---

[1] Compare the introduction of 'Julius' as a name for whoever invented the zip; see Evans (1982). For many purposes, we could use '(2)' instead of 'Lari'. But '(2)' is not a human language name for sentence (2); and '(2) is not true' may not be an expression of any human I-language, as opposed to semi-technical analog.

may want to speak of propositions when talking in idealized ways about the truth-evaluable contents of judgments/assertions. Though a decision to talk this way is not an argument for the hypothesis that sentences of a human I-language determine (function from contexts to) propositions. On the contrary, sentences like (1) and (2) suggest that this hypothesis is false, even if we agree to waive concerns about which proposition is allegedly expressed by (6)

(6) Al chased Theo.

in contexts where the possibility of mutual chase has not been excluded. And whatever we say about judgments/assertions/propositions, Liar Sentences tell against thesis (D1).

(D1) for each human language **H**, there is a correct theory of truth for **H**.

There are potential replies. But as we'll see, these replies make it harder to defend (D).

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

Employing certain technical notions, familiar to logicians, can help in providing truth-theoretic *models* that accommodate *analogs* of (1) and (2). But especially given thesis (C),

(C) Each human language is an I-language in Chomsky's sense.

these technical notions seem out of place in theories of meaning for child-acquirable languages.

It can be tempting to just set Liar Sentences aside, along with talk of skies and sunrises, as special cases. But if human languages are biologically implemented *procedures that generate* expressions—then we shouldn't be surprised if the generable expressions are *not* related to the world in ways required for truth-evaluability. Sentences like (3) may be tools that speakers use, in assembling thoughts from available concepts, rather than end-products that themselves have truth values (even relative to contexts). If (3) is a biologically generable combination of lexical items, which humans can use to assemble a thought of a certain sort, then (3) may not be any more truth-evaluable than (1). It may just be easier to see that (1) is not truth-evaluable.

1.2 Contexts

It is worth remembering that even sincere assertions can fail to be true or false.

A competent speaker might sincerely utter (2)

(2) The second numbered sentence in chapter four of
Conjoining Meanings: Semantics Without Truth Values is not true.

while justifiably but mistakenly thinking that (7)

(7) 2 + 2 = 5

was the second numbered sentence in the relevant chapter. Such a speaker might also utter (8),

(8) Vulcan has an iron core.

presupposing that Vulcan exists in order to say that it is dense enough to affect Mercury's orbit. No such utterance is true. But falsity requires more than grammaticality, sincerity, and nontruth; see Strawson (1950), Evans (1982). Likewise, an utterance of (9) may be neither true nor false

(9) That is a dog.

if the utterer fails to demonstrate anything. More generally, making a false claim requires a significant kind of success, in that various presuppositions have to be met.[2]

In section three, I argue that these familiar points bolster "Liar Arguments" against (D),

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

[2] Cp. Armour-Garb (2001, 2004) on what he calls the "no proposition expressed" strategy. Though it is hard to stop (2) from expressing a proposition if (2) *would have expressed* a proposition *had* (7) been the second numbered sentence, unless the relevant notion of expressing is technical and presupposes an independent sense in which (2) is meaningful. See Armour-Garb and Woodbridge (2013, forthcoming).

in part by highlighting the question of whether a theory of truth/meaning for a human language should have theorems concerning utterances, sentences, or both. This foundational question is interwoven with questions about whether human I-languages generate expressions with elements that track the *many* ways in which the truth or falsity of utterances can depend on context; cp. Stanley (200x). Section three connects these sources of difficulty for (D) with others that are more specific to complementizer phrases, as in (10).

(10) Everyone who thinks that there are unicorns on Hesperus
wants to join Paderewski in hunting them.

As Kripke's (197x) puzzle highlights—and as Davidson's (196x) initial discussion of "samesaying" already suggested—reports like (10) and the simpler (11)

(11) Someone denied that Hesperus is Phosphorus.

do not merely present the challenge of providing a model according to which replacing one name for Venus with another need not preserve truth. Even given a theory according to which 'that'-clauses can be used to denote (or describe, or quantify over) suitably fine-grained truth-evaluable entities, there remains the question of whether words like 'think' and 'deny' have Tarski-style satisfaction conditions. Prima facie, the truth or falsity of an utterance of (11) can be depend on the context in many complicated ways. So even if 'chase' has a satisfaction condition that can be specified, perhaps allowing for some dimensions of context sensitivity that can be plausibly accommodated in a specification of what 'chase' *means*, one might doubt that 'think' and 'deny' ('say', 'doubt', etc.) have any such satisfaction conditions.

It can be tempting set these lexical items aside, along with 'sky' and 'Venice', as special cases. But at some point, advocates of (D) have to get beyond sketching idealized models of various constructions, and provide a model that is *accurate* for at least a few simple cases. We don't expect models of the hydrogen atom to extend easily to all of chemistry, much less biology. But a theory of the atomic table cannot abstract away from the rows, the columns, and each of the elements. Likewise, a truth-theoretic account of meaning for a human I-language cannot idealize away from each lexical item while also bracketing general questions about truth, denotation, predicates, relations, and contexts.

A main theme of this chapter and its predecessor is that nearly fifty years after Davidson's (1967a, 1967b, 1968) seminal work, a cluster of objections to (D) remain as serious as they were then. Some weaker objections in the vicinity have been addressed, and various proposals have been offered, showing that the stronger objections are not decisive. But we can and should ask whether the ancillary assumptions required to defend (D) are plausible. For truth-theoretic semantics is neither in its infancy nor the only game in town when it comes to describing the meanings of human language expressions. As discussed in chapter three, we can reject Davidson's conception of meaning, but welcome his replacement of CHASED(X, X′) with CHASED(X, X′, X″) in logical forms. Given a hypothesized logical form *L* for a human language sentence S, we can view *L* as a model thought that might be constructed from *ideal* atomic concepts by *executing* the assembly instruction that S provides, instead of saying that *L* specifies the truth condition that S itself exhibits. On this view, the meaning of S does not determine a truth condition for S. So we need not abandon all the good work on action reports, Liar Sentences, context sensitivity, and the meanings of complementizer phrases. But we need to locate the real value of this work, as opposed to automatically construing it in terms of (D).

2. Liars, Truth, and Meaning

In §2.1, I discuss the difficulty that Liar Sentences present for (D1).

(D1) for each human language **H**, there is a correct theory of truth for **H**.

I then turn to an independent worry—via Foster (1976)—about (D).

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

As discussed in §2.2, a certain response to "Foster's Problem" has become standard. But this response makes it even harder to see how advocates of (D) can accommodate examples like (1)

(1) Lari is not true.

2.1 Ways of Not Being True

Recall that Lari is none other than sentence (2).

(2) The second numbered sentence in chapter four of
Conjoining Meanings: Semantics Without Truth Values is not true.

The basic point regarding such sentences is familiar. If (2) is true, then since (2) *is* the second numbered sentence in the relevant chapter—viz., this one—then (2) is not true. But if (2) is false, then it is false that (2)—a.k.a. Lari—is not true; yet if (2) is false, then given how falsity related to truth, (2) is not true. So *if* (2) is true, *then* (2) is true, and (2) is not true; and *if* (2) is false, *then* it is false that (2) is not true, and (2) is not true. So (2) is neither true nor false. Likewise, (1) is neither true nor false. Hence, Lari is not true, and (1) is not true.

That conclusion is not paradoxical. Many things are neither true nor false. Consider the galaxy Andromeda, my horse Sadie, the number 5, etc. If (1) and (2) are among the things that are neither true nor false, that is not deeply puzzling. If (1) and (2) *are* among the things that are true nor false, then we face puzzles; but this is one reason, among many, for thinking that (1) and (2) *are not* among the things that are true nor false. There is nothing paradoxical in an especially good argument that Andromeda is not true. Likewise, I claim, there is nothing paradoxical in an especially good argument that Lari is not true. Given the further assumption that (1) is true iff Lari is not true, contradiction follows; but this is one reason, among many, for thinking that the further assumption is false. There are nearby puzzles concerning truth, judgment, and assertion. But the premises of a paradox are supposed to be compelling. And we can consistently reject the hypothesis that sentences of a human language are true or false. Though before drawing morals from (1), we need to be clear that there is more than one way of being neither true nor false.

In noting that Andromeda is neither true nor false, one is not asking which other truth value(s) it exhibits. Many things are just not truth-evaluable. In certain contexts, one might be able to use (12) as an indirect way of saying that Andromeda is not true.

(12) Andromeda is true if and only if there are finitely many prime numbers.

But absent the stipulations of a special invented language, no galaxy has a truth condition. Galaxies, unlike sentences of PL, do not have satisfaction conditions. Nor do galaxies express thoughts and thereby inherit truth conditions. Similar remarks apply to horses, numbers, etc. In my view, sentences of a human language fail to be true or false in this boring sense. They are not among the truth-evaluable things.[3]

We can *invent* a language, ηPL, such that: its sentences are truth-evaluable, and a sentence is false if and only if its negation is true; it allows for symbols of type <e> that do not stand for any d-entity; and sentences of ηPL that contain such symbols denote a third truth value η, distinct from ⊤ or ⊥. In various ways, one can also stipulate that the invented sentence (13)

(13) False(Triskaideka)

[3] Likewise, each sentence/galaxy/horse is neither greater than 5 nor less than or equal to it, and yet—unlike 5*i*—not a special kind of number. Say, if you like, that *some* true claims are contradictory; see Priest (1979, 2006). We still need *very* good reasons for thinking that each one of some contradictory claims is true. And in my view, the reasons for thinking that human language sentences are truth-evaluable do not motivate acceptance of contradictions.

denotes η if 'Triskaideka' denotes (13); cp. Kleene (1950), Kripke (1975). One can then hypothesize that human languages are like ηPL in this respect. On this view, (14)

(14) The fourteenth numbered sentence in chapter four of *Conjoining Meanings: Semantics Without Truth Values* is false.

is relevantly like (13). So if 'Aril' is a name for the fourteenth numbered example in this chapter, then perhaps (15) has the truth condition specified by (15-T), both sides of which are false.

(15) Aril is false.

(15-T) True('Aril is false.') ≡ False(Aril)

In which case, both (15) and Andromeda fail to be true or false, but (15) has the truth value η.

So one can coherently hypothesize that (15) is true if and only if Aril is false. One can also say that a homophonic/homographic analog of (15) would have been true, had the fourteenth numbered example in this chapter been sentence (7), repeated below.

(7) 2 + 2 = 5

And one speculate that (8) has a truth condition, or least a truth value, without being true or false.

(8) Vulcan has an iron core.

But whatever the attractions of this idea, not mine, (1) does not merely fail to be true or false.

(1) Lari is not true.

For sake of argument, let's grant that sentences can have the truth value η. If (1) is like (13) in being truth-evaluable, then presumably, (1) is true if Lari is not true; and Lari is not true, even if there is a difference between being *false* and being *truth-evaluable but not true*. So if (1) is truth-evaluable, then presumably, (1) is true. But (1) isn't true. Perhaps (1) is truth-evaluable in some technical sense without having the truth condition specified by (1-T).

(1-T) True('Lari is not true.') ≡ ~True(Lari)

But it needs showing that any such technical sense has anything to do with linguistic meaning. And note that even if sentences can have the truth value η, (1-T) is not true; its right side is true, but its left side isn't. So (1) seems to be among the things that are not truth-evaluable, and likewise for boundlessly many other Liar Sentences.

I suspect that if (1) is not truth-evaluable, then neither is (15), even if we cannot *prove* that (15) doesn't have the value η. Logicians have reasons for distinguishing "mere" Liar Sentences like (14) from their "strengthened" counterparts that replace 'false' with 'not true'. But it's hard too see how a human I-language could respect this relatively subtle distinction, somehow generating expressions in a way that leads to (15) having a truth condition—and perhaps the truth value η—while (1) ends up not being truth-evaluable. Though likewise, it's hard to see how (16-18) could be truth-evaluable if (1) and (15) are not; see §2.4 below.

(16) Lari is a sentence.

(17) Bessie is not blue.

(18) Bessie is a brown cow.

So perhaps *no* human language sentences are truth-evaluable, and Liar Sentences are just especially vivid examples, as opposed to sentences that call for special semantic treatment.

Of course, it's *possible* that only some declarative sentences have truth conditions. One can speculate that the declaratives of a human I-language have truth conditions unless they are defective in a way that (1) is. But absent a good account of how this possibility is realized, it remains a mere speculation. If sentences like (1) were the only examples of declaratives that seem not to have truth conditions, then it might be reasonable to set them aside as special cases. But as we've seen, doubts arise even with regard to apparently humdrum cases like (6);

(6) Al chased Theo.

and (3), while famous, is hardly a good example for Davidsonians.

(3) Snow is white.

Still, it's widely held that *modulo miscreants*, human language sentences are true or false *relative to contexts*. This leaves multiple degrees of freedom with regard to accommodating Liar Sentences: either the sentences themselves or the contexts in which they are used are somehow exceptional in a way that can be ignored when specifying the meanings of expressions like (3) and (6), which do not include "semantic" vocabulary like 'true'. Arguing against this vague response to Liar Sentences will take a while. Though in the end, I think the basic point is about as simple as it seems to be at first glance: (1) has a perfectly fine meaning, but no truth condition;

(1) Lari is not true.

and there is no relevant difference between (1) and (17), at least not for purposes of theorizing about human I-languages and the meanings they connect with pronunciations.

This is not to deny that context matters with regard to examples like (19).

(19) That is not true.

Logicians can use (19) to illustrate a paradox concerning the assertions that speakers can make in certain contexts. Some theorists will want to say that such contexts are somehow defective, and that (19) has a truth value relative to each nondefective context. I disagree; but this is a conclusion, not a premise. I readily grant that (19) can be used to make claims about many things, including (1-20), and that many of these claims are true or false.

(20) 2 + 2 = 4

And one can talk about relations that a sentence S bears to the pairs ⟨**v**, C⟩ such that **v** is the truth value of the assertion that a speaker would make in the non-defective context C by using S to make an assertion. I don't think (19) *has* a truth value relative any to any context. But my claim is not that (19) fails to have a truth condition because of Liar Paradoxes, and that (1) is similar. My claim—which does not rely on special assumptions about truth, assertions, or contexts—is that (1) does not have a truth condition. If this is correct, then advocates of (D) face the challenge of providing a plausible theory according to which (17) has a truth condition but (1) does not.

(17) Bessie is not blue.

Again, one can *invent* languages whose sentences include analogs of (1); where truth conditions for these sentences are stipulated in terms of technical notions like Kripke's (1975) fixed points, or Gupta and Belnap's (1993) revisions. Such a language is related to the relevant domain in a logically sophisticated way, as specified by the inventor. But for just this reason, the inventions may be unlike human languages. So even if certain natural languages of thought are biological variants of (C)PL—with open sentences like ABOVE(X, X′) and COW(X), as discussed in chapter two—it hardly follows that some such mentalese is related to the world in the way characterized by certain technical notions that logicians designed with the aim of consistently assigning truth conditions to invented analogs of (1).[4]

[4] As discussed in chapter one, it's hard enough to explain how any animals could have a mental variant of (C)PL, even one that doesn't support logically sophisticated truth theories. A concept may not be *true of* anything unless the concept is sufficiently integrated with others in ways that make it possible for the concept-user to recognize *logical relations* among concepts that are sufficiently numerous and diverse. Achieving such integration is no small task. And it's hard to see how words could be true of things if not via concepts. It's puzzling enough that COW(_) applies to all cows, but nothing else. Any corresponding relation that 'cow' bears to the cows would seem to be more complicated. Children apparently use COW(_) to acquire 'cow'; see Bloom (200x). And as Fodor (1986, 2003) stresses, echoing Chomsky (1959), the presence of cows often elicits uses of COW(_) but not 'cow'; though cows

Tarski (1933, 1944) had his own reasons for denying that there are theories of truth for human languages. But even if these reasons can be evaded by quarantining sentences like (1), we shouldn't evade the question of whether sentences like (17) are truth-evaluable. As many philosophers have stressed, it's remarkable that humans can make *any* claims that are true or false. Once a thinker achieves this level of clarity and contact with her environment, truth is at most a mere negation sign away. If only for this reason, it is very optimistic to think that sentences of a human language—expressions that children can generate and comprehend—are themselves truth-evaluable. In the right settings, humans can use these sentences to make and express truth-evaluable *judgments*, given suitable concepts. But truth/falsity may be downstream of linguistic meaning, in that certain *acts of using* meaningful expressions are *candidates* for being true or false, subject to further constraints that are not specifically linguistic.

2.2 Same Truth Value, Different Meanings

In thinking about the implications of Liar Sentences for (D),

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

we need to think about the ways in which (D) goes beyond (D1).

(D1) for each human language **H**, there is a correct theory of truth for **H**.

In particular, if a theory of meaning for a human language is a theory of understanding for that language, one might wonder if a Tarski-style theory of truth that accommodates (11)

(1) Lari is not true.

can also be a theory of understanding for English. But let's go slowly and try to be concessive.

If we abstract away from any context sensitive elements of (21) and (22),

(21) Ernie snores.

(22) Bert yells.

then a Tarski-style theory of truth for English—or better, for any version of English that generates analogs of (21) and (22)—will have theorems like (21-T) and (22-T).

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

(22-T) True('Bert yells.') ≡ Yells(Bert)

In fact, (21) is complicated in at least three ways. First, 'snores' has a generic character that 'snored' does not. (Tense introduces its own complexities, as does 'is a snorer'.) Second, there are many Ernies; see, e.g., Burge (1973). Third, 'Ernie' is being used here to talk about a muppet. And in many contexts, it is clear that muppets are silent inanimate things that cannot snore. Yet it doesn't follow that (21) is false; cp. Travis (1985). Similar points apply to (22) and other instances of the '*NAME VERB*' form. But let's pretend that (21-T) and (22-T) are actually true, with invented sentences like 'Snores(Ernie)' having suitable interpretations by stipulation.

Once we consider more complex sentences, or homophony, it becomes clear that for purposes of deriving T-sentences, mere quotation is inadequate as a device for depicting object language sentences. We need structural descriptions like [see [the [man [with [a telescope]]]]]. But let's view quotation as standing in for suitable structural descriptions of the intended expressions; though it may turn out that by appealing to such descriptions, defenders of (D) incur a commitment to a conception of human language syntax that presupposes (C).

(C) each human language **H** is an I-language in Chomsky's sense

Let's also recognize that (D) does not imply that *every* theory of truth is the core of a correct theory of meaning. Following Davidson (1967b, 1984), one might say that a meaning

---

rarely if ever elicit 'cow' absent uses of COW(_). So *if* the word is true of what the concept applies to, then presumably, this semantic relation is conceptually mediated.

theory requires a *suitably formulated* truth theory that meets certain *empirical constraints*. Indeed, there are at least two clusters of reasons for saying this. I'll soon return to the second. But first, and perhaps most obviously, human languages allow for nondeclarative sentences. A theory of truth/meaning for English needs to be formulated in a way that makes it possible to *supplement* the theory in ways that accommodate (23-28).

(23) Ernie snores? (24) Does Bert yell?
(25) When does Ernie snore? (26) Does Bert yell when Ernie snores?
(27) Please color Kermit green. (28) Don't color Kermit red!

One obvious thought, developed in various ways by many authors, is that each sentence inflects a sentential "radical" with a grammatical mood; where the radical has a truth condition, perhaps relative to assignments of values to variables, and the mood defeasibly indicates the kind of speech act being performed. Abstracting from context sensitivity, the idea is that (21) and (23) share a *part* that is true if and only if Ernie snores, while the difference in pronunciation reflects a difference in mood that is meaningful, at least with regard to how acts of uttering these sentences are construed. Some theories (e.g., Hamblin 1958, 1973) go farther, treating interrogatives and declaratives as expressions that denote things of different sorts—e.g., questions and propositions.[5] It may be that an adequate truth-theoretic conception of linguistic meaning will need to take some such form, in order to describe the systematic relations exhibited by matrix questions like (25), relative clauses as in (26), and embedded questions as in (30).

(30) Does Bert know when/where/why/if/whether Ernie snores?

But let's not fuss about this. Given a theory of truth that can serve as a decent theory of meaning for English declaratives, it would be churlish to complain if the theory didn't explain everything we would like to explain about the meanings of nondeclarative sentences.

However, a correct theory of meaning for a "declarative fragment" of a human language **H** has to be formulated in a way that permits a correct *extension* of the theory to the unboundedly many interrogative and imperative sentences of **H**. So while (D) does not imply that every truth theory for **H** can be supplemented in ways that accommodate the nondeclarative sentences of **H**, (D) does imply that some truth theory for **H** is so supplementable. It's not enough if some theory of truth *might* be supplemented, in the sense that there is no barrier *in principle* to using a truth theory as the core of a theory of meaning.[6] Empirically discovered features of human languages may preclude extending a truth theory for declaratives to nondeclaratives. I won't press such objections here, since the grammatical details are complicated, and any proposal is sure to be controversial; see Lohndal & Pietroski (2011). But objections to (D) need not establish that either (D1) is false, or a truth theory *cannot* be a meaning theory.

(D1) for each human language **H**, there is a correct theory of truth for **H**.

This matters, because objections to (D) take various forms. Some objections, regarding specific constructions, target (D1) without challenging the idea that such a theory (if there is one) could serve as a theory of meaning. Some objections to this latter idea, concerning how meaning is related to truth, are compatible with there being theories of truth for human languages. But objections to (D) can also take a more dilemmatic form: given certain features of **H**, no plausible truth theory for **H** can be formulated in a way that makes it plausible meaning theory. In

---

[5] See Dummett (1976) on the need for a theory of "content" that meshes with a theory of "force," whether or not 'True(_)' is the central predicate in a theory of meaning for a Human Language; cp. Wright (197x). See also Segal (1991), Ludwig & Lepore (200x), Lohndal & Pietroski (2011), and further references there.

[6] Even a kid from Kenosha can become the president. But it doesn't follow that some kid from Kenosha will be elected if suitably formulated and supplemented with campaign funding.

particular, a truth theory that accommodates Liar Sentences might be too sophisticated to be a meaning theory for a language that a child can naturally acquire. Though to develop this thought, we need to consider a general concern about using truth theories as meaning theories. And this brings us to the second reason for thinking that only *suitably formulated* truth theories that meet certain *empirical constraints* can serve as theories of meaning for human languages.

Sentences (21) and (22) are not synonymous. But they may be alike with regard to truth.

(21) Ernie snores.

(22) Bert yells.

Suppose that Ernie snores and Bert yells, and hence that the invented sentence (31) is true,

(31) Snores(Ernie) & Yells(Bert)

along with the T-sentences (21-T) and (22-T). Then (32-34) are also true.

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

(22-T) True('Bert yells.') ≡ Yells(Bert)

(32) True('Ernie snores.') ≡ Yells(Bert)

(33) True('Bert yells.') ≡ Snores(Ernie)

(34) True('Ernie snores.') ≡ True('Bert yells.')

Similarly, since (35a) and (35b) both fail to be true, (35c) is true.

(35a) There are finitely many primes.

(35b) Five precedes two.

(35c) True('There are finitely many primes.') ≡ True('Five precedes two.')

Indeed, endlessly many instances of 'True(*S*) ≡ p' are true but "uninterpretive" in the following sense: the metalanguage sentence used on the right does not specify the meaning of the object language sentence mentioned on the left. So one wants to know how a theory of truth could ever be a theory of meaning for a language whose sentences have meanings that are individuated more finely than truth values. There may be possible worlds at which (32-34) are false, holding the meanings of (21) and (22) fixed. But given endlessly many examples like (35), sentence meanings seem to be more finely individuated than sets of possible worlds.[7]

In response, an obvious initial thought is that one can construct a theory whose *theorems* include (21-T) and (22-T), but not (32-34). Suppose there is a truth theory Θ for English such that *no* uninterpretive T-sentences can be *derived* from Θ. Then Θ might serve as the core of a meaning theory for English, since Θ would—via its theorems—pair each sentence S of English with (and only with) a sentence of the metalanguage that specifies what S means. But this line of thought raises at least three clusters of questions that advocates of (D) must address.

First, are T-sentences like (21-T) meaning-specifying in the relevant sense, or do "interpretive" T-sentences need to be more *disquotational* and hence more like (36)?

(36) 'Ernie snores.' is true if and only if Ernie snores.

If the latter, then it seems that no interpretive T-sentences will be *derivable*, given the need for at least some technicality in the metalanguage. This point becomes especially vivid once we consider context sensitivity and quantification, as illustrated with (37); see section three.

(37) I chased something.

But let's grant that T-sentences like (21-T) are interpretive. More generally, let's not fuss if the right sides of derivable T-sentences exhibit logical *structure*, so long as they don't employ notions that are "foreign" to the sentences mentioned (or structurally described) on the left, as in

[7] Given any independently plausible notion of possible worlds; see Kripke (1982), Lewis (1986).

(32-34); though the interpretive/uninterpretive (or domestic/foreign) distinction may have to be drawn in terms of the cognitive resources that humans employ when acquiring using languages.

Second, there is the question of whether any theory has enough deductive power to generate interpretive T-theorems, yet not enough to generate uninterpretive T-theorems. The issue here is not merely the technical one of writing appropriate axioms and inference rules. Tarski envisioned a theory of truth as an *addition* to a nonsemantic theory (e.g., of arithmetic) that is formulated in an object language governed by a logic that licences deduction of theorems concerning the *domain* (not just the object language). One can instead view a hypothesized truth theory for English as a theory in its own right, *but not* part of a larger theory-of-the-world that is formulated in the same metalanguage and subject to the same logic. I think it's odd to view truth theories as segregated in this way from theories of the things that words are allegedly true of.[8] But let's grant that a theory of truth for English need not be interwoven with assumptions that would make it possible to derive (38) from (21-T).

(38) True('Ernie snores.') ≡ Snores(Ernie) & Precedes(Two, Five)

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

This still leaves the question of whether any theory Θ that generates theorems like (21-T) can fail to generate uninterpretive theorems like (39a) or (39b); where **Γ** is itself a theorem of Θ.

(39a) True('Ernie snores.') ≡ Snores(Ernie) & **Γ**

(39b) True('Ernie snores.') ≡ **Γ** ⊃ Snores(Ernie)

The scope of this point depends on the relevant background logic. But consider a Fregean (second-order) logic that licenses instances of (40); where **A** is a theorem of arithmetic,

(40) HP ⊃ **A**

and 'HP' stands for (the conjunction of suitable Fregean definitions and) the Humean Principle that some things correspond one-to-one with some things if and only if the former have the same cardinality as the latter; see Heck (2011) for extended discussion. If instances of (40) can be instances of **Γ** in (39a) or (39b), then generating uninterpretive T-sentences is easy. One can, of course, try to formulate theories of truth against the background of a weaker logic. But a truth theory that has (21-T) and (22-T) as theorems will also have (4) as a theorem

(17-T) True('Bert yells.') ≡ Yells(Bert)

(41) True('Ernie snores.') ≡ Snores(Ernie) & [True('Bert yells.') ≡ Yells(Bert)]

if the theory permits replacement of a sentence 'p' with the conjunction of 'p' and a theorem. So the background logic needs to be *very* weak, perhaps to the point of being a logic in name only.[9] I return to this point in section three. But the requisite isolation of a truth theory is quite severe. Elementary principles of logic can yield a theory that generates uninterpretive T-sentences.

We are, recall, reviewing some of the questions raised by the following idea: if there is a (suitably supplementable) truth theory Θ for English such that *no* uninterpretive T-sentences can be *derived* from Θ, then Θ is the core of a theory of meaning for English. The first two clusters

---

[8] It's worse than odd if one also says that 'there is no boundary between knowing a language and knowing our way around in the world' (Davidson 1986, p. 446). It seems that (D) requires a significant boundary between general world knowledge and speaker knowledge concerning the implications of truth theories, *pace* Quine (195x). But for now, let's not worry about whether advocates of (D) are entitled to isolate truth theories for human languages from theories of numbers, spacetime, matter/energy, life, animal psychology, human language syntax, etc.

[9] If derivational principles apply only to premises of one sort, corresponding to one aspect of cognition—without supporting extraction of consequences from premises that reflect diverse forms of thought—why view the principles as a *logic*, rather than mere pattern matching within a single (perhaps computational) cognitive module?

of questions concerned the plausibility of this conditional's antecedent. A third cluster grants the antecedent, for the sake of argument, and challenges the conditional itself.

2.3 Fostering Liars

Foster (1974) observed that given a truth theory which generates T-theorems like (21-T),

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

one can easily construct *extensionally equivalent* theories whose theorems include corresponding instances of (42), where 'T' can be replaced with any true sentence of the metalanguage.

(42) True('Ernie snores.') ≡ Snores(Ernie) & T

For illustration, suppose that (21-T) follows from a theory that has (43) and (44) as axioms,

(43) ∀x[TrueOf('Ernie', x) ≡ (x = Ernie)]

(44) ∀x[TrueOf('snores', x) ≡ Snores(x)]

along with some schema like (45), whose instances include (45a).

(45) True(*'NAME VERB.'*) ≡ ∃x[TrueOf(*'NAME'*, x) & TrueOf(*'VERB'*, x)]

(45a) True('Ernie snores.') ≡ ∃x[TrueOf('Ernie', x) & TrueOf('snores', x)]

Replacing (44) with (46), or (45) with (47), yields a theory from which (48) follows.[10]

(46) ∀x[TrueOf('snores', x) ≡ Snores(x) & Yells(Bert)]

(47) True(*'NAME VERB.'*) ≡
∃x[TrueOf(*'NAME'*, x) & TrueOf(*'VERB'*, x)] & Yells(Bert)

(48) True('Ernie snores.') ≡ Snores(Ernie) & Yells(Bert)

So even if *some* truth theory for English has (21-T) as a theorem, this doesn't yet explain why (21) means what it does. An alternative theory has (48) as a theorem.

(21) Ernie snores.

At least one of these truth theories is not the core of a correct theory of meaning for English, since (21) and (49) differ in meaning.

(49) Ernie snores and Bert yells.

To be sure, (46-47) are not interpretive *axioms*, while (44-45) may be. But it's no defense of (D) to say that a truth theory can be a meaning theory so long its axioms are interpretive.

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

If a theory of truth is to *explain* why human language expressions mean what they do, then we need some account of what makes (44-45) *better* than (46-47), as *theoretical descriptions of* the semantic properties allegedly exhibited by the expressions in question.

Davidson (1984) evidently agreed, since he suggested that uninterpretive axioms fail to meet certain *empirical* constraints, which he tried to specify by talking about what an idealized interpreter would ascribe to speakers on the basis of certain evidence; see section four. Many authors have rightly objected to the verificationistic (and Euthyphronic) conception of meaning and language acquisition that Davidson encoded by appealing to what a "Radical Interpreter" would say.[11] So let's pass over this aspect of Davidson's proposal. In my view, unabashed cognitivism is a far better option. One can hypothesize that speakers of **H** have a language of

---

[10] For simplicity, I assume we can derive 'Snores(Ernie)' from '∃x[(x = Ernie) & Snores(x)]' and 'Snores(Ernie) & Yells(Burt)' from '∃x[x = Ernie & Snores(x) & Yells(Burt)]'. But this is not essential to the general point.

[11] See Fodor and Lepore (1992), Heck (2007), Pietroski (2005b), Ludwig and Lepore (200x). And of course, it doesn't help to replace the Radical Interpreter with a semanticist who inscribes a theory Θ—especially if the metalanguage is as powerful as Church's lambda calculus—absent a justification for adopting Θ as opposed to extensionally equivalent alternatives.

thought in which they encode truth-theoretic axioms, from which consequences can be extracted in a constrained way. The idea is that certain axioms like (44-45) are better than (46-47) because competent speakers *encode and deploy* the former but not the latter.

As noted in chapter two, Larson and Segal's (1995) system for deriving T-theorems is an admirably explicit paradigm. It permits instantiation of axiomatic schemata that specify hypothesized contributions of phrasal syntax to phrasal meaning. But the rules for deriving further consequences, given axioms that specify the hypothesized contributions of lexical items, only permit replacements of established equivalents: *if* $\vdash P \equiv Q$ *and* $\vdash Q \equiv R$, *then* $\vdash P \equiv R$; *if* $\vdash \alpha = \beta$, *then* $\vdash \Phi\alpha \equiv \Phi\beta$. So the inference from (21-T) to (42) is not licenced, not even if 'T' is itself a theorem.

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

(42) True('Ernie snores.') ≡ Snores(Ernie) & T

And while theorists inscribe theorems like (21-T), the idea is that speakers can derive mental analogs like (16-M) from more basic representations; where small capitals still indicate concepts.

(21-M) TRUE('Ernie snores.') ≡ ∃X[(X = ERNIE) & SNORES(X)]

(43-M) ∀X[TRUE-OF('Ernie', X) ≡ (X = ERNIE)]

(44-M) ∀X[TRUE-OF('snores', X) ≡ SNORES(X)]

(45-M) TRUE(*'NAME VERB.'*) ≡ ∃X[TRUE-OF(*'NAME'*, X) & TRUE-OF(*'VERB'*, X)]

Heim and Kratzer's (1998) system can be spelled out in a similar way, permitting derivations of mental sentences like TRUE('Ernie snores.') ≡ SNORES(ERNIE) from Church-style axioms like the following: ‖'snores'‖ = λX.T ≡ SNORES(X)]; TRUE('...') ≡ ‖'...'‖ = T.

Of course, appeal to a language of thought abandons certain ambitions for theories of meaning; cp. Dummett (1975). And by focusing on how speakers actually represent the alleged truth conditions of English sentences, one goes beyond the task of specifying a theory such that knowledge of it would suffice for knowing that these sentences have those truth conditions; cp. Davidson (1967b, 1976), Foster (1976). But these are not yet objections to a cognitivist gloss of (D) and how human semantic competence can be applied to specific expressions.[12] Some ambitions are unattainable, and some tasks are unduly modest. Advocates of (D) are free to offer other conceptions of the constraints that are allegedly satisfied by *at least one* truth theory for each human language, yet not by extensionally equivalent uninterpretive theories. But let's grant, at least for now, that some constrained system of derivation provides an answer to Foster's Challenge.

Still, an obvious worry remains. If derivability is logically blind, apart from replacement of established equivalents, then (1-T) and its mental analog will be as derivable as (13-T).

(1-T) True('Lari is not true.') ≡ ~True(Lari)

(17-T) True('Bessie is not blue.') ≡ ~Blue(Bessie)

---

[12] And given such a gloss, theorists can say that a hypothesized truth theory Θ is *more interpretive* than alternatives if Θ *better* reflects the ways that speakers of **H** represent the hypothesized semantic properties of **H**-expressions. As Heck (200x) notes, a cognitivist gloss also makes room for the idea that 'true' expresses "the very concept of *truth* that plays a central role in the semantic theory we tacitly know (p. 343)." Though as he shows, even if the mental language is the relevant metalanguage, very modest assumptions about the concept TRUE(_) still lead to paradox. Heck himself offers a different cognitivist gloss on (D), drawing on Higginbotham (198x) to suggest that one overt truth theory for **H** is more interpretive than another if the former better reflects how speakers of **H** use their semantic competence in communication. I have doubts about emphasizing communication; see section four. But this debate is intramural. And while we may not know how speakers represent truth conditions, evidence of various kinds can be relevant; see Evans (1982), Peacocke (1986) and Davies (1981, 1987), Pietroski et. al. (2009), Lidz et.al. (2011).

Note that if Lari had turned out to be a horse, as opposed to a Liar Sentence, one might have held that both sides of (1-T) are true. Correlatively, one might have thought that (1-T) is no more problematic than (17-T). And in my view, there is something deeply right about this last thought: (1-T) is not less true than (17-T). The difference is that we can approximate a proof that (1-T) is false, given that Lari is a Liar Sentence. And to repeat an earlier point, it isn't paradoxical that (1-T) is false.

The right side of (1-T) is true, since Lari is not true. And since (1) is not true,

(1) Lari is not true.

the left side of (1-T) is false. But it's OK for (1-T) to be false. This *invented* sentence lets us express a certain *theoretical* claim about the English sentence (1). And theoretical claims are often false, with no special exemptions for claims about human languages. We can invent languages in which the analogs of 'true' and 'blue' are of importantly different types, with the analog of (1) being ill-formed. But (1) is grammatical, and I see no reason for thinking that its grammatical structure is importantly different from that of (17).

(17) Bessie is not blue.

The word 'true' may well be polysemous, and perhaps homophonous. But this hardly shows that 'true' cannot have its ordinary semantic sense, whatever that is, in (1).

Perhaps (50) is true, and (1) is not TruthEvaluable in the relevant sense. That's also OK.

(50) TruthEvaluable('Lari is not true.') ⊃
True('Lari is not true.') ≡ ~True(Linus)

One can say that (1) fails to be true in the way that horses and galaxies fail to be true—viz., by not being the sort of the thing that has a truth condition. If (1) doesn't have a truth condition, then we need not and should not say that (1) is true if and only if Lari isn't true. We can say that (51) is true, since both sides of the biconditional are false.

(51) True('Lari is not true.') ≡ Precedes(Five, Two)

Just don't replace the right side with a truth, as in (1-T), and try to maintain the biconditional.

Foster's Challenge suggests that a meaning theory has to be governed by a naïve background "logic" that only permits derivation of theorems like (17-T) from simple schemata and substitutions. Liar Sentences suggest that a truth theory has to be sophisticated enough to not generate (1-T). Defense of (D) requires a plausible response to this apparent tension.

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

This doesn't show that (17-T) is false. But if a plausibly interpretive/cognized theory of truth has (17-T) as a theorem, then it is hard to see how the theory could fail to have (1-T) as a theorem. That is an indirect argument that no plausible theory of meaning for English will have (17-T) as a theorem. Similar remarks apply to (52-53).

(52) Lari is true. (52-T) True('Lari is true.') ≡ True(Lari)
(53) Bessie is blue. (53-T) True('Bessie is blue.') ≡ Blue(Bessie)

In my view, (52-T) is true, since both sides of this biconditional are false. And if Bessie is not blue, then (53-T) is also true. But if a plausibly interpretive/cognized theory of truth has (52-T) as a theorem, then it is hard to see how the theory could fail to (1-T) as a theorem; and if it has (53-T) as a theorem, then it is hard to see how the theory could fail to have (52-T) as a theorem. It would be tendentious to *assume* that the left side of (53-T) is false no matter what color Bessie is. But even if one can consistently maintain (53-T), it doesn't follow that one should do so. The left side of (53-T) may be false because (53) doesn't have a truth condition any more than (52) or (1) does. Correlatively, we need not say that (54) is true.

(54) ‘Bessie is blue.’ is true if and only if Bessie is blue.

For we can say that (54) doesn’t have a truth condition any more than (1) does.

Let me stress this point. If (54) is a human language sentence—and so not confined to a written form governed by philosophical conventions—then it’s not at all clear that (54) is *true*, as opposed to a sentence that can be *used to report* the fact that (53) can be used to *make a claim* that is true if and only if the Bessie in question is blue in the relevant sense. It is often convenient to simplify, and say that (54) is true. But convenience is not an argument that instances of (55)

(55) *S* is true if and only if S.

are true, perhaps modulo context sensitivity; where ‘S’ is replaced by an English declarative and ‘*S*’ is replaced with that declarative quoted.

Even if ‘modulo context sensitivity’ can be cashed out plausibly for examples like (56),

(56) ‘I am hereby quoting him.’ is true if and only if I am hereby quoting him.

saying that instances of (55) are true is tantamount to adopting (D). *If* there are endlessly many true instances of some context-sensitive version of (55), as Davidson and many others suggest, *then* the obvious and perhaps only explanation for all these metalinguistic truths is that they follow from generalizations that constitute a truth theory for English. But the antecedent of this conditional is as much in question as (D) itself. Note that schema (55) has endlessly many instances like (57), whose quoted sentence can be used to report (some of) what Bert thinks.

(57) ‘Bert thinks that Ernie snores.’ is true if and only if Bert thinks Ernie snores.

If all these English biconditionals are true, then presumably, they follow from a finite number of basic principles that constitute a truth theory for English. But that’s no *argument* for the hypothesis that some theory of truth for English accommodates attitude reports.

Initially, it might seem that truth conditional semantics can be defended as the best explanation of the “fact” that instances of (57) are true, modulo context sensitivity. But this description of the explananda is tendentious. Liar T-Sentences like (58) highlight the point.

(58) ‘Lari is not true.’ is true if and only if Lari is not true.

This instance of schema (55) might seem to be true if you don’t know what Lari is. But upon reflection, (58) seems to be a counterexample to (55). So one can’t insist that (54)

(54) ‘Bessie is blue.’ is true if and only if Bessie is blue.

is true on the grounds that all instances of (55) are true. On the contrary, one needs to take seriously the possibility that (54) is no more true than (1).

If **H** is a Tarski-style language, then **H** is related to the world in a certain (perhaps sophisticated) way. If **H** is a human language, then a human child has the capacities required to understand each expression of **H**, even if these capacities are quite unsophisticated. One can hypothesize that human biology threads this needle, letting members of our species acquire Tarski-style languages without special training. But prima facie, Liar Sentences tell against this hypothesis. And as we’ll see, “repairing” a simple theory—so that it is plausibly true—can be at odds with preserving its virtues as a hypothesized account of linguistic meaning.

2.4 Complicating Complicates

It is tempting to look for a theory that has theorems like (53-L) and (1-LT);

(53-LT) Legit(‘Bessie is blue’) ⊃ [True(‘Bessie is blue.’) ≡ Blue(Bessie)]

(1-LT) Legit(‘Lari is not true’) ⊃ [True(‘Lari is not true.’) ≡ ~True(Lari)]

where both of these conditionals are true, as is the antecedent of (53-LT), but the antecedent of (1-LT) is false. Let’s not worry here about how to spell out the technical notion ‘Legit’—e.g., via fixed points (Kripke 1975), revisions (Gupta & Belnap 19xx), or whatever. For even if all the troublemakers can be filtered out, two related concerns remain *if* the goal is to defend (D).

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

First, it seems that any such filtering will yield consistency at the cost of explanation. If there are endlessly many English sentences such that the proffered theory does not assign interpretations to them, then we are left with the same kind of explananda that we started with. In thinking about human languages, the theoretical task is not merely to explain how a finite mind can understand boundlessly many expressions. One wants to know, among other things, how we humans understand the expressions that we do understand. If a theory of meaning accommodates endlessly many expressions like 'chased a cow' and 'chased every brown cow', but not 'chased a cow today', then the theory does not accommodate 'chased a cow' in the right way. And if a truth-theoretic account of meaning assigns an interpretation to (53) but not (15),

(53) Bessie is not blue.

(1) Lari is not true.

that is worrisome, since (1) is as meaningful/comprehensible as (53).

Moreover, and this is the second concern, theorems like (53-LT) and (1-LT) don't seem to be interpretive. One can *say* that competent speakers recognize the antecedent of (53-LT) as true, that they discharge it to obtain (53-T), and that (1) is understood by analogical extension.

(53-T) True('Bessies is blue.') ≡ Blue(Bessie)

But if speakers can discharge the antecedent of (53-LT), then we need to revisit the concession that speakers tacitly deploy some truth theory Θ that is governed by a very weak background logic that does not generate uninterpretive instances of (39a); where **Γ** is itself a theorem of Θ.

(39a) True('Ernie snores.') ≡ Snores(Ernie) & **Γ**

And in this regard, it's important to remember that troublemakers like (1) cannot be filtered out in terms of *grammatical* properties; see Parsons (1974), Kripke (1975).

Adapting an example from Kripke, suppose that nine people wrote a sentence on a bit of paper that was put into an otherwise empty box called 'Box'. The first four people wrote (3).

(7) 2 + 2 = 5

The next four wrote (20). The ninth person considered three options: (7), (20) and (58).

(20) 2 + 2 = 4

(58) Five bits of paper in Box carry inscriptions of a sentence that is not true.

Once the last bit of paper is deposited, someone utters (58)—or another problematic sentence, say one that replaces 'Five' with 'More than four' or 'An odd number of'. Unlike (7) and (20), (58) is context sensitive, if only because it is tensed. But it's hard to see how any context sensitive element of (58) could track which sentence the ninth person chose to inscribe. Yet the *utterance* of (58) is false, true, or neither depending on this choice. So prima facie, (58) does not itself have a truth condition, not even if we fix values for any context sensitive elements.

If the ninth person inscribed (7) or (20), it might seem that (58) is false or true. But if that person inscribed (58), we see the problem with assuming that (58) has a truth condition. One can say that (58) is Legit in many but not all contexts. But then one needs to say how and when speakers discharge the antecedents of conditionals of the form LEGIT(S, C) ⊃ [TRUE(S) ≡ P].

Lycan (2012) recognizes that there is no syntactic characterization of the human language sentences that are Liarish troublemakers for truth-theoretic conceptions of meaning. But he offers a variant proposal: English is not *a* human language; rather, "English" is a hierarchy of sublanguages, each of which has its own proprietary truth predicate. The idea is that each speaker of English has a core competence that can be extended with a kind of metacapacity.

Think of the core competence as one that lets us generate (use, and comprehend) the expressions of a basic language λ that is like the many dialects/idiolects we think of as versions of English, except for having a lexicon that contains no semantic vocabulary. The meta-capacity lets us generate the expressions of a closely related language λ′ that includes some semantic vocabulary, and in particular, a language-relative truth predicate 'T_λ' that is true of—or if you prefer, whose extension includes—all and only the sentences that are true-in-λ. For each expression of λ, there is a *correspondingly structured and homophonous* expression of λ′; and according to Lycan, a sentence of λ is itself a sentence of λ′. So a sentence can be true-in-λ and true-in-λ′. Sentences of λ′ that are *not* sentences of λ cannot satisfy 'T_λ'. But each sentence that is true-in-λ′ satisfies the language-relative truth predicate '$T_{\lambda'}$' that appears in the language λ′′, whose expressions are said to include those of λ′. And so on.

On Lycan's view, "English speakers" have access to many truth predicates.[13] But these predicates are not mere homophones, since a common conceptual thread runs through the hierarchy; although one cannot replace '$true_\lambda$', in a sentence of λ′, with '$true_{\lambda'}$'. Lycan thus adopts a version of the idea that (1) is understood via analogy to sentences like (17).

(1) Lari is not true.

(17) Bessie is not blue.

The idea is that while (1) is not itself a sentence of any language in the "English hierarchy," replacing 'true' with '$true_\lambda$' yields a sentence of λ′ that is true-in-λ′ but not a sentence of the basic language λ; though one cannot say this in λ′. Replacing 'true' in (15) with '$true_{\lambda'}$' yields a sentence of λ′′ that is true-in-λ′′ but not a sentence of λ or λ′; and so on. This may not preserve the idea that English (or any language in the hierarchy) has a *truth* predicate, at least not if such a predicate needs to have the meaning of 'true' (or be true of all and only true things). But advocates of truth-theoretic semantics might say that semantic words—perhaps including lexical items like 'correct', 'honest', 'coherent', and 'heteronymous'—are special in this respect.

A deeper concern, in my view, is that Lycan's proposal requires a single sentence to be a *sentence of* different languages. As discussed in section four, I think this is at odds with the idea that human languages are I-languages in Chomsky's (1986) sense. Though as Lycan reminds us, Davidson held that a single utterance *u* might be transcribed as /empedəkliyz liypt/, classified by one interpreter as an utterance of the English sentence (59),

(59) Empedocles leaped.

and classified by another interpreter as an utterance of the German sentence (60).

(60) Empedocles liebt.

If *u* can be true-in-English iff Empedocles leaped, and also true-in-German iff Empedocles loved, then *u* can be true-in-English and true-in-German. So perhaps an utterance of (1) can be true-in-λ′, true-in-λ′′, etc. We don't view each utterance that might be transcribed as /bank/ as an act of uttering several words simultaneously. But in special cases, a person might perform a speech act that counts as "doing two things" with one utterance; and perhaps this point extends to cases of (unwittingly) uttering a sentence of both λ′ and λ′′.

[13] Only a few of which will ever be used in ordinary discourse. The suggestion is reminiscent of Davidson's (1986) characterization of linguistic communication in terms of "passing" theories that speakers can deploy and share, in contexts, by adjusting "prior" theories that may not ever be used without adjustments of some kind. See Ludlow (2011) for related discussion.

Still, I don't think a *sentence of* one I-language can be a *sentence of* another, at least not in the way that Lycan requires. It's hard enough to make sense of the idea that (17) is a sentence of *both* a basic I-language and a lexically enriched I-language. (Does each speaker of English have an I-language that lacks the word 'water'—and for any lexical items, an I-language that lacks them?) In any case, I don't think (1) can be a sentence of two human I-languages that connect the articulation of 'true' to different semantic properties: humans acquire procedures that generate meaningful expressions, not mere syntactic structures; there is no "meaning-neutral" analog of (1) that a human I-language can generate. Nonetheless, Lycan has a serious proposal. So I'll come back to it, after an initial discussion of context sensitivity and Davidson's (1967b) claim that theories of truth for human languages will have theorems concerning *utterances*.

3. Relativizations or Actions

One might have hoped that the right account of context sensitivity will help in dealing with Liar Sentences. But as we'll see, the two main strategies for accommodating sentences like (34)—

(37) I chased something.

relativizing the truth of sentences to contexts, or specifying truth conditions for sentential utterances—heighten the worry that no truth theory for a human language **H** is the core of a correct meaning theory for **H**. Focusing on truth favors the second strategy, according to which certain *actions* are truth-evaluable. Focusing on linguistic meaning favors the first strategy, according to which *expressions* are the real bearers of semantic properties, which are relativized. Thinking about the contrast reveals a tension in the idea that a single theory, concerning ordinary human beings, can be a good theoretical characterization of both truth and meaning.

3.1 Two Kinds of Theorems

As briefly discussed in chapter zero, we can extend a Tarskian language like PL (discussed in chapter two) in a Kaplanian way that yields models of sentences like (61) and (62).

(61) I chased it.

(62) I am here.

Add three indices—s, p, t—and a pointer ☞ such that each pointer can be extended, with '′', to create another. Then expand each Tarskian sequence by adding a zeroth element that is itself an ordered triple of domain entities. Each expanded-sequence, or *assignment* of values to variables, is of the form: $\langle\langle e^s, e^p, e^t\rangle, e, e', e'', ...\rangle$. For each assignment $\mathcal{A}$: $\mathcal{A}(s)/\mathcal{A}(p)/\mathcal{A}(t)$ is the first/second/third element of $\mathcal{A}$'s zeroth element; and each pointer has the value of the corresponding variable—$\mathcal{A}(☞) = \mathcal{A}(x)$, $\mathcal{A}(☞') = \mathcal{A}(x')$, etc. This allows for object language sentences like (61a) and (62a), along with metalanguage theorems like (61b) and (62b).

(61a) CHASED(s, ☞)

(61b) Satisfies[**A**, CHASED(s, ☞)] ≡ $\mathcal{A}(s)$ chased $\mathcal{A}(x)$

(62a) LOCATION(s, p)

(59b) Satisfies[$\mathcal{A}$, LOCATION(s, p)] ≡ $\mathcal{A}(s)$ is located at $\mathcal{A}(p)$

This leaves room for further analysis.[14] But whatever the details, (58a) and (59a) are not true or false, since these sentences are satisfied by some but not all assignments. One can, however, say that contexts *license* assignments. Suppose that whatever contexts are, each context

[14] See chapter three. For example, (61a) might be decomposed as $\exists x'$((BEFORE(x', t) & CHASE(x', s, ☞)), which is satisfied by $\mathcal{A}$ if and only if: some $\mathcal{A}^*$ such that $\mathcal{A}^* \approx_{x'} \mathcal{A}$ is such that $\mathcal{A}^*(x')$ occurs before $A^*(t)$, *and* $A^*(x')$ is a chase by $\mathcal{A}^*(s)$ of $\mathcal{A}^*(x)$; or abbreviating, something "was" a chase by $\mathcal{A}(s)$ of $\mathcal{A}(x)$.

determines exactly one speaker, place, and time, along with some (perhaps null) n-tuple of deictic actions. Then one can say that context **c** licenses assignment $\mathcal{A}$ if and only if: $\mathcal{A}(s/p/t)$ is the speaker/place/time of **c**; and if Δ is the *i*th deictic action determined by **c**, then the *i*th element of $\mathcal{A}$ is the domain entity demonstrated via Δ. If **c** determines a deictic action in which nothing is demonstrated, then **c** does not licence any assignment. But bracketing this possibility, one might adopt the context-relative notion of truth characterized in (63).

(63) True(S, **c**) ≡ for each assignment $\mathcal{A}$ that **c** licenses, Satisfies($\mathcal{A}$, S)

If each conversational situation determines a unique context, then modulo vagueness regarding the relevant speaker/place/time, one can say that a sentence S is true in that situation if and only if S is satisfied by each assignment that is licensed by the determined context. This allows for various conceptions of how (knowledge of) meaning is related to (knowledge of) pragmatic principles concerning the use of expressions in conversational situations.

This shows, among other things, that mere inclusion of an indexical or deictic expression does not preclude a sentence from having a context-relative truth condition. Given a Kaplanian extension of a Tarskian language, T-theorems still concern the truth conditions of *sentences*. And one can hypothesize that human languages are variants of some such extension of PL. But it doesn't follow that (61) and (62) are true or false, relative to contexts or in situations, since human language sentences may not be truth-evaluable at all. Providing *models* for (61) and (62) does not show that these sentences have truth conditions relative to contexts. Even ignoring the possibility of failed acts of demonstration, issues concerning chases and event variables remain; see chapter three. Moreover, even if objections specific to (58) and (59) can be addressed, our question is whether there are meaning-specifying theories of truth for human languages, not whether sentences like (61) and (62) themselves favor a negative answer.

It is also worth remembering that according to Davidson (1967b), a truth theory for a human language will have theorems concerning *utterances*; cp. Burge (197x), Higginbotham (1985), Larson and Segal (1995), Lepore and Ludwig (200x). Regarding (61), one might suggest generalization (61c); where 'True[*u*, S]' means that *u* is a true utterance of sentence S,

(61c) True[*u*, CHASED(*s*, ☞)] ≡

the author of *u* chased the (first and perhaps only) thing demonstrated in *u*

and an utterance of (61) can be an utterance of CHASED(*s*, ☞). The idea is that an utterance is an *action* that can be described as a spatiotemporally located articulation, by a speaker in a place at a time, of a sentence that has a certain logical form. For many purposes, the D(avidson)-strategy and the K(aplan)-strategy are interchangeable. But here, I want to stress that they reflect different ways of thinking about how human languages and their users are related to truth and contexts.

On the K-strategy, a theory of truth for a language is a theory of a *language*. And for these purposes, one can even take languages to be sets of expressions rather than generative procedures. The important point here is that on the K-strategy, theories of truth are theories of linguistic objects, independent of how they are used. Uttering a sentence, in a certain situation, may determine which assignments are licensed on that occasion of use; and in this sense, truth depends on use. But the context-sensitive truth conditions of a sentence S are specified without reference to actions of using S, except perhaps with regard to deixis.

By contrast, on the D-strategy, a theory of truth for a language is a theory of certain *spatiotemporally located actions of using* the language. The theory ascribes truth conditions to *utterances of* generable sentences. This approach may locate truth in a more plausible place. But

it is no small task to formulate *axioms* from which biconditionals concerning utterances *follow*, given a weak logic that avoids unwanted/uninterpretive theorems.

Accommodating quantificational constructions like (37) is hard enough,

(34) I chased something.

even with logical forms like (37a), given the need for sequence relativization as shown in (37b).

(37a) $\exists$X(CHASED(s, X))

(37b) Satisfies[$\mathcal{A}$, $\exists$X(CHASED(s, X))] $\equiv$

some $\mathcal{A}$ * such that $\mathcal{A}^* \approx_x \mathcal{A}$ is such that $\mathcal{A}(s)$ chased $\mathcal{A}^*$(x)

So, one might prefer a Kaplanian theory that delivers theorems like (37b) rather than (37c),

(37c) True[u, $\exists$X(CHASED(s, X))] $\equiv$

some $\mathcal{A}^*$ such that $\mathcal{A}^* \approx_x \mathcal{A}$ is such that the author of u chased $\mathcal{A}^*$(x)

here ignoring event variables; cp. note 14. The added complexity of (34c) has to be reflected, one way or another, in axioms from which such theorems are allegedly derived. Specifying sentence meanings in terms of truth-evaluable utterances also seems to confuse linguistic competence with language use, even if we bracket utterances that are neither true nor false.

Following Wittgenstein and others, one can hypothesize that linguistic competence *is* competence regarding the use of expressions, as opposed to knowledge of expressions whose properties constrain how they can be used. But recognizing implications, for example, does not seem to be a matter of generalizing across utterances. Moreover, many—perhaps most—uses of human linguistic expressions are intrapersonal. And even if truth-evaluable utterances play a central role communication, truth-evaluable judgments may not play an analogous role in cognition. One can speculate that using expressions to entertain thoughts ("just thinking") is derivative on using expressions to endorse thoughts. But prima facie, it's the other way around.[15]

It can be tempting to say that languages are for making assertions. Though we shouldn't assume that human languages are "for" any particular use, especially if they turn out to be I-languages in Chomsky's (1986) sense: biologically implemented expression-generating *procedures* that pair articulations with meanings; see section four. One can hypothesize that human I-languages pair articulations with mental analogs of (37c), as opposed to (37b), despite intrapersonal uses of I-languages. But another possibility is that I-languages are used for many purposes, including the generation and construal of certain utterances, without speakers encoding any theory of *meaning* that employs a truth predicate that applies to utterances. In short, it may not be independently plausible that I-language expression meanings are specified by axioms concerning utterances (or any truth-evaluable uses) of the expressions in question.

3.2 Evaluating Actions

On the other hand, the D-strategy can seem reasonable if the goal is to specify theories of *truth* for human languages. With regard to such languages, it may well be that truth is more plausibly ascribed to certain *actions of using* sentences than to sentence-context pairs.

From a Strawsonian perspective, (58) vividly exemplifies a general point.

(58) Five bits of paper in Box carry inscriptions of a sentence that is not true.

A declarative sentence S can fail to be true or false, even given values for any variables in S. One can say that *sentences* like (64) carry presuppositions—e.g., that the name denotes something—

(64) Vulcan is bigger than that.

[15] Perhaps all mental uses of language are somehow derivative on episodes of interpersonal communication. But I don't of any good reasons, as opposed to "private language arguments," for thinking that this is so.

and that derivable specifications of truth conditions are correspondingly conditionalized. But then the allegedly meaning-specifying theorems (and axioms) are correspondingly complicated. Moreover, while some presuppositions seem to be grammatically constrained, it is hard to account for even these constraints on truth-evaluability in terms of properties of the expressions themselves, as opposed to general knowledge of how expressions can be used to convey information; see Schlenker (2009). In any case, it's hard to see what presupposition of *sentence* (58) is violated if one bit of paper in Box contains an inscription of (58).

Prima facie, some constraints on truth-evaluability are constraints on actions of using sentences to make assertions or judgments in particular contexts. I think that many observations regarding vagueness also point in this direction. In a context where it is clear what counts as being green—or least as clear as it can be for normal humans—there can still be vague cases. Following Sainsbury (1996) and many others, I conclude that 'green' does not have an extension, not even relative to contexts. No set is such that for each thing x, x is an element of that set if and only if x is green. And I see no independent reason for thinking that the meaning of 'green' determines a function from contexts to extensions, as if 'green' contains a covert variable whose values correspond to ways of replacing a vague predicate with a precise one.

Supervaluation can, of course, be useful for certain purposes; see, e.g., Fine (19xx). In particular, when humans try to make truth-evaluable claims by using expressions that do not have extensions, vague cases can present complications or even paradoxes. In some theoretical contexts, it can be helpful to regiment such usage and bracket certain questions about meaning. This may improve our understanding of how the practice of *making an initial thought more precise* is itself sensitive to context. But as Wright (19xx) and others have discussed, many uses of human language predicates are intentionally *tolerant*. In using 'green', we often want to avoid making an assertion whose truth or falsity could depend on a distinction that ordinary humans cannot make.[16] In many contexts, a person who thinks that Kermit is green can be described as someone who endorses the truth-evaluable proposition that Kermit is green. But such description does not commit us to saying that 'green' is true of Kermit.

These considerations suggest that the true/false distinction, like some other right/wrong distinctions, targets actions that meet the conditions for being evaluable in a certain way. We can idealize, say that human language sentences are true or false relative to contexts, and then speak of propositions as abstracta that are unrelativizedly true or false. Given episodes of thinking, we can speak of thoughts thought. Given episodes of judging, we can speak of propositions judged true; see, e.g., Cartwright (1962). But it doesn't follow that when actions of assertion/judgment are true or false, they are truth-evaluable *by virtue of* having propositional contents, much less that theories of meaning should specify such contents. If truth lies downstream of meaning—in that acts of using sentences are *candidates* for being true or false, subject to further review—then perhaps theories of meaning should not deploy the predicate 'True(_)', even if humans enjoy a mental predicate TRUE(_) that applies to certain assertions/judgments. That leaves the hard task of describing the (presumably context sensitive) conditions on being true or false, and explaining

[16] Similar points apply to the practice of *making laws* by using words like 'painful', 'lethal', or 'intentional' in official contexts. See Graff (2000) for important related discussion. Note that inscribing '{x: x is green}' does not establish that the inscription has a denotation; cp. '{x: ~(x ∈ x)}'. See Pietroski (2005a) for related discussion in the context of Boolos' (1998) conception of plural quantification. Perhaps when doing logic, uses of concepts are subject to normative constraints according to which GREEN(_) is to be used as if it has an extension, even if don't know which extension this is; see Williamson (1994). But even given some version of "epistemicism" for ideal(ized uses of) concepts, it doesn't follow that the human language word 'green' has an extension.

why these are conditions on truth/falsity; but see Glanzberg (2004, forthcoming) for a promising start.

One can agree with Davidson that a *truth* theory for a human language **H**, if such there be, will have theorems that specify the truth conditions of certain spatiotemporally located *actions* of using expressions of **H** (rather than use-independent properties of those expressions). But one might also think that a good theory of *meaning* for **H** needs to have theorems that specify use-independent properties of the boundlessly many expressions of **H**; where these features of **H**-expressions constrain and partly explain how the expressions can and cannot be used. For purposes of studying specific constructions like (68),

(68) Two chipmunks chased each other around the park for an hour.

it may not matter which way advocates of (D) accommodate indexicals and demonstratives.[17]

(D1) for each human language **H**, there is a correct theory of truth for **H**.

But in evaluating (D), one cannot ignore questions about how meaning and truth are related,

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

or how they are jointly related to sentences, utterances, and contexts.

Recall that Russell's (190x) notion of a proposition led Frege (190x) to ask if concrete particulars like Mont Blanc could be constituents of truth-evaluable entities. Likewise, Strawson rejected the idea that meanings are referents. From his perspective, Russell mischaracterized the sense in which judgments about mountains are spatiotemporally located: the judgeable *contents* do not have constituents that can be climbed. In this respect, judgments about mountains and their snowfields are as abstract as judgments about numbers and their successors; contents are individuated finely, like *ways of thinking about* things (see chapter one). But an *episode* of judging that every number has a successor is still spatiotemporally located, and evaluable as true, even though the constituent concepts let us think about abstracta in certain ways.

Like 'thought', 'expression' exhibits symbol/episode polysemy. An invented formula can be an expression of PL. Writing a note of condolence can be an expression of sympathy. Human sentences are expressions, as are many actions of using such sentences. This can be confusing, since acts of endorsing or conveying truth-evaluable thoughts can be evaluated in many ways. A sentence can be used to make judgments/assertions that are rash, rude, rebellious, true, etc. But are the sentences themselves rash, rude, rebellious, or true relative to contexts? In addressing such questions, we need to think about what human languages and their expressions *are*.[18]

One can hypothesize that humans have both a capacity for acquiring I-languages, and an ancillary system that specifies—or in some other sense determines—truth conditions for certain *uses of* I-language expressions that *already* pair articulations with meanings. And one can try to specify these meanings, which human languages pair with articulations, in truth-theoretic terms via some version of the Kaplan-strategy that allows for certain dimensions of context sensitivity. This is an interesting hypothesis, especially given an explicit proposal about the requisite *very weak* background logic deployed to generate theorems like (21-T); see section two.

(21-T) True('Ernie snores.') ≡ Snores(Ernie)

---

[17] Many of the virtues and vices translate. See Glanzberg (201x) for related discussion in the context of comparing truth-theoretic and model-theoretic conceptions of semantics; cp. Lepore (1983).

[18] It may be no accident that when Davidson (1986) addressed the question of what truth theories are theories of, he concluded that they are not theories of languages in any traditional sense; see Pietroski (1994) for discussion.

But then one can't deal with Liar Sentences by saying that truth and falsity are really properties of spatiotemporally located assertions/judgments, and that truth theories for human languages assign semantic properties to utterances *rather than* expressions relativized to contexts.

3.3 Contexts, Kettles, and 'That'-Clauses

In chapter zero, I reviewed some reasons for thinking that the truth or falsity of a thought expressed with a sentence like (69) or (70) can depend on the context of use in subtle ways.

(69) The kettle is black.

(70) Sam denied that Hesperus is Phosphorus.

I want to recapitulate this point before getting back to Liar Sentences in section four. While this may seem like a digression, my intention is to connect the details of this chapter with the larger theme of the book.

In my view, the meaning of a human linguistic expression Σ—μ(Σ) for short—is an instruction that determines the *form* of a concept assembled by executing this instruction. A concept that is assembled by executing μ(Σ) in a given context can also be a concept that is expressed by using the pronunciation of Σ—μ(Σ) for short—in that context. But executing μ(Σ) on other occasions may lead to different results if at least one of Σ's lexical constituents has more than one fetchable concept at its address. Lexical meanings *constrain without determining* concept choice. In this sense, the concept associated with Σ can vary across contexts, even if no index in Σ tracks this dimension of variation. Likewise, the sound associated with π(Σ) can vary across contexts, even if no index in Σ tracks the difference between shouting quickly at the top of one's vocal range or speaking slowly and softly at the bottom of one's vocal range.

If a word like 'France' (or 'Venice') can be used to access a concept of a certain geographical region, or a concept of a certain political unit, then we can make sense of why sentence (71) is weird in a way that (72) is not.

(71) France is a hexagonal republic.

(72) France is hexagonal, and France is a republic.

For each use of 'France' in (72), there is a way of executing the instruction that leads to a reasonable thought; but not so for the single use in (71). This also helps explain why inferring from (72) to (71) feels wrong, or at least risky. A thought constructed in accord with (72) can be true; but a thought constructed in accord with (71) is likely to be false, at least in many contexts. But if 'France' does not determine which concept is to be fetched via this proper noun, then (73)

(73) The author of *Conjoining Meanings* likes France.

expresses neither a thought about certain terrain nor a thought about a certain nation. More generally, I claim, there is no thought that (73) expresses and no truth condition that (73) has.

In a particular context, there may be reasons for using 'France' to access the political concept. But even in such a context, I don't think the meaning and context jointly determine the concept accessed, unless one builds the accessing into the context. And I see no reason for thinking that 'France' somehow includes an index whose character somehow tracks which concept one *ought* to access in a given context. In many contexts, there may be a fact of the matter concerning which concept one ought to access for purposes of reasonable conversation; and the meaning of 'France' certainly constrains the space of possible (human) answers to the question of how this instruction should be executed in a given contex. But the meaning/instruction is not itself a mapping from contexts to answers.

Meaning constrains the use of expressions while leaving room for various kinds of cognitive freedom. Travis (199x) makes this point with examples like (69).

(69) The kettle is black.

And while he does not describe lexical meanings as instructions for how to access concepts, I think an important moral of his discussion can be put in these terms: when a hearer encounters a sentence that contains a certain lexical item L, she often has a choice regarding which concept to access with L; and often, her choice will be influenced by the rest of the sentence and the larger context. This suggests that we need to distinguish the compositionally determined meaning of a sentence from the use of that meaning in constructing a potentially truth-evaluable thought.

Chomsky (1977, 2000b) makes similar points; see Pietroski (2003, 2005b). Carston (2002) speaks of "ad hoc concepts," in order to stress that in speech or comprehension, one may restrict or relax available concepts for specific temporary purposes; see also Wilson and Sperber (200x). This is fully compatible with my conception of meaning. Lexical addresses do not harden and refuse to admit new residents. On the contrary, if children use some concepts to introduce others in the course of lexicalization, then we shouldn't be surprised if the use of a lexical item L in communication often involves using the concepts at L's address to introduce yet another concept whose utility may be fleeting. Indeed, the processes that support lexicalization may be continuous with those that support communication; cp. Davidson (1986), Ludlow (201x).

One wants to know *how much* freedom we have to introduce concepts in ad hoc fashion. One also wants to know what concepts are. But our ignorance is not a reason for skipping over concepts and saying that lexical items map contexts to extensions. That said, let me stress that I take meanings to be instructions, not their executions. A hearer might recognize the meaning of an uttered expression Σ, use Σ to build a concept, yet also recognize that the speaker would use Σ to build a different concept of the specified form (with constituents that are somehow more restricted/relaxed than those currently available to the hearer). Sometimes, understanding a speaker's *words* triggers appreciation of the fact that the speaker uses those words to access concepts that are subtly but perhaps importantly different than any concepts one currently has access to. Reading philosophy seems like a paradigm case.

It is a fantasy—arguably a horrific one—that all speakers of English attach the *same* concepts (or the same mappings from contexts to extensions) to English expressions. Even ignoring the differences across I-languages that count as English, acquiring such an I-language leaves room for variation in the concepts that individual speakers connect with particular lexical items; and it leaves room for choice with respect to which concepts one accesses on a given occasion of using some lexical items to access some concepts. Pace Dummett (1986), this does not imply that successful communication is accidental in any worrying sense. As members of the same species, humans share many cognitive dispositions. The fact that other primates do not acquire pronounceable lexical items, while children acquire them quickly and with apparent ease, suggests specialized circuitry that every human enjoys. (If we meet intelligent aliens who are not biologically based, talking with them may require more luck.)

Speakers of English connect lexical items with concepts in ways that are good enough for successful human communication, at least often. But success differs from accuracy; and for most practical purposes, I prefer success. The concept a speaker uses, on a given occasion of using 'black' to talk about a kettle, probably doesn't have the same extension as the concept that any hearer uses. That doesn't preclude (tacit) agreement that the kettle in question counts as black for the purposes at hand. Correlatively, while human linguistic communication is often successful in many ways, this does not justify the claim that 'black' maps contexts to extensions.

One can hypothesize that lexical items have (or are) indices that have Kaplanian characters. But even if the lexical item 'black' is always used with an accompanying index *i*, yielding a complex homophonic expression of the form [black-*i*], that leaves the task of

specifying the meanings of the lexical item and the index. This task is difficult, even if we assume that [black-*i*] is of the Fregean type <e, **t**> and impose *no* constraints on lexical types. This allows the index to be of any Fregean type <α>, with the lexical item being of the corresponding type <α, <e, **t**>>. But however one encodes the alleged dimension of contexts, allegedly tracked by the alleged index, one still has to encode the context-invariant contribution of the lexical item. Moreover, even if this unconstrained homework problem is soluable, there *are* constraints on lexical types, even if we don't yet have a good account of them. Positing indices of any unusual type—say, other than the one(s) posited for 'I', 'here', and 'now'—raises the question of whether human I-languages permit such indices; see Chierchia (198x).

One can speculate that all open-class lexical items combine with unpronounced indices of type <e>. But human I-languages may not permit lexical items that express the requisite *relations* between entities like kettles and the (presumably abstract) entities that allegedly reflect dimensions of contexts (e.g., the alleged blackness-dimension) in the way that 'I' reflects the speaker-dimension of contexts. Even if theorists can specify such relations, and stipulatively assign them to invented models of lexical items, it hardy follows that lexical items specify such relations. I can't prove that children don't understand 'black' in terms of a dyadic relation that a kettle can bear to some entity that effectively selects one of the potential extensions for 'black'. But prima facie, the context dependence of truth goes beyond the context sensitivity of meaning.

As noted in chapter zero, reports made with sentences like (70)

(70) Sam denied that Hesperus is Phosphorus.

present analogous but more severe difficulties for thesis (D).

(D) For each human language **H**, there is a theory of truth
that is the core of correct theory of meaning for **H**.

This is so, even if the "opacity" of 'that'-clauses can be accommodated. Initially, one might think that the context sensitivity of indirect speech reports (as opposed to direct quotation) is due to a special composition rule—a rule that has the effect of introducing a covert index—for phrases in which a verb combines with a 'that'-clause. It can also be tempting to say that combining an adjective with a noun, as in 'black kettle', has the effect of introducing an index; though even if this helps with (74), (75) also illustrates the context sensitivity of 'black'.

(74) The black kettle is on the table.

(75) The kettle on the table is black.

And while indirect speech reports may be special cases, in several respects, the phenomenon of opacity is not a license to posit a composition rule that is sensitive to context. More generally, the context sensitivity of truth is not a license to assume that meanings are compositional only in a weak sense that is compatible with (D) given examples like (70) and (75).

We can invent languages that allow for sentences like (76),

(76) Sam denighed glonk that Hesperus is Phosphorus.

in which 'denied' is a ditransitive verb whose indirect object 'glonk' is an index that tracks which substitutions within 'that Hesperus is Phosphorus' are permissible in the relevant context. We can stipulate that 'denigh' is true of <ε, β, γ> relative iff ε is an event of denial whose "content" is β relativized to γ; where β is the sort of thing denoted by a 'that'-clause, and γ is a potential value of the index 'glonk'. Then one can say that (76) is true relative to a context **c** iff some denial is such that it was done by Sam prior to **c**('now'), and its content is the denotation of 'that Hesperus is Phosphorus' as relativized to **c**('glonk'). One can hypothesize that (70) has the

same truth condition.[19] This is to posit a sophisticated and hard-working composition rule for phrases in which a verb combines with a 'that'-clause, raising the question of whether human linguistic meanings *can* combine in this way—and if they can, why this kind of sophisticated and hard-working composition rule is not more often attested in human languages.

For these purposes, let the meaning of 'that Hesperus is Phosphorus' be whatever you like. I take 'that'-clauses to be instructions for building mental predicates of a special sort: predicates that apply to certain *concepts* that have been prepared for declarative (as opposed to interrogative) use; see Lohndal and Pietroski (2011). As discussed in chapter six, I think the meaning of 'deny that Hesperus is Phosphorus' is an instruction for how to build a concept of the form DENY(_)^∃[CONTENT(_, _)^Φ(_)]; where CONTENT(_, _) is a thematic concept akin to PATIENT(_, _). But if you like, let 'that Hesperus is Phosphorus' denote some fine-grained proposition, perhaps the Fregean sense of an idealized version of 'Hesperus is Phosphorus'.

The point here is not merely that the meaning of (70) *differs* from that of (77).

(77) Sam denied that Hesperus is Hesperus.

The point is that the meaning of (70) does not determine a mapping from contexts to truth values. In light of work by Davidson (1967a) and others, we should not reject thesis (D) just because certain substitutions do not preserve truth. Obviously, (78) does not imply (79).

(78) Slim is so called because he is thin; and Slim is Jim.

(79) Jim is so called because he is thin.

Likewise, (80) does not imply (81), and (82) does not imply (83).

(80) Slim is Jim; and Kim said, 'Slim is thin'.

(81) Slim is Jim; and Kim said, 'Jim is thin'.

(82) Slim is Jim, and Kim said so.

(83) Jim is Jim, and Kim said so.

But these nonimplications just highlight the fact that human languages allow for logophors that are in many respects like markers for direct quotation; see Quine (195x), Cartwright (1962), Forbes (199x), Pietroski (1996). So even substitution of synonyms is not always valid. We can say that 'lawyer' does not rhyme with 'attorney'. And while quotation raises puzzles, I don't think that the quotational aspects of (78-83) present especially deep difficulties for (D).[20] But this isn't to grant that (77-84) have compositionally determined truth conditions.

(84) Kim said that Slim is thin.

Suppose that Kim uttered (85) at some time *t*, using 'James' as a name for Slim/Jim.

(85) James is gaunt.

Then an utterance of (84) might be true. If the intended audience for (84) only knows the man in question as Slim, and Kim used 'gaunt' only after checking a thesaurus, then (84) may be a good tool for reporting Kim's utterance of (85). But in a different situation, an utterance of (84) might be false. If Kim's choice of 'gaunt' as opposed to 'thin' was medically important, and everyone

---

[19] Cp. Schiffer (1992), Ludlow (199x). Citations to Higginbotham, Segal, Larson/Ludlow on appeals to similarity.

[20] See Cappelen and Lepore (2005). Sentence (89) can be pronounced as 'Kim said, quote, Slim is thin'. And many human languages have a distinctive complementizer for direct speech reports; see Munro (198x) for interesting discussion. Of course, complementizers differ from demonstratives in many ways. In particular, a complementizer has to target the sentence it introduces, as illustrated with (92), which does not have the meaning of (92a); see Segal and Speas (1986), Seymour (1994).

(92) Slim is thin. His mother said that he needs to gain weight.

(92a) #Slim is thin. His mother said that. He needs to gain weight.

But the spirit of Davidson's account does not require that complementizers be demonstratives; see Pietroski (1996).

thinks that Slim is Jim's overweight cousin, then an utterance of (84) would not correctly report Kim's speech act. Likewise, if Kim uttered (86), some utterances of (87) might be true,

(86) Slim is thin.

(87) Kim said that James is thin.

while other utterances of (87) are false.

Davidson himself stressed such points, and so appealed to a notion of "samesaying" that dovetailed with his inclination to ascribe truth to certain *actions* rather than sentences relativized to contexts. On his view, an utterance of (84) is true iff the embedded utterance of (86) "samesays with" some action of saying done by Kim. As noted in chapter zero, one can recode this idea by saying that 'said' expresses a relation that sayers bear to sentences relativized to contexts. But there remains the question of what it is to "say" a sentence relative to a context. So while I see how to invent languages that allow for opacity, that doesn't speak to real challenge that sentences like (87) and (75) present for the idea that such sentences have truth conditions.

(75) The kettle on the table is black.

3.4 Contexts and Commonsense

One can try to argue that sentences of a human language contain the indices required by (D),

(D) for each human language **H**, some theory of truth
is the core of a correct theory of meaning for **H**

by offering independent arguments for (D) *and* the further claim that the most plausible version of (D) is one that preserves the following idea: if an utterance of some sentence S is true in one context, but an utterance of S is false another context, then S contains an index that has different values relative to the different contexts; see Stanley (2000, 2002). But then the argument for the requisite indices relies on defense of a thesis *bolder* than (D). And note that even if sentences of a human language include *enough* indices to track the ways in which truth can depend on context, it doesn't follow that these indices can take the requisite *values*, or that human languages can support the requisite modes of composition. And given (C),

(C) each human language is an I-language in Chomsky's sense

limitations on lexical meanings and semantic composition may severely constrain the possible values of any posited indices.

As noted in chapter one, and further discussed in chapter six, it's hard enough to figure out which conjunction and closure operations are employed in human semantic composition. But these biologically implemented operations are surely limited in ways that impose constraints on the space of possible meanings for expressions (atomic and complex) of human I-languages. So to argue for covert indices by arguing for (D), one needs to argue for (B)

(B) each human language **H** is an I-language such that
some theory of truth is the core of a correct theory of meaning for **H**

without *assuming* that human I-languages generate expressions with the requisite indices. If (B) was independently plausible, modulo a few special cases, it might be reasonable to posit some otherwise unmotivated indices. But if every sentence of a human language seems like a counterexample to (B), unless one idealizes away from the problematic aspects of that sentence, then I don't see how to use (D) as a bootstrap for justifying appeal to covert indices that help respond to some—but by no means all—of the objections to (B).[21]

[21] At this point, debate can degenerate into trading hunches about whether to describe particular sentences as counterexamples to (D), or as invitations to pursue Stanley's strategy, treating every example as a potential research program. So I don't think it will be useful to rehearse more examples. But there remains the challenge of defending

Some advocates of (D) explicitly reject appeal to covert indices, absent independent motivations for such appeal, and conclude that a single utterance of a sentence can be an act of expressing more than one proposition; see Cappelen and Lepore (2005). If 'black' contains no index, one can hypothesize that each utterance of (69)

(69) The kettle is black.

has a "minimal" (or primary) truth condition that can be specified in some preferred theoretical context, but that an utterance of (69) can also have *another* truth condition that is more specific to the context of that utterance. But if *uses* of 'black' exhibit *various* truth-theoretic properties, why think that the lexical item also exhibits a *truth-theoretic* property that is constant across such uses? Why posit the alleged minimal/primary truth conditions?

In principle, one might propose algorithms for computing "secondary propositions" from "primary propositions" and representations of certain aspects of situations that speakers recognize as relevant, despite the absence of corresponding linguistic indices. But as Cappelen and Lepore admit, examples like (69) leave little hope of specifying plausible algorithms of this sort. Cappelen and Lepore defend their proposal as the best option for advocates of (D), absent evidence that words like 'black' include indices that track the relevant aspects of contexts. But this is to admit boundlessly many cases of a sentential utterance having a truth condition that is *not* characterized by a theory of truth/meaning for the language in question. Prima facie, this threatens the conception of meaning characterized by (D).[22] And especially given thesis (C), Stanley might argue that his Kaplan-esque strategy for defending (D) by appeal to covert indices—and taking truth to be a property of sentences relativized to contexts—is better than taking truth to be a property of utterances and saying that an utterance can have more than one truth condition. But even if one version of (D) is more plausible than others, that is not yet an argument for (D) or a response to the apparent counterexamples.

We can speak of first meanings, second meanings, and so on; see Davidson (1986). Or drawing an analogy to Marr's (1982) description of visual processing, one might say that executing what I call a meaning yields a "primal sketch" of a thought, which can be an input to further processing that yields a multi-dimensional representation—perhaps one that includes a Schein-style logical form of the sort discussed at the end of chapter three. But prima facie, the meanings that human I-languages connect with pronunciations do not track all the ways in which the truth of falsity of assertions (made in part by using the meanings) can depend on the context.

It is also worth remembering that examples like (69) and (88)

(69) The kettle is black.

(88) Al chased Theo.

were supposed to be cases that *motivate* (D), despite all the examples suggesting that the distinction between metaphorical and "literal" commonsense talk is vague and context sensitive. So if (69) and (88) already illustrate reasons for rejecting (D), that is significant, since ordinary speech is often much *less* friendly to (D). For example, consider the lines below, from Robert Mellin's lyric for 'My One and Only Love'.

(D) without positing powerful apparatus that is not independently motivated. And this challenge is heightened given an alternative conception of semantic composition that is more constrained.

[22] It seems like a Pyrrhic victory for Davidsonians if (D) is preserved by saying that (i) a single utterance of a sentence can have more than one truth condition, and (ii) the truth condition that matters for interpretation of the utterance—a human action—differs from the truth condition determined by the lexical axioms and composition principles, even given contextually determined values for any variables in the expression.

(89a) The very thought of you makes my heart sing
like an April breeze on the wings of spring.
(89b) And you appear in all your splendor.
(90a) You fill my eager heart with such desire.
Every kiss you give sets my soul on fire.
(90b) I give myself in sweet surrender.

Let's agree to set (89a) and (90a) aside as metaphorical. Hearts don't sing; neither are they eager or filled with desire. Spring has no wings. There are no souls; and even if there were, they couldn't be set on fire. So nobody owes a theory according to (89a) and (90a) are true in the relevant context. By contrast, while (89b) and (90b) are stylized, I don't think they are metaphorical, at least not in the ways that (84a) and (85a) are. One can extend the notion of metaphor so that it covers just about all ordinary discourse, and then try to characterize the apparent distinction between the (a)-cases and (b)-cases in some other way.[23] But if advocates of (D) end up saying that almost all sentences are false in most contexts, and ordinary intuitions of "correctness" are correspondingly unreliable indicators of truth, then we need to hear much more about the motivations for (D). So I assume that at least prima facie, advocates of (D) should count (89b) and (90b) as potential truths.

In which case, a person α—or α's appearance, or an event of α appearing—can be *in* all α's splendor, and so α has some splendor that x can be in; cp. talk of chases, sunrises, cites, and skies. Likewise, an event of β giving β (perhaps to α) can apparently be *in* sweet surrender. Regardless of whether these are examples of polysemy or homophony, a theory of truth for English would need to specify satisfiers for 'in' accordingly. In my view, disquotational specification is adequate. But in any case, (D) seems to require a domain that includes entities like appearances/givings that are somehow "in" things (e.g. splendor/surrender) in ways that are harder to comprehend than the corresponding expressions. And even if one posits the requisite "things," there remains the question of *how* 'in' could be true of the requisite ordered pairs.[24]

One can say that (89b) and (90b) are quasi-metaphorical special cases. But even if no single family of examples shows that (D) is false, each family can add to the overall argument that meaning is not so tightly connected to truth. In my view, the connection is indirect—via the use of assembled concepts that are used in context dependent ways—and in many cases, quite tenuous. I think Liar Sentences make it possible to see this point through a narrow but sharply focused lens. The argument that (1) has no truth condition approximates a proof.

(1) Lari is not true.

This doesn't establish that (53) has no truth condition.

(53) Bessie is blue.

---

[23] Cp. Lakoff (19xx). Perhaps "in the hush of night while you're in my arms" is doubly metaphorical: night has no hush to be in; and maybe the singer's beloved, unlike his blood vessels, cannot really be in his arms. For what it is worth, I think Davidson (197x) was right to describe metaphorical usage as a kind of speech act, in which an expression is used (with the only meaning it has) to trigger cognitive processes beyond those that are plausibly involved in grasping any thought expressed with the sentence. But just as I wouldn't say that querying is a matter of using a sentence that is true or false to ask a question, I wouldn't say that "metaphoring" is a matter of using a sentence that is true or false to trigger cognitive processes that a hearer/reader might find interesting or aesthetically pleasing. Though singing (89a) is presumably a case of metaphoring that leads a hearer to form a thought that can be recognized as obviously false—given sober reflection of a sort that makes it hard to enjoy a lyric, a horror movie (see Carroll 1990), and many other valuable things.

[24] If 'in' has satisfiers, it can be used to specify them. Theorists might work out what the satisfiers would have to be, invent a language in which 'In(_, _)' has those satisfiers, and hypothesize that 'in' is truth-theoretically equivalent.

But given (C), it's not easy to see how (53) could have a truth condition while (1) does not.

(C) each human language is an I-language in Chomsky's sense

Other considerations regarding context sensitivity, illustrated with (69) and (87),

(69) The kettle is black.

(87) Kim said that James is thin.

make it even more plausible that sentences of a human language do not have truth conditions. Reflection on examples like (88) suggests that action reports do not have truth conditions,

(88) Al chased Theo.

even if the reports concern episodes of chasing as opposed to episodes of saying. Examples like (8) and (71) bolster the case that (53) and (69) are like (1) in not having truth conditions.

(8) Vulcan has an iron core.

(71) France is a hexagonal republic.

3.5 Contexts and Communication

It might seem that simple observations regarding human communication provide the basis for an argument in other direction. Stanley (200x) tries to defend his version of (D) in part by showing how it preserves the idea that "from an utterance of a sentence, one gains information about the world." He offers a familiar kind of example that is instructive.

> If Hannah utters to Esther the English sentence "There is some chocolate in the kitchen", and if Esther wants chocolate, she will go to the kitchen. Even if Esther has never heard that particular sentence before, if she speaks English, she will understand what information it conveys about her physical environment (p. xx).

It's not entirely clear, from this passage, just what Stanley takes the relevant explanandum to be: that sentences convey information; that one can gain information from sentential utterances; or that given a sentence, one can gain information from noting that it was uttered. The details may be important, given the differences between Davidsonian and Kaplanian strategies for accommodating context sensitivity, and corresponding debates about whether proposed truth theories for human languages should have theorems concerning utterances. It may also be important to be clear about the relevant notion of information, which is presumably not quite Shannon's (1948) technical notion. I'm also not sure how Stanley's point is supposed to generalize to the English sentences (69) and (89), in light of Travis' (199x) discussion;

(89) There is some milk in the refrigerator.

see chapter zero. But I can grant that Hannah very likely conveyed some kind of information to Esther by pronouncing an instruction for how to build a thought.

If Esther speaks English and is suitably attentive, she will construct a thought in response to Hannah's pronouncement; and Esther may reasonably believe that Hannah expected Esther to construct such a thought, that Hannah is not being cruel, etc. If 'chocolate' and 'kitchen' are lexical instructions for how to access atomic concepts, then hearing these words will lead speakers of English—i.e., those who have acquired an I-language that counts as an idiolect of English—to access similar concepts. And *if* Hannah knows what she is talking about, as opposed to just making stuff up, then Esther gets some information (as opposed to misinformation) about the location of some chocolate. But I don't see any reason for thinking that the sentence itself is a *source* of this information as opposed to a tool for conveying it in complicated human way.

Again, the real issue is about how *tightly* meaning is connected to truth. Observations regarding human communication don't settle this issue. Nor do they counterbalance the reasons for rejecting (D). Stanley goes on to say that it isn't clear how one can construct a verifiable theory of meaning if one denies that sentences have truth conditions. I think this "only game in

town" argument is overrated. We can use speakers' judgments—concerning implications, ambiguity, and whether or not contextualized *uses* of sentences are correct—as *evidence* for or against specific hypotheses about human I-languages, even if these expression-generating procedures are indirectly related to the truth conditions of truth-evaluable thoughts.

Stanley suggests that certain facts about communication can *only* be explained via theories according to which sentences bear a certain relation to information, in some sense that motivates truth-theoretic semantics. I disagree. If talk of information is standing in for talk of truth conditions, or a technical notion of proposition (cp. Soames 1987), then it is question-begging to assume that sentences bear any natural relation to information. If 'information' is being used in its ordinary way, then Hannah clearly conveyed information to Esther. But I don't see why such facts *can't* explained if meanings are instructions for how to build concepts with which humans can think about things in certain ways, often with the result that speakers and hearers think about the same aspects of a shared environment.

That said, Stanley's appeal to unpronounced indices provides a possible answer to the question of how *anything* could be true given the context sensitivity of truth. This is important. For one might think that even if certain actions are paradigmatic examples of things that are true: (i) *some* sentences have to be true, at least relative to assignments of values to variables, on pain of *nothing* being true; (ii) truth has to be characterized indirectly, via satisfaction of sentences by assignments; but (iii) it's not plausible that mental sentences exhibit the systematicity required to impose satisfaction conditions on sequences. So one might think that sentences of a human language are the best candidates we've got for sentences that have satisfaction conditions, at least if the context sensitivity of truth can be accommodated with many indices, in roughly the way that Stanley suggests. I'm inclined to agree with (i) and (ii); and I feel the pull of (iii). But my suggestion, developed in chapters five and six, will be that there are various kinds of mental sentences: those we share with other animals; those we can form by virtue of having a language faculty that lets us introduce lexical items, perhaps thereby introducing formally new concepts that combine systematically; and those we can form later, having already acquired a human language, when we engage in serious enquiry.

The existence of truths doesn't show that sentences of a human I-langauge are among them. Many sentences in a math/logic/physics/biology/semantics textbook, or this chapter, are hybrids that mix vocabulary from a human language that a child could acquire with invented symbolism whose interpretation is stipulated in a way that only cognoscenti can follow. These hybrids are used in specialized contexts where it is understood that formulating truths—and hence, achieving the kind of precision required to formulate truths of the sort in question—is the goal, even if it is achieved only rarely and partially. Thoughts assembled via these hybrids are, in my view, better candidates for having Tarskian satisfaction conditions than the expressions generable by I-languages that children can acquire. But this is advertisement. Having digressed, in order to connect the details of this chapter with other considerations that bear on the question of how meaning is related to truth, let me return to some remaining details.

4. Back to Troublemakers

In this section, I consider a few attempts to reconcile thesis (D) with Liar Sentences.

(D) for each human language **H**, some theory of truth
is the core of a correct theory of meaning for **H**

4.1 Hierarchy Revisited

Recall Lycan's (2012) suggestion that a sentence can be a sentence of both a "core" language λ that contains no semantic vocabulary and an "expanded" language λ' that contains such

vocabulary, including a truth predicate for **λ**. A more expanded language **λ′′** contains a truth predicate for **λ′**. And so on. On this view, (17) can be true-in-**λ** *and* true-in-**λ′**,

(17) Bessie is not blue.

but (1) is not a sentence of **λ**, and so not true-in-**λ**. Although (1) may be true-in-**λ′**.

(1) Lari is not true.

Initially, this seems like an attractive idea. But if λ and λ′ are human *I-languages*, then it's not clear how it is supposed to work. If λ and λ′ both *generate* expression (17)—a particular articulation/meaning pair—and λ′ generates (1) while λ does not, then it's hard to see what λ can *be* apart from an arbitrarily restricted variant of λ′: the same generative procedure, apart from somehow not having access to a few of the lexical items included in the specification of λ′. So it's hard to see how a truth theory for the *generative procedure* λ′ could, except by arbitrary stipulation, fail to be a truth theory for the *generative procedure* λ. We can, if we like, describe a lexicon **L′** and combinatorial system **C** as the result of adding lexical items to an I-language specified in terms of **C** and a smaller lexicon **L**. We can then introduce two extensionally distinct notions: *true-in-**CL*** and *true-in-**CL′***. But it's hard to see how this bears on the question of whether there is, independent of what semanticists choose to do, a Tarski-style theory of truth for the generative procedure characterized by the generative procedure **CL′**.

Advocates of (C) can agree that 'English' is not a name for any one human language.

(C) each human language is an I-language in Chomsky's sense

Those who speak English have acquired I-languages that are very similar in many respects that matter for communication and various socio-political purposes. And a "monolingual" speaker of English is not limited to a single I-language. We can acquire new lexical items and use different "registers" in different conversations. In these senses, we can and do use distinct I-languages that massively overlap. But I see no reason, apart from fealty to (D) in the face of (C),

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

for thinking that human I-languages exhibit natural "cuts" corresponding to Lycan's hierarchy.

It can be tempting to slide into E-language thinking, characterize languages in terms of disposition to utter, and say that an utterance can be both true-in-English and true-in-German. But even if (59) and (60) sound alike, it doesn't follow any I-Language generates both.

(59) Empedocles leaped.

(60) Empedocles liebt.

And if it turned out that the sound of 'leaped/liebt' had both meanings in two I-Languages, that would be the way to describe the discovery: in this respect, distinct procedures turned to be alike. But our theories of the languages acquired would remain theories of the procedures acquired—not theories that target (alleged) corresponding sets of articulation-meaning pairs and specify the meanings by assigning truth conditions to the relevant articulations relative to a syntax.

Lycan admits that his proposal is *ad hoc* in some respects. But his stated aim is to provide an alternative to Ludwig and Lepore's (200x) suggestion that derivable biconditionals like (1-T)

(1-T) True('Lari is not true.') ≡ ~True(Lari)

can be interpretive (i.e., correct meaning specifications) without being true, and that the axioms of an interpretive truth theory can yield contradictions given true assumptions. On this view, a truth theory for a human language **H** doesn't have to be *true* to be the core of a correct theory of meaning for **H**. It's good enough if the theory generates an interpretive T-theorem for each declarative sentence of **H**, but no uninterpretive theorems; see also Eklund (2002). I share Lycan's suspicion of the idea that false truth theories can serve as correct meaning theories. This

seems like an option of penultimate resort, with embracing contradiction as the last resort. And perhaps Lycan's defense of (D) is, all things considered, the better one. But I also sympathize with Ludwig and Lepore, who recognize that given Liar Sentences, (D) suggests that correct meaning theories will be based on false theories of truth.

In the end, my response to Ludwig and Lepore (and Lycan) will be that we can abandon (D) in favor of the claim that meanings are concept construction instructions. But here, let me note that while Ludwig and Lepore are not explicit about what they take human languages to be, they seem to assume that acquiring a spoken language is a matter of acquiring a capacity to interpret certain utterances. From this Davidsonian perspective, a truth theory for a human language **H** is part of a larger theory of action for a speaker of **H**; where for these purposes, actions include episodes of speech and comprehension. If the theorist's goal in formulating a truth theory is to *ascribe it* to speakers of **H**—as opposed to describing how **H** is related to a certain domain, about which the theorist wants to hypothesize by using **H**—then the formulated truth theory need not be true, much as beliefs ascribed to speakers need not be true. Indeed, it can be appropriate to ascribe an incorrect truth theory for a language to a person who has made a certain kind of mistake. But this suggests that the person is trying to use a language whose character is determined independently of that person. And whatever such a language is, it is presumably *not* an I-language that the person already implements.[25]

If a child thinks that '÷' signifies addition, and infers that '8 ÷ 2 = 10' is true, then it may be appropriate to explain the child's endorsement of this sentence by ascribing to the child a partly false theory of how the arithmetic notation is to be understood. Dummett (197x, 198x) suggested that any speaker could likewise be wrong about certain aspects of her naturally acquired human language. On his view, I might be wrong about what 'livid' means in my own language—and not just wrong about how certain "experts" would use a homophonous word, or wrong about certain conventions governing public discourse. Davidson (1986) rejected this suggestion, along with the underlying idea that a speaker's native language is independent of the speaker's grasp of that language. As an illustration, Davidson said that when Mrs. Malaprop utters 'That is a nice derangement of epitaphs', her sentence (and not just she) means what our sentence 'That is a nice arrangement of epithets' means. On this view, homophonic translation of Mrs. Malaprop would be incorrect. Her utterance is true if and only if the demonstrated thing is a nice arrangement of epithets; and if she pragmatically implies anything else, that is in part because her utterance has this (non-homophonically reported) truth condition.

That debate about the possibility of "linguistic error" (e.g., that 'epitaph' is true of epithets) focused on specific lexical meanings. It is far more radical to abstract from such errors and ascribe a false truth theory to all speakers of English—especially in the absence of a true truth theory and a plausible account of the facts (psychological, social, causal, or whatever) in virtue of which such a theory would be true. Given that human languages include "semantic vocabulary," which speakers use as they do, we may find ourselves committed to contradictory policies concerning how expressions of a human language can be used to make true claims. That is one way of describing the Liar Paradox. But if only because we don't know which of our commitments are false, the contradictory character of these commitments is no justification for doubling down on (D), and concluding that false truth theories can be correct meaning theories.

[25] Though perhaps is it a "Ψ-language" in Ludlow's (2011) externalistic sense. One can stipulate that I-languages are "private languages"—expressions of which do not have "objective" correctness conditions—and hence not "real languages." But then human languages may not be real in the stipulated sense.

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

This leaves us with no clear idea of what the hypothesized false truth theories are theories of.

One can take an instrumentalist attitude towards truth theories, treating them as mere calculation devices that pair object language expressions with "translations" in a certain formalism. But even if one can show that the formal translations are independently attractive as meaning specifications—and this is no small task—that would leave the question of whether the appeal to truth plays any important role in the explanation of why human linguistic expressions mean what they do. Why not instead view the object language expressions as instructions for how to assemble thoughts of the sort depicted with the metalanguage formalism?

4.2 False Theories and I-languages

One might, however, try to combine the spirit of Ludwig and Lepore's suggestion—and at least some of Davidson's reply to Dummett—with thesis (C) by adopting (BF).

(C) each human language is an I-language in Chomsky's sense
(BF) each human language **H** is an I-language such that some *false*
truth theory for **H** is the core of a correct theory of meaning for **H**

One might say that Mrs. Malaprop not only has an I-language with a lexical item 'epitaph' that is true of epithets, she has and endorses a truth theory (for this I-language) that includes a true axiom whose content is that 'epitaph' is true of epithets. On this view, Mrs. Malaprop is right about her word, not merely wrong about ours. But the idea would be that each speaker of a human language **H** can be wrong about the truth conditions of certain sentences of **H**, at least if we each have and endorse a truth theory whose theorems include biconditionals like (1-T).

(1-T) True('Lari is not true.') ≡ ~True(Lari)

This suggestion can be tempting if one views children as interpreters, trying to figure out the truth conditions of alien utterances. One imagines a little field linguist with a concept of truth that gets "overextended" by a simple theory, with the result that the child is apt to draw incorrect conclusions like (1-T), as a side effect of being able to draw correct conclusions like (17-T).

(17-T) True('Bessie is not blue.') ≡ ~Blue(Bessie)

But this is to view the quoted sentence as an expression of the language used by the speakers who are being interpreted, not as an expression generated by the interpreter's own I-language. And our goal is not merely to explain why each us comes to *think*—perhaps wrongly—that the sentences used *by others* have certain truth conditions. Given a certain I-language (or a family of I-languages that are the same in certain respects), our goal is to explain why the expressions of that I-language (family) mean what they do. But if we want to know what the semantic properties of (17) are, and how they are compositionally determined,

(17) Bessie is not blue.

it doesn't help to be told that (17-T) follows from some interpreter's false theory.

One can say that (17-T) follows from a false but interpretive theory of truth for English. But this reminds us that the truth theory allegedly represented by ordinary speakers is governed by a *very* weak logic, making the theory impervious to logical pressure.[26] Once again, this raises

[26] When theorizing about a cognitive competence whose application to cases is subject to logically contingent constraints, we need to distinguish (at least) two notions of implication. One concerns what a theory implies given the rest of science, logic and mathematics very much included. Ideally, one wants to make *all* the implications of a theory manifest, whatever implications turn out to be; see, e.g., Frege (1879, 188x, 189x, 189x). So we want a logic that is more powerful than the propositional calculus. And for purposes of figuring out what follows from a theory, mathematical implications usually count, even though not even arithmetic—much less analysis or geometry—is

the question of whether the appeal to truth plays any important role in the explanation of why human linguistic expressions mean what they do, and why this "error theory" is preferable to viewing human linguistic expressions as concept assembly instructions. Advocates of (BF) need to say *which* false implications of a hypothesized truth theory Θ do not tell against the hypothesis that Θ is the core of meaning theory for a given human language. And again, Liar Sentences are not the only barriers to providing true theories of truth for human languages.

That said, let me conclude this section by noting my appreciation for Lycan, Ludwig and Lepore's recognition that Liar Sentences present a serious challenge for advocates of (D).

(D) for each human language **H**, there is a theory of truth
that is the core of a correct theory of meaning for **H**

In my view, their proposals honestly reflect the difficulties and the substantive ancillary assumptions needed to accommodate sentences like (1).

(1) Linus is not true.

In retrospect, Davidson's (1967a) discussion seems overly brief and optimistic.

> The semantic paradoxes arise when the range of the quantifiers in the object language is too generous in certain ways. But it is not really clear how unfair to Urdu or to Wendish it would be to view the range of their quantifiers as insufficient to yield and explicit definition of 'true-in-Urdo' or 'true-in-Wendish'....In any case, most of the problems of general philosophical interest arise within a fragment of the natural language that may be conceived as containing very little set theory (28-29).

Restricting the range (or domain) of human language quantifiers doesn't help with (1). I don't know how Davidson was counting "problems of general philosophical interest," But one interesting question is whether (D) is true. And in thinking about other reasons for being suspicious of truth-theoretic semantics, we shouldn't assume that examples like (1) present mere technical problems for a proposal that shines with regard to other examples.

---

reducible to "pure logic." By contrast, Foster's Problem suggests a far more parochial notion of implication, corresponding to the *constrained* capacities of ordinary speakers to deploy semantic competence in ways that support the comprehension of boundlessly many novel expressions. As with other natural competences, deployment is constrained by the relevant representational forms (which are biologically implemented), and so unlikely to be closed under whatever forms of deduction a theorist uses to evaluate theories. Indeed, the constraints on an interpretive truth theory are so severe as to raise the question of whether the theory really specifies *truth* conditions for sentences, especially if it is granted that the conjunction of the axioms is false.