Semantics for Semantics*

James R. Shaw

Draft, please don't cite

What drives our investigations of semantic circularity are logical puzzles, stemming from the liar. No wonder, then, that such investigations center predominantly on the question of how semantic circularity impacts our logical theorizing. But this focus tends to defer an important question, that I take as my topic here: what relevance does the liar, and semantic circularity more generally, have for the shape of our compositional semantic theories?

Questions about what a compositional semantics for a self-applicable truth-predicate should look like, as compared to questions about the logic of paradox, have received relatively little attention. The current literature on truth and circularity, for example, reveals nothing like the state of research in epistemic logics, which are accompanied by highly active parallel research programs centering directly on the compositional semantics of epistemic modal language, or knowledge ascription.[1] There is, I think it is fair to say, no comparably active, linguistically driven research program centering on the compositional semantics of truth-talk.[2]

*For helpful comments and advice I'm grateful to Bradley Amour-Garb, Michael Glanzberg, Anil Gupta, Doug Patterson, David Ripley, James Woodbridge, audience members at the *Truth at Work* conference in Paris, and several anonymous reviewers.

[1]For a mere sampling of only recent work on the semantics of epistemic modals, for example, see Anand & Hacquard (2013), Dorr & Hawthorne (2013), von Fintel & Gillies (2007, 2008), Hacquard (2006, 2010), Hacquard & Wellwood (2012), MacFarlane (2011), Moss (forthcoming), Willer (2013), Yalcin (2007, 2010, 2011, 2012).

[2]This is not to say that there is no recent work on truth of relevance to compositional theorizing. There have been discussions of the relationship of truth-conditional semantics to deflationism, for example in Horisk et al. (2000), Collins (2002), Patterson (2005), and Burgess (2011). But such investigations tend to focus on how truth is employed in giving our semantic theories, not in how "true" figures as an object of study within them. There are also some more direct, linguistically driven treatments of the semantics of "true" itself—a recent case being Moltmann (forthcoming). But there, the critical issue of semantic circularity tends to drop out of the picture. A topic closer to my theoretical interests here is that of how to accommodate the threat of paradox within something like a Davidsonian truth-theoretic semantics. But the focus on vicious, as opposed to virtuous, circularity in such investigations tends to lead to pessimistic outcomes: broadly hierarchical treatments that seem to preclude circularity, rather than accommodate it (Lycan (2012)), treatments that take our interpretive truth theories to simply be inconsistent (Lepore & Ludwig (2007)), or treatments that abandon broadly truth-theoretic approaches to meaning altogether (Pietroski, this volume). As will eventually become clear, an aspiration of this paper is to consider both virtuous and vicious circularity together, and to exploit compositional lessons about the former to illuminate the latter. The eventual hope is that this methodology will help us safeguard interesting forms of semantic

There are very good reasons, I believe, why investigations into truth and semantic circularity have been lopsided in this way. Compositional semantic investigations tend, methodologically, to be descriptive in character. And they aim, in part, to account for data that includes relatively stable acceptability judgments, like truth-value judgments. But because circularity can lead layman, like theorist, into confused or contradictory pronouncements, it's not clear that there are stable acceptability judgments to explain or, if so, what they are. And the problems are serious enough to raise worries that the usage of ordinary speakers simply evinces incoherence, so that there is no descriptive theory of a coherent use of "true" to begin with.[3] Moreover, even those investigations that treat semantically circular languages with tools conducive to compositional investigation, like the model-theoretic tools of Tarskian or Kripkean theories, tend to apply those tools in ways that treat "true" compositionally on a par with other predicates by associating it with an extension, or extension/anti-extension pair. If these theories are on the right track, there isn't obviously anything of specifically compositional interest in the truth predicate's behavior.

Despite all this, I believe that there are some very important lessons about truth that we can appreciate only by scrutinizing semantic circularity from the perspective of the compositional semanticist. The goal of this paper is to substantiate that claim.

I'll begin with a broad theoretical explanation of why a compositional semantics for a language admitting semantic circularities is important to have: the viability of such a theory is presupposed by foundational work on meaning in the philosophy of language. And that foundational work, in turn, is what gives significance to any formal investigations into truth, including strictly logical investigation (§1). This opening discussion is meant to help convey the urgency of a specifically compositional challenge that drives the rest of the paper. The challenge is to explain, consistently with linguistic productivity facts, relatively stable truth-value judgments concerning two classes of virtuous semantic circularities (§2–§4). This challenge, I argue, strikingly compels us to abandon compositional theories that give the truth-predicate an extension assignment as part of its semantic value.

This negative contention is the central claim of the paper. But the argument for it will give us some clues about what shape semantic theories must take to accommodate the virtuous circularities in the appropriate way (§5). In particular, the relevant frameworks end up positing an extremely unusual form of sensitivity in the semantics of uttered sentences to the circumstances of their

circularity in a broadly truth-conditional setting. It is also worth flagging that all work on logics of truth can potentially be construed as indirectly constraining our compositional theories. The trouble is that it is uncommon for theorists to frame their investigations in terms of distinctively compositional goals, such as explaining natural language productivity facts, so it is not always easy to tell if theorists mean to commit themselves to such explanatory goals in the course of their logical pursuits.

[3] I have in mind here the tradition of 'inconsistency views' of paradox such as Chihara (1979), Eklund (2002), and Patterson (2007, 2009). Such views can sometimes motivate revisionary, prescriptive methodologies, unlike those of compositional semanticists, that seek to engineer replacement semantic concepts (e.g., as in Scharp (2013)).

tokening, revealing empirical motivations for an unprecedented form of context-sensitivity. Or, at any rate, nearly unprecedented: theories with precisely this highly unusual form of sensitivity have been advanced, not in connection with the compositional problems I consider from virtuous circularity, but as frameworks for understanding the logical behavior of viciously circular liar sentences (§6). Thus, intriguingly, understanding the structure required of any viable compositional theory accommodating semantic circularity indirectly provides us with important independent motivations for a narrow class of solutions to the liar paradox. This class of solutions has historically had, to my knowledge, just a single adherent.

1 Foundational and Compositional Semantics

As noted, I want to begin with some broad reflections on why we need a compositional theory of meaning for a language admitting semantic circularities—subsuming vicious circularities like the liar, and 'virtuous' circularities more like the truth-teller. My purpose in doing so is not merely to convey the importance of providing such a theory, but to highlight methodological constraints relevant to some of my later problem cases (especially in §4).

A tempting reason to maintain that we need a compositional theory admitting circularities is because true circular claims may be required of any language capable of discussing all its own semantic properties at a suitable level of generality. But we need to be careful about why we would expect there to be a language capable of doing this. One could maintain that it is a fact about natural languages that they have this degree of expressive power. But I do not find such claims obvious, nor altogether easy to justify.[4] My own motivations arise not from the language of the layman, but the language of the theorist: any proper study of language which presupposes a theoretical understanding of meaning, presupposes the existence of languages with true semantically reflexive statements.[5] To explain why I think this, I need to develop some interconnections between compositional semantics and more abstract theories in foundational semantics.[6]

[4] A representative expression of the attitude I have in mind here is given by Priest, who maintains that "a natural language... can give its own semantics." (Priest (2006), p.70), but without supplying much by way of justification. I think it is unclear what the empirical or *a priori* grounds for a claim like this are supposed to be (cf. Gupta (1997)).

[5] I take roughly similar considerations underlie some of the remarks of McGee (1991) p.ix that expressive breadth is required for human language to be within the reach of scientific inquiry, though McGee's remarks are mixed with more controversial claims about the expressive richness of English.

[6] I reserve the term *foundational semantics* for all investigations into foundational issues in semantics, including (e.g.) questions about the sources of intentional content (including mental content), and about the nature of semantic properties, like truth. On my preferred usage, foundational semantics subsumes, but is not identical to, metasemantics, which investigates what it is to know, or 'cognize', a language whose primitive expressions bear certain semantic values, governed by certain rules of composition. (See Yalcin (forthcoming) for a sympathetic understanding of metasemantics, and a discussion of why questions in metasemantics should be conceptually separated from questions about content, e.g.). It is worth flagging that the

Philosophers working in foundational semantics are forced to talk about meaning, representation, and truth in extremely general terms. For example, a classic project in foundational semantics is to give the conditions under which any pronouncement or gesture counts as meaningful. Another is to say what meanings are. Projects like these commit the foundational accounts to an extremely high level of generality. For example, we should expect completed foundational accounts to at least make claims about every actual candidate assertion from every actual language.

The generality of these foundational projects inevitably gives rise to an interesting kind of circularity. The accounts themselves are a form of meaningful linguistic communication—the very kind of meaningful communication that the foundational accounts aim to illuminate. So if the accounts are successful, the linguistic tools they advert to should fall within the accounts' explanatory purview. In this relatively straightforward way, the success of work in foundational semantics seems to turn on the coherence of a broad range of reflexive applications of semantic vocabulary like "true", "means", or "represents". To succeed at all, the accounts must successfully speak, in part, about themselves at a general level.[7]

The reflexivity seen here to be a component of standard foundational accounts taken isolation, can actually be pressed more urgently by tracing the circularity through the compositional theories that take the theorist's own language as an object of study.

Compositional semantics involves developing formal accounts that illuminate the way semantic properties of whole sentences depend on the semantic properties of their parts. As such, foundational and compositional semantics sometimes may overlap significantly, for example in the semantics of belief reports.[8] Belief reports (seemingly) relate believers to propositional objects—the kinds of content that foundational semantics constrains and illuminates. A compositional semantics for belief reports inevitably imposes its own constraints on propositional objects. This raises the issue of whether these constraints harmonize with the constraints given by our foundational accounts. As such, the success of a compositional semantics for belief ascriptions ends up being tied to the success of distinct accounts in foundational semantics.

Terms like "true", "means", or "represents", however, generate an even more intricate two-way entanglement between foundational and compositional semantics. This arises as follows:

(A) Foundational semantics must use terms to characterize assertoric content

ensuing discussion is not, or not uniquely, about metasemantics so-construed.

[7]It's worth noting that this reflexivity is unavoidable even for reductive accounts that supply analyses of semantic notions in non-semantic terms. A reductive analysis, no less than any other claim, involves meaningful symbolic manipulation. Accordingly, any proposed analyses must be properly classified using otherwise reduced vocabulary if the analyses are sufficiently general. Naturalistic reductive accounts are in fact especially hostage to this problem, since if reductions lack the relevant generality, they won't in any way allay the naturalistic worries they were designed to address.

[8]See Stalnaker (1987) for a helpful discussion of this interaction.

or effects that make assertions significant. Completed accounts will exhaust properties that are intelligibly assigned in our compositional theories.

(B) Every semantic term from the account in (A) must figure in some systematic compositional semantics if the original foundational account is in fact coherent and has significance.

(A), in a way, merely restates one key explanatory goal of foundational semantics: to characterize the kinds of semantic properties that would be attributed in a formal theory, like a compositional theory.[9] Formal theories, whether merely logical or compositional, have no significance independently of an interpretation of the kinds of semantic properties (like truth, or validity) that the formalism ascribes.[10] And foundational semantics is charged, in part, with the task of telling us what those properties are like. (B), on the other hand, follows from the claim that the theories within which we give our foundational accounts themselves have compositional structure. If so, there is some story to be told as to how the statements of our foundational accounts gain significance, and that story can be given in compositional terms.

(A) and (B) represent, as I say, a kind of entanglement of compositional and foundational semantics, with each playing a role in illuminating or explaining the other. This intertwining requires the sentences of our completed foundational theories to discuss foundationally significant semantic properties, which in turn are attributed in the compositional formalisms that show how those very composite sentences acquire the semantic properties they do. That is where the need for semantic reflexivity, and a compositional theory modeling its behavior, comes from.[11]

Abandoning (A) or (B) is in danger of crippling linguistic theorizing. If (A) fails, properties attributed in some of our compositional semantic theories outstrip those detailed in foundational semantics, and our compositional theory becomes mere uninterpreted formalism, and will cease to do explanatory work.[12]

[9] Strictly speaking, these properties may be assigned by a compositional theory *via* a post-semantics. But it is harmless, I hope, to gloss over this complication here to get a simplified sense of the entanglement I have in mind.

[10] Cf. Dummett (1959).

[11] I haven't yet argued that "true" itself is among the properties treated in this way, but I take for granted that our compositional theory should inform us about the assertions used in our foundational accounts in part by helping us to see how they are "good", or appropriate, or coherent in ways we would like the assertions in our foundational accounts to be. Consequently the theory should afford us linguistic tools which partition assertions in the desired ways. And any such linguistic tools seem like they will be in danger of generating something resembling liar-like phenomena. Focusing on "true", then, can serve us well in assessing how to accommodate (A) and (B) even if truth-talk is not ultimately of appropriate foundational significance.

[12] Importantly, logical theories might not even yield consistency proofs. Without requisite foundational work, the notion of 'consistency' yielded by a formal theory is unexplained. This is significant because recently several theorists have aimed to stave off ostensible commitments of their formal tools by insisting that their theories merely supply consistency proofs. See, for example, Field (2008) p.356 and Beall (2009) pp.56–7. Even 'mere' consistency results require some input, however minimal, from foundational semantics.

If (B) fails, properties mentioned in our foundational semantics outstrip those which can be given a compositional semantics, so that there is no systematic account of the significance of the foundational account itself, as if its sentence-meanings were had by magic. This raises strong worries about the coherence of the foundational account.

Of course, if one works on the paradoxes, one may reasonably be concerned that either (A) or (B) do fail, in spite of all this. I think the paradoxes are so recalcitrant that we should bear this in mind as a drastic option to contend with. But I want to stress that the damage accruing to this option can go unappreciated by those positing expressive and logical restrictions in their formal theories of truth. The foundational programs just discussed won't recognizably survive weakenings of the assumptions I've been making. If the semantic paradoxes show the reflexivity required for (A) and (B) to be incoherent, then we should probably abandon hope of resolving the problem of intentionality, especially in any naturalistically satisfying way. Some intentional, representational, and linguistic phenomena are essentially inexplicable, unlike any other known object of study. Moreover, whatever formalisms we give in supplying our theories of truth will be essentially uninterpretable, or open ended. We'll lack any justifiable theory drawing explicit ties between the formalism and the mental and linguistic phenomena that the formalism is clearly supposed to somehow illuminate. This threatens the whole point of giving a formalism in the first place.

So: better to proceed, provisionally, on the assumption that such a compositional semantics admitting circularity can be found. Let's turn now to the question of what it should look like.

2 Compositional Circularities

In explaining why we need compositional theories admitting semantic circularities, I leaned only on the weak claim that compositional theories help explain how the meanings of whole expressions depend on the meanings of their parts. But to draw out my puzzles, I will need to make some pronouncements on what kind of data and methodological assumptions specially constrain our compositional theories. In this section, I'll say what those constraints are, and why semantic circularity may interact with them in special ways. The assumptions I make will not be completely uncontroversial, but should serve well enough as a basis for the ensuing discussion.[13]

A compositional semantic theory is a component of a broader theory about our capacity to speak and understand a language.[14] This understanding is construed as a kind of cognitive state, aspects of which are modeled by the formal apparatus of the theory.[15] The key components of the apparatus are an as-

[13] A helpful recent, and more detailed presentation of parts of the perspective on compositional theorizing I'm adopting can be found in Yalcin (forthcoming).

[14] Working, e.g., very broadly in the vein of Chomsky (1965, 1986).

[15] Familiarly, we will need to think of this cognitive state as different in kind from ordinary

signment of semantic values to the minimal interpretable constituents of the language, and rules for combining those semantic values in accordance with the syntactic rules governing the combination of the constituents. The structure of the theory is constrained by empirical facts that are taken to be the product of the hypothesized cognitive state of language-understanding, at least generally and in part. The facts include speaker judgments of interpretability or markedness, judgments about entailment, and facts about speaker communication. Most importantly, for my purposes, they also include productivity data, and speaker judgments of truth or falsity.

'Productivity data' comprise speakers' ability to understand a wide range of sentences they have never encountered before, on the basis of limited training. The compositional theory models aspects of the cognitive state of language understanding that engender productivity by grounding the understanding of the semantic value of whole sentences in an understanding of a finite stock of semantic values, and syntactic and semantic composition rules.

'Truth-value judgements' comprise information about how speakers pronounce, often uniformly, on the truth or falsity of a sentence (or utterance of it) relative to actual and counterfactual scenarios. The theory models aspects of the state of language understanding that engender truth-value judgments by associating whole sentences with semantic values that determine the truth-value of an utterance of that sentence, in part as a function of the context in which that utterance is produced.[16]

As I noted, the semantic theory models aspects of the cognitive state of language understanding that *generally* and *in part* explains these kinds of facts. 'Generally', because (e.g.) speakers can come to understand a sentence, or recognize its truth-value, by ostensibly non-compositional means (perhaps having been told its meaning or truth-value). Or speakers may make mistakes by (e.g.) confusing implicated information with information literally asserted. 'In part', because (e.g.) speakers may arrive at the actual truth-value judgment of a sentence in part owing to information they have from non-linguistic capacities, like perception.

The foregoing exhausts the understanding of compositional theorizing that will be necessary to appreciate the cases to come. But before turning to them, it will be helpful to locate why semantic circularity might present special challenges to developing a theory of the sort I've just described.

Let's begin by considering a simple case where a compositional theory could be called on to do some basic explanatory work. Suppose we show a group of

propositional attitudes. I will sometimes speak of what speakers 'know' (e.g.) in understanding a language for simplicity, but I do not mean to presuppose the cognitive state in question has anything in common with ordinary knowledge as studied, e.g., by the epistemologist.

[16]For a discussion of the importance, and commonality, of truth-value judgments and productivity data as constraints on a compositional theory, again, see Yalcin (forthcoming). I follow Yalcin in speaking of the theory as being constrained by truth-value judgments, as opposed to facts about truth-conditions (cf. Lewis (1970), Heim & Kratzer (1998)), so as not to beg questions against non-factualist or non-truth-conditional accounts of certain branches of discourse. However, if we take discourse about truth to have truth-conditions, as I will here, these subtleties won't matter much for the ensuing discussion.

consultants the following array

and ask whether (1) is true.

(1) The triangle is white or the square is grey.

Sighted English-speaking consultants report (1) is true, even if they never encountered this sentence before. Others are not as reliable. What explains these patterns?

A natural explanation appeals to two theories. First, a theory of non-linguistic cognition: the sighted consultants have the perceptual and cognitive capacities to tell the shapes and colors of objects in their near environment. Second, a theory of linguistic competence for English—a competence appropriately modeled by a theory associating semantic values with the words "grey" and "square", and comprising rules for composing those meanings in line with English syntax.

The two theories obviously must interact, in a fairly straightforward way, to explain the regularity in speaker judgements. Here's a very rough story of how the interaction goes: Our theory of linguistic competence for English models that competence compositionally, by recursively associating semantic values with whole English sentences on the basis of the semantic values of their parts. The semantic values for whole sentences, coupled with information from context, determine conditions under which (what is expressed by) an utterance of that sentence is true.[17] The conditions themselves are generally non-linguistic, as with (1): our semantics for English should derive, roughly, that (1) is true just in case there is a single triangle in the array that is white or a single square that is grey.

Now, our aforementioned theory of non-linguistic cognition should be able to tell us when and how a consultant can distinguish whether conditions like this do or do not obtain. The ability to distinguish colors and shapes is not a compositionally or linguistically mediated competence—at least not one involving the compositional competences of English. So our theory of non-linguistic cognition should explain how our sighted consultants distinguish that there is a single triangle that is white or a single square that is grey.

The reliability in speaker judgments then gets explained by the two theories working in tandem: the verdict that the English-speaking, sighted consultants

[17] I will assume, for convenience, that we can treat utterances, at least derivatively, as truth-bearers. But even if "true" only properly applies to propositions, for example, the same worries I'm about to raise in the next two sections will arise if we consider the predicate "expresses a true proposition".

give is a simple entailment from two pieces of information that the linguistic and broader cognitive theories respectively attribute to them. The linguistic theory models tacit, compositionally mediated information that speakers have of the conditions under which (1) is true.[18] The cognitive theory explains how speakers gain the information that the conditions obtain. The latter information, coupled with the former, entails that (1) is true.

The aspect of this explanation which interests us is the clean separability of the linguistic compositional theory and the non-linguistic cognitive theory. As soon as a compositional theory of English recursively assigns (1) a semantic value determining appropriate truth-conditions, its work in explaining the regularities in truth-value judgments is done. Any further explanatory burdens are immediately 'passed on' to the theory of non-linguistic cognition. The theories are cleanly separable in this way owing to our earlier plausible assumption that distinguishing between the conditions under which (1) is, and is not, true involves only non-linguistic cognitive capacities.

This assumption holds for the majority of used natural language sentences. But it will not hold for many sentences containing semantic terms like "true". The capacity to distinguish situations in which there are true or untrue English utterances is often a linguistic, compositionally mediated ability. Accordingly, in many ordinary situations, explanations of truth-value judgments for sentences containing "true" will not always hand off explanatory burdens to a separate cognitive theory so quickly. Instead, a speaker's knowledge of whether novel, structured utterances are among the truths is precisely what is supposed to be explained by the compositional theory itself. So sometimes an explanation of regularities in speaker judgments using the compositional semantics will be forced to advert to the compositional machinery a second time.

Suppose, for example, we add another sentence to be assessed.

(2) The sentence labelled "(1)" on your questionnaire is true.

We find the same regularity for (2) as for (1). Why? Again, our compositional theory should determine the obvious truth-conditions for (2): that the sentence labelled "(1)" on the questionnaire express a truth. But here, unlike with our sentence about shapes or colors, the explanatory work of the compositional theory in explaining the judgements about (2) is not quite done. We need to explain the speakers' prior acquaintance with (1), and how speakers compositionally settled *its* truth, to fully explain how they have the information which, paired with the information modeled by the cognitive theory, entails (2)'s truth. The compositional theory, not merely the separate cognitive theory, is partly

[18] Again, the information modeled by the semantic theory need not be 'known' by the speaker as part of their propositional attitude psychology. It may not even, properly speaking, *be* 'information' that is modeled—perhaps (to take one example) the theory models something more like a network of causal processes. And we need make no commitments here about the level of abstraction at which the theory models the target cognitive faculties. Still, the idea is that some aspect of the complexity captured by the representation of semantic values in the theory is 'retrievable' or 'exploitable,' in *some* sense, in the story about the origins of the speaker assessment. That is all I will be assuming in what follows.

responsible for explaining how our speakers are aware of (1)'s truth, so as to settle (2)'s truth. If, to take a strange case, our consultants knew English up the meanings of "grey" and "white", they would obviously not be able to settle (2)'s truth in the way they typically do. The fact that the compositional theory re-engages with (1) in the explanation of the judgments about (2) is what helps explain why and how this bit of linguistic competence with "grey" and "white" figures in the consultants' judgments about (2), which contains neither of those words.[19]

The simple case involving (2) reveals a basic pattern in using compositional theories to explain regularities in truth-value judgments about sentences containing semantic terms. To explain the regularities, the compositional theory may be set up to require speakers to have information about truths. But to explain how speakers have that information, we may need to appeal to the compositional theory again. For any set of attributions, these round-about explanatory moves may iterate many times before bottoming out in our non-linguistic cognitive theories, which in turn ground the explanation in a satisfactory way.

But note that the structure of these explanations reveals a possibility, and hidden danger, unique to our compositional theories of semantic terms. The danger is that we structure our theory in such a way that explanations fail to bottom out. Our theory may be structured so that using it to explain various truth-value assessments unendingly redirects us to the compositional theory, thus affording no explanation of the assessments at all. This outcome is obviously unacceptable provided it arises for an empirically confirmed regularity in compositionally mediated, correct truth-value assessments. I'll call such a situation one in which a (vicious) *compositional circularity* arises in our theory.

Though it may be hard to appreciate now, in the abstract, it is worth stressing that the question of whether a theory exhibits compositional circularities is not settled by the theory's being compositional. In particular, a theory may ensure that semantic values of parts and wholes satisfy a compositional principle of any desired strength, while that theory nonetheless fails to avoid compositional circularities in the sense I've just described. If so, avoidance of compositional circularities would place an additional desideratum on our compositional theories of semantic vocabulary. As I say, this may not be easy to see merely from the definitions given here. But rather than continue in abstraction, the best way to appreciate the separability of the constraints is to examine some actual theories which generate compositional circularities, while seemingly satisfying compositional constraints on relations between parts and wholes. Let's turn to this task now.

[19] As always, there are many ways that consultants could come to the information that (2) is true. The example here is only claiming that *one* of those ways, albeit a natural one given the set-up, will re-engage the speakers' linguistic competence for (1).

3 Defaulting

In §1, I argued that we have a very strong need for a compositional semantics for languages that admit semantic circularities. What should such theories look like?

One of the most successful semantic programs that fits into the paradigm given by §2 is the broadly model-theoretic tradition informed by generative grammar. Working within this program, a tempting way to model the compositional effects of "true" would be to assign it a familiar semantic value borne by any common predicate: one which at a context-index pair (say) fixes an extension consisting of the entities which bear the property it is used to record.[20] I'll call this view "Truth-Extensionalism".[21]

> **Truth-Extensionalism**: The compositional contribution of "true" is an extension (relative to a context/index pair, perhaps along with an anti-extension, etc.).

I'll presently argue that Truth-Extensionalism must be rejected because it generates two kinds of compositional circularities. In this section, I'll present the more marginal and controversial of these two cases. These marginal cases have the virtue of being simpler, giving us a clearer view of at least what a compositional circularity would be, and why it would be important. In the next section, I'll present more central cases, though these will be more complex.[22]

Suppose Sid has promised Marie that he will speak only the truth today. A friend asks about his uncharacteristic honesty, and he explains.

(3) I made a promise which I'll have kept just in case everything I say today is true.

Then he continues, with confidence.

(4) Everything I say today will be true.

Suppose Sid's other pronouncements throughout the day are uncontroversial truths (about what he ate, etc.). In this scenario, Sid's utterances of (3) and

[20] I don't mean "property" in any inflationary sense. I only presume that "true" partitions utterances in accordance with the aims of our foundational accounts in semantics.

[21] As mentioned in n.17, I'll work on the simplifying assumption that truth applies to utterances. The arguments to follow could easily be reinstated for "expresses a true proposition".

[22] Since I'm arguing against Truth-Extensionalism, one may wonder who holds such a view. The answer is that it is not obvious that anyone does. But this is not because the position is clearly rejected *in favor* of some specific alternatives. Rather, as I noted at the outset, the question of what a compositional theory for "true" should look like, and how it is integrated with our theories of linguistic productivity for other terms, is most often ignored, or deferred. Still, Truth-Extensionalism is a natural starting point since, as just mentioned, the model-theoretic tradition presents us with one of our more successful frameworks for compositional investigation. Moreover, there is at least some hope, given the pervasive use of model-theoretic tools in investigations of the logic of truth (notably in Tarskian and Kripkean frameworks, and those that build on them), that the use of those tools could substantially inform the final shape of our compositional theory.

(4) seem to be simple truths. And speakers treat them as such. In informal consultations, for example, ordinary speakers only had these judgments and expressed some confidence in them (sometimes perplexity at what other status (3) and (4) could have). Consultants also seem to give these verdicts for variants where, for example, promise fulfillment wasn't at issue.

These judgments could be important, because (3) and (4) appear to attribute semantic properties to themselves—(3) conditionally, and (4) directly. Let's focus on (4), which I'll call a *defaulter* because it appears to 'default' to truth despite a significant truth-teller-like semantic circularity. Of course, until we consider the matter more closely, this is just a *prima facie* appearance. In particular, the judgments about (4) raise three related questions: First, is (4) in fact attributing semantic properties to itself? Second, what is that status of speakers' judgements about (4)? Third, in light of our answers to the first two questions, how are we to explain the judgments?

To answer the first question, note that both (3) and (4) must attribute semantic properties to themselves if Sid is to properly speak to the conditions of his promise—the case is set up to 'pressure' a broad reading of the quantifier "everything". Several consulted speakers were able to confirm that their judgments involved interpreting the quantifier with this range. These judgments, if indeed pervasive, are significant, despite the fact that quantifiers are known to exhibit a slippery form of contextual domain restriction. If someone says "John has shaken hands with everyone in the room" she may naturally mean that John has shaken hands with everyone in the room but John himself. But, in simple cases like this (that don't involve quantification over large numbers of objects, or groups with vague boundaries), how broadly a quantifier ranges, either as meant, or as interpreted, tends to be transparent. A speaker uttering the aforementioned sentence, or a hearer interpreting it, can often say whether she was taking John to be included among those whose hands were shaken by John (usually: not), and that judgment, when firmly given, seems to be a reliable indicator of at least how the speaker or hearer was interpreting the quantifier. As such, nothing about ordinary cases of domain restriction gives us reason to discount speakers' judgments about how broadly they are interpreting the quantifier to range here in the case of (4).

Speakers seem able to naturally interpret (4) as a truth about everything Sid said, (4) included. But this raises our second question. We know speakers' truth-value judgments aren't infallible. Most importantly, speakers could be confusing the proposition expressed by Sid's (4) with information that is implicated or otherwise pragmatically conveyed by Sid's utterance of it. Might the broad interpretation of (4) be false or defective, with speakers confusing its truth with the truth of separate content that is implicated or otherwise conveyed?

The only motivation to treat the judgments in this case as based on implicatures seems to be to avoid assigning relevant, true, *reflexive* content to (4). After all, if there were such relevant, true, reflexive content, what possible grounds would there be not to treat (4) as expressing it? If this is the motivation for an appeal to pragmatics, though, such appeals are essentially self-defeating, because they are themselves helpless to avoid positing such con-

tent. Sid could have equally well promised that what he said *or* implied (or conveyed, etc.) would be true. If he reports his success on Monday, it continues to seem true, quantifiers seemingly interpreted broadly just as with (4). What is the explanation of *this* judgment? We face a dilemma. We could grant that the new utterance literally and truthfully expresses a bit of reflexive content. But then there is no reason not to attribute similar content to (4). Alternatively, we could try to say that the new utterance merely implicates a truth about all things said and implicated. But then we will have a bit of content implicated that is both reflexive (an implicature in part about the implicature) and true. There is no 'hyper-pragmatic' content to avoid the reflexivity in the implicated content. And so the motivations for appealing to pragmatics in the first place are undermined—it needlessly shifted the bump in the rug.

Speakers don't seem to be confusing the truth-value of (4) with that of implicated content. But this doesn't yet tell us speakers are making correct judgments. There is still room for a broad error theory. Perhaps speakers are systematically mistaken about what is said or implicated in defaulting cases. On this view, for example, speakers are simply in error if they take Sid to communicate something that, on its own, entails that his promise will be fulfilled.

This maneuver is quite radical. After all, we would be claiming that speakers systematically make mistakes about what they all stably think that they are communicating to each other (whether by saying, or implicating, and so on). How in this case could a speaker think they are conveying something with certain truth-conditions (remember: not necessarily *literally* conveying them, merely conveying them *somehow*), a hearer think that those truth-conditions are conveyed, and both be in error? One might be concerned that it is close to constitutive of communicating certain truth-conditions, in a broad sense, that it is common belief between communicator and interpreter that precisely those truth-conditions are being conveyed.

The radical nature of this position means that it calls for strong justification. There is perhaps one set of considerations that could play this justificatory role: reflections on reflexivity in relation to paradox. The thought goes roughly as follows: Speakers may think they are conveying certain truth-conditions. But we know from discussions of reflexivity and paradox that sometimes reflexive statements can't obviously have truth-conditions at a given world (in the case of the liar) owing to logical problems, or they have truth-conditions that are awkward to assign (in the case of the truth-teller). Perhaps similarities between the problem cases and (4) should lead us to assimilate the latter to the former, in spite of how speakers seem to use sentences like (4) in communication.

Obviously, though, (4) is importantly dissimilar from liar sentences. There is no logical obstacle to (4) being true. On the other hand, (4) does look much like a truth-teller, such as (T).

(T) The first line labelled "(T)" in this paper is true.

The problem is that there is a clear and striking asymmetry in how speakers react to (4) and (T).[23] This asymmetry needs to be accounted for somehow.

[23]Note: the asymmetries persist even for 'contingent truth-tellers' (that ascribe truth to a

Granted, we have existing theories that treat (T) as untrue (often 'defective'). And these theories will likely treat (4) as untrue as well. But overturning the judgments about (4) with an error theory for these reasons is putting the cart before the horse. If we are creating a compositional theory, it should be responsive to the data extrapolated from speaker judgments, not overturning it. We should of course allow irregularities to be smoothed over by appeals to other plausible explanations of the data—most notably pragmatic ones. But when such explanations aren't available, we can't simply pretend the data isn't there in order to hold on to our current favored theories, even if they are simpler. The data must be accounted for somehow, so the error theory, even bolstered by reflections on paradox, must be accompanied by some further explanation of the regularities in speaker judgments. I'm not sure what this further explanation looks like, given that pragmatic avenues seem to be ineffectual.

There is another, I think slightly better, way of pressing this kind of worry.[24] The position grants that (4) is typically treated as a truth by ordinary speakers. But it goes on to contend that the speaker judgments are inherently unstable. Perhaps, for example, speakers haven't properly reflected on the similarities between cases like (4) and those like (T), and if they did, they would be willing to retract their judgments, or at least become confused.

I think this is a more plausible route to pursue with the theoretical reflections on paradox—not to simply overturn speaker judgments, but question their stability. A problem is that the methodology here is a little murky. The paradoxes are extremely complicated. Even careful, highly trained theorists can make mistakes about how they operate. And we do not yet have anywhere near an uncontroversial theory of how paradox is to be treated. Consultants may be unduly influenced by the complexities of the paradoxes, or prejudices of theorists, in forming 'sensitized' judgments (whatever the proper methodology for producing them would be). For example, I think it would not be entirely surprising if discussions of paradox could (without proper coaching) problematize speakers' intuitions about clear truths like those used in the famous Dean/Nixon cases of Kripke (1975), discussed below. If they did, I think it would be clear we should ignore those judgments. More generally, if we get confused judgments *after* complex theoretical tutoring, there is an open question about which of the two conflicting sets of judgments should be empirically fundamental. Perhaps more importantly, even if the 'sensitized' judgments are the more fundamental ones, the *original* judgments (in particular the original asymmetries between (4) and (T)) have still been left largely unexplained. We only have the claim that speakers get confused by circularity—but why so much so with (4), and not at all with (T), even when in the former case they are broached of all non-semantic facts, and consulted about the range of quantification? So even this more nuanced form of error theory still needs to be supplemented with some additional explanatory considerations.

Still, I see that the contention that we should be wary of sensitized judg-

single utterance), provided speakers are given full awareness of the non-semantic facts, as I have done for (4).

[24] I'm grateful to Doug Patterson for helping me see the stronger version of the worry.

ments may be controversial. And though I do not know what kind of theory can be paired with the error theory to systematically predict the regularities in speaker judgments, I do not want to presuppose that no such theory can be devised. We've bottomed out in a mix of empirical and theoretical considerations I can't hope to resolve here. Thankfully, there's no strong need to. As I flagged earlier, the cases I'll consider in the next section are much more theoretically costly to ignore in the way proposed here, though they effectively require a 'benign' reflexivity just like (4) seems to exhibit. This will actually break the alleged symmetry between (4) and (T) on which the error theory relies. Still, we shouldn't turn to those cases immediately. Our discussion so far has presented at least a noteworthy *prima facie* case for (4)'s truth. It will be helpful to first discuss why this simple feature of (4) would matter to our compositional theories, before turning to more complex examples.

Now, working provisionally on the assumption that speakers are correctly assessing (4) as true, I take it to be uncontroversial that those assessments are compositionally mediated. Speakers who judge (4) true haven't, for example, colluded with some third party who has told them its truth-value. Awareness of the truth of (4), if speakers have it, depends at least in part on appreciating (4)'s structure, and the meanings of its parts.

Granting these claims, let's first note a minor consequence. If (4) is true, an understanding of "true" in English is not exhausted by a grasp of the Tarski biconditionals. This is because one can never use a Tarski biconditional for (4) to establish (4)'s truth. This biconditional, along with basic empirical premises, can establish that (4) is true if several utterances, including that of (4), are true. But this mild circularity ensures (4)'s truth cannot be established using only uncontroversial empirical premises and truth biconditionals. This may have implications for those deflationist positions which take something like a grasp of the Tarski biconditionals to exhaust an understanding of "true". Such theories would require added rules to capture defaulting conditions.

I mention this logical fact because it is connected with the separate claim that I want to make my focus. If (4) is true, truth-extensionalist views cannot explain the speaker judgments about (4) because those theories generate compositional circularities.

Recall the discussion of §2: compositionally mediated speaker judgments are to be explained as the product of linguistic competence, modeled by our semantic theory, in conjunction with the non-linguistic information speakers possess, as explained by our non-linguistic cognitive theories. But if we adopt a standard model theoretic semantics, for example, and "true" is merely assigned its extension at a context, then our theory will encapsulate the information that (4) is true just in case a certain set of utterances is in that extension—i.e., among the truths. As we back out into our theory of cognition, we can say how speakers settle more details about what that set of utterances is, and gradually explain how their truth-values could have been established (and so, effectively whether they are or are not in the relevant extension[25]). Some will be explained like

[25]This kind of talk of 'knowing' or 'settling' whether some utterance is in the extension

(1), by direct appeal to assessors' cognitive faculties. Others will be explained like (2), by round-about interactions between our cognitive and compositional theories. But after all this, we still won't be able to establish how speakers have settled (4)'s truth in Sid's case. The recursive characterization of (4)'s truth will ultimately hinge on its own status. Put another way: the information encapsulated in the theory modeling speakers' linguistic competence, coupled with the information attributed to them by our non-linguistic cognitive theory, simply does not entail that (4) is true—the judgment that we find speakers systematically making. But the linguistic and non-linguistic information is *all* the information the speakers could have.

There are two avenues to pursue in completing an explanation of how speakers are aware of (4)'s truth, consistently with Truth-Extensionalism. But neither seems satisfactory. First we can try again to complete the explanation by appealing to our cognitive theory. Though (4)'s truth hinges compositionally in a model-theoretic framework on its own truth, perhaps that is unproblematic because speakers have an awareness of (4)'s truth that is explained by a separate non-linguistic theory of cognition. But this is highly implausible. It requires speakers' knowledge of (4)'s truth to be non-linguistic and non-compositional, as if speakers had a separate cognitive faculty to 'see' its truth independently of what its words meant. This implausibly makes any work used to settle (4)'s value—including that of finding out the truth of the other things Sid said—superfluous.

The only alternative is to resort back to our compositional theory. The problem is that one is then trapped in an immediate, and vicious, circle. The theory pushes the explanation back to a point where we face exactly the same question: how do speakers have an awareness of (4)'s truth to begin with?

Merely assigning "true" an extension as part of its semantic value thus makes any account of (4)'s truth circular in a special way: it obscures how speakers could, with their linguistic and cognitive faculties, coordinate on (4)'s truth when they do. This is why, if speakers' judgments indeed reflect (4)'s actual semantic status, we must reject Truth-Extensionalism.

To clarify what I mean in saying that (4) is circular in a 'special way', let me contrast two cases of semantic circularity which are not *compositional* circularities, and so don't raise problems for Truth-Extensionalism.

Consider first interdependences like (J) and (N), familiar from Kripke (1975).

(J) Most of what Nixon says about Watergate is true.

(N) Everything Jones says about Watergate is false.

Kripke noted that while utterances of (J) and (N) by Jones and Nixon respectively can be paradoxical, under favorable circumstances they are not. If Nixon

of "true" shouldn't be taken too seriously, of course. It is merely meant to help give an understanding of why the information encapsulated in the compositional theory, coupled with the information speakers have by perception and other non-linguistic means, do not entail (4)'s truth.

utters only clear falsehoods besides (N), and Jones only utters (J), then Jones' utterance is false and Nixon's is true.

One might think that such utterances resemble defaulters. The interdependent semantic ascriptions make (N) in some sense, in part, 'about itself'. But none of this casts doubt on Truth-Extensionalism. Every normal case where (J) and (N) are truth-evaluable is one where we can settle, with the information contained in a suitable compositional theory, enough truth values to settle the semantic status of one utterance before the other.

For example, our compositional theory will explain that speakers can settle (J)'s truth if they can settle that the majority of Nixon's utterances are untrue. If Nixon has only uttered falsehoods besides (N), our cognitive and linguistic theories are in a position to explain how speakers get that information in a non-circular way, before using it to settle (J)'s truth, and then using *that* information to settle (N)'s truth. There is no threat of a circular explanation here. There is no grounds to think that the information modeled with an extension assignment to "true", paired with the information speakers have by other means, will not entail the correct truth-value allotments to (J) and (N) (in the relevant circumstances).

So (J) and (N), though semantically interdependent, do not generate compositional circularities for Truth-Extensionalist theories the way defaulters do. Utterances that rely more directly on their own truth come closer to generating the problem. Simple truth-tellers like (T), discussed above, at least have the right kind of circular structure. But (T) can't be used to argue against Truth-Extensionalism, for two reasons. First, it's unclear that there is speaker concurrence on (T)'s status. In fact, it seems there is *no* concurrence for (T) on a 'standard' semantic status. Second, and relatedly, (T) is commonly taken to be semantically defective. But generating an unavoidable explanatory circularity seems like a perfectly legitimate way to explain the presence of semantic defect. It's not clear why the status of defective utterances would have to be determined by non-defective compositional means.[26]

So the problem generated by (4) requires a special confluence of features, borne neither by Kripke-style interdependences, nor by self-dependent truth-tellers. Note that the problem is not that Truth-Extensionalist theories fail to be compositional. A model-theoretic semantics, like that associated with a Kripkean fixed-point construction, may show how the semantic value for (4) harmoniously depends on (is a function of) the semantic values belonging to (4)'s expressions, in essentially the standard way (up to any special accommodations needed for anti-extensions). The problem is that such a theory does not afford

[26](T) of course raises the important question of why utterances of (4) sometimes default whereas those of (T) do not. My arguments rely only on the existence of some defaulters, so I won't give a detailed answer here. The following rough explanation should suffice: Defaulting seems to be a convention that we adopt to avoid certain kinds of expressive limitations when we try to express a pattern in distributions of semantic properties. One case doesn't constitute a pattern, so the convention never engages for sentences like (T). Since the line between mere scattered instances and a pattern is amenable to debate, so will questions about where defaulting is natural or not. But the extreme cases, like (T) and relevant uses of (4), will be clear.

an explanation of how speakers exploit these compositional dependencies to establish the semantic facts that they can. The information encapsulated by the theory isn't sufficient, given the other information speakers have, to determine the relevant semantic facts.

Recall that these problems for the truth-extensionalist can be framed as a dilemma in explaining judgments concerning (4), once the truth of other utterances (4) speaks about is established. Either she can try to derive (4)'s truth within the compositional semantics, representing speakers' linguistic competence, which is hopelessly circular. Or she can appeal to some bit of non-linguistic cognitive awareness to settle (4)'s truth, which I've claimed to be implausible. Before concluding our discussion of defaulters, I want to address one last attempt to avoid trouble by gripping the second, cognitive horn of this dilemma.

I've claimed non-linguistic cognition can't get us far enough to settle (4)'s truth. Perhaps it could (coupled with information possessed in virtue of linguistic competence) explain how we arrive at the truth of the other utterances (4) speaks about. But we don't know (4) is true in the same way that we know, for example, that the sky is blue—we don't simply 'perceive' (4)'s truth without any compositional work.

But the analogy with perception here, though I think forceful, only points out that speakers' *a posteriori* non-linguistic knowledge is unhelpful. But what if (4)'s truth is known with the help of non-linguistic *a priori*, conceptual knowledge? Consider: Most English speakers may correctly judge "green resembles blue more than red" a true sentence of English. Plausibly, this regularity isn't explained by their linguistic competence, combined only with knowledge gained by *a posteriori* cognition. Instead, it is the product of their linguistic competence and *a priori* cognition of conceptual truths about colors and their relative similarities. The objection I want to consider asks whether there could likewise be *a priori*, conceptual truths about the concept of *truth* that help speakers settle the value of (4). For example, perhaps it is a just an *a priori*, conceptual fact about truth that, when a sentence like (4) is uttered and all other utterances (4) speaks about are true, then (4) itself expresses a truth. The idea is that linguistic, compositional resources and *a posteriori* forms of cognition are generally needed only to get us information about the truth of the other utterances (4) speaks of. But the final step of settling (4)'s truth owes neither to *a posteriori* cognition, nor to the compositionally mediated competence with the language in which (4) is couched. Rather, it involves an appeal to conceptual facts about truth that speak to the semantic properties of defaulters under special circumstances.

This view at least improves on our other suggested appeals to cognitive faculties in that it leaves room for compositional work to play some role in settling (4)'s truth, before hypothesized conceptual truths come into play. However, there are two worries for the proposal. The first is simply that introducing irregular conceptual truths about the linguistic properties of special uses of "true" seems *ad hoc*. But I want to press a second, more instructive problem with the strategy: it is clearly a conventional matter whether, and when, defaulting oc-

curs. There is nothing incoherent about a set of compositional conventions for "true" on which (4) comes out false, or defective. Indeed, truth-tellers like (T) show us that such compositional conventions (or lack thereof) are likely already in force for other cases resembling (4).

Why does this matter? The objection we're considering leans on the following alleged conceptual truth about truth: "As a matter of conceptual necessity: if the rest of (4)'s semantics and the relevant empirical facts are held fixed, and "true" in (4) expresses the concept of truth, then (4) *must* be true". But what I've just claimed is that this conditional is in fact false, as is witnessed by the possibility of a community speaking a language just like ours, but in which defaulting conventions for "true" are different, or entirely absent. In such a language, (4) could have an identical semantics up to the use of the word "true". And all should agree that the utterance of (4) in relevantly similar circumstances is no longer genuinely true in such a language—whether because it is false, or defective. After all, compositional conventions for a word's use can be stipulated however we like, at least provided they are coherent.

If this is right, we can change (4)'s truth-value while holding fixed the empirical facts (other than those concerning the conventions governing "true") and the semantics of words other than "true". That is, we've satisfied two of the three conditions in the antecedent of the allegedly true conditional and shown that its consequent does not follow. If so, the only recourse for the objector is to claim that the third and final condition in the antecedent of the conditional is false. That is, she must claim that in this hypothesized language without our forms of defaulting—call it "$\mathcal{L}$"—"true" no longer expresses a concept of truth. The problem is that there are no grounds to maintain this, given that the use of "true" in $\mathcal{L}$ and our use of "true" can *agree* in all applications, including to utterances of (4) by $\mathcal{L}$-speakers. (4)-*as-used-in*-$\mathcal{L}$ falls neither under the concept $\mathcal{L}$-speakers associate with "true" in their language, nor under the concept we associate with "true" in ours. $\mathcal{L}$-speakers' use of "true" fails to apply to (4)-as-used-in-$\mathcal{L}$ because defaulting conventions for their use of "true" in (4) are different, or non-existent. This much has already been conceded: one can stipulate one's conventions for how a word is used however one likes. For us, by contrast, "true" fails to apply to (4)-as-used-in-$\mathcal{L}$ because we acknowledge that when these defaulting conventions in $\mathcal{L}$ are different, or non-existent, the result is precisely that (4) has a semantics in $\mathcal{L}$ on which is ceases to be true in the relevant circumstances—that is, ceases to be true according to *our* use of "true". Ultimately, then, their concept agrees with ours in application to their utterances of (4), despite the different compositional conventions in each language. The lack of defaulting conventions in their language makes their uses of (4) untrue (so "true" in their uses of (4) fails to self-apply). But our recognition of this fact *requires* our use of "true" also not to apply to their utterances of (4). That's precisely what is required by our recognition that their language adopts different conventions for the behavior of (4) (as used in their language).

So, if there were different, perhaps less permissive defaulting conventions in a language, the use of "true" in such a language needn't conflict in its applications with our applications of "true" to utterances of (4)-in-that-language, or to any

other utterances. If a linguistic community ceases to allow utterances of (4) to default, and hence for "true" to apply to such utterances 'from within' their uses of (4), we (must) correspondingly cease to apply our "true" to those selfsame utterances 'from without'. As concerns their utterances of (4), our use of "true" and theirs work in lockstep. If defaulting is the *only* way their use of "true" differs from ours, nothing prevents the rest of the applications of "true" in both languages from also lining up. And if the words "true" in each language truthfully apply to *exactly the same utterances*, what grounds could there be for saying that one of them expresses the genuine concept of truth while the other does not?

What all this shows is that the idea of explaining the truth of (4) in our language by appeal to *a priori* conceptual truths about truth is misguided.[27] I noted the *ad hoc* character of this move at the outset. The view is taking what looks like a plainly linguistic, compositional convention and tries implausibly to rebrand it as a conceptual truth. But possession of the concept of truth won't settle what kind of linguistic community we are in, and what kinds of defaulting our community's use of "true" condones.

Though the appeal to *a priori* cognition is perhaps weak, it is worth discussing in some detail in order to draw out a general lesson which will be of help in the next section. What we've learned is that there are different, *intra*-linguistically incompatible (but *inter*-linguistically compatible) compositional conventions that are equally good at reporting truths. Two linguistic communities can agree on what things should be called "true", but differ in whether defaulting for "true" is permissible within their respective languages. This sounds surprising, until we realize that this possibility is afforded only by the odd features of certain special reflexive applications of semantic vocabulary. Normally, if we shift about the proper application of a predicate (as we do in taking up, or abandoning, defaulting conventions for "true"), the application of that predicate ceases to line up with any property it originally reported. But as we shift the application of "true", we thereby unusually shift about instances of the property we were reporting as well. Sometimes this pair of concurrent shifts occurs in a harmonious way, accounting for the plurality of acceptable compositional conventions for talking of truth.

What this means is that for truth, like for no other property we know of, we must effect a sharp separation between two things: the property we are out to record and the conventions governing the word we use to do the recording.[28] For truth, there is no immediate, simple derivation of the latter from the former. This would be one important reason, given (4)'s truth, why Truth-Extensionalism would fail: Truth-Extensionalism *only* models the information

[27] Our consideration of defaulting conventions also, relatedly, yields an objection to what Gupta & Belnap (1993) call the *supervenience of semantics*—the claim that the application of "true" in a language is settled by non-semantic facts and the meanings of other words in the language. Compare a related discussion in Kremer (1988) of the grounds for taking the best construal the position of Kripke (1975) as involving a rejection of the supervenience claim.

[28] I don't mean to be using the word 'property' here in an inflationary sense. This is, in effect, just a restatement of the claim that we need to separate extension or intension from semantic value.

that "true" correctly applies to all and only the truths. But if that holds of our language, it can equally hold of the language of non-defaulters. Something else—something conventional and compositional—is what distinguishes us.

4 Semantic Generalities

I've argued that if defaulters are true, they create compositional circularities within truth-extensionalist frameworks. That is, the linguistic, compositionally mediated information encapsulated by the framework modeling speakers' linguistic capacities, coupled with their non-linguistic knowledge, is insufficient to determine defaulters' truth-values owing to semantic circularity. But if (4) is true, speakers ostensibly have that knowledge, in part owing to their linguistic competence. So truth-extensionalist frameworks as theories of linguistic competence must be rejected.

As I say, all this holds *if* defaulters are true, as they appear to be. I've argued that the case for taking appearances at face value here is strong. But I also acknowledged that there is a complicated mix of empirical and theoretical concerns which could motivate a special error theory concerning the relevant speaker judgments. If this error theory pans out, we are forced to conclude that we have a little bit less expressive flexibility in English than we thought. Sid loses the ability to state his promise will be fulfilled, while fulfilling it, for example. Maybe this is just something we have to live with.

But even if defaulters fail to create compositional circularities, examining them clarifies the preconditions for other utterances to create such circularities. The key lesson was learned at the end of the previous section. A truth-extensionalist theory effectively tells us that speakers' semantic competence with "true" is exhaustively modeled by the information that "true" (as used in a context, and evaluated at a world, etc.) applies to all and only the truths. That, after all, is the only information that is supplied in associating "true" with a semantic value that merely maps a context-index pair to a set of truths. The problem is, as recently noted, that what the truths *are* might *depend* on how we report them. There may be different, intra-linguistically incompatible methods to have "true" apply to all and only truths. If there are two such methods, then were a speaker to come to 'understand' only that their use of "true" should apply to all and only the truths, they wouldn't yet be in a position to know which of those methods their use of "true" follows. After all, two languages with relevantly different conventions (e.g., defaulting conventions) could agree that "true" applies to all and only things among the set of truths. So different extension assignments for "true" in each language cannot be what differentiates them. Accordingly, if there are possible languages that differ with respect to any such conventions, and a speaker only understands that "true" applies to the truths, some information about how "true" is to be applied will be logically independent of all linguistic information the speaker has acquired. That is, some truth-evaluable, productive uses of "true" will be essentially underdetermined by our model of the speaker's linguistic competence.

Note that this is not a problem for a truth-extensionalist theory if speakers exhibit uncertainty in how "true" is to be applied to the undetermined cases, or if there is a sense that the term "true" is defectively applied to them. Giving a semantic value for "true" that fails to specify how it is to be applied in such cases may be the perfect way to model the relevant speaker behavior. So the real problems for Truth-Extensionalism only arise when we have a use of "true" whose application is non-defective (simply true or false) *while* the application of "true" is underdetermined by the stipulation that "true" apply to all and only the truths. When these two conditions are met, a truth-extensionalist theory will not be in a position to explain how speakers correctly allot the truth-values in question. There will be a non-defective, conventional use of "true" which the theory cannot help predict.

So examining defaulters helps reveal the conditions under which a compositional circularity arises for truth-extensionalist views. We need a sentence S for which two claims hold.

(i) *Empirical/Semantic Claim:* Speakers correctly judge S true (or false) through normal compositional means.

(ii) *Logical Claim:* S would be non-true (non-false) in a language $\mathcal{L}$ which differed from ours *only* in adopting a different conventional use of "true", compatible with its correct application to all and only truths.

Let me spell out why these two conditions are sufficient to generate compositional circularities. If speakers can become aware that some sentence S containing "true" is true (or false), by conventional compositional means (condition (i)) we need an explanation of how they can become so aware. Let I denote the sum of the information encapsulated by the theory of speaker competence *excluding* that mediated by the semantic value of "true", along with all non-linguistic *a priori* and *a posteriori* information a given speaker possesses. To this we add the aspect of speaker competence modeled by the semantic value of "true"—call it i—to I. Together, I and i should conjointly determine that S is true (false), if we've chosen our semantic value for "true" correctly. After all, I and i together exhaust the sources of information speakers have, both tacit, linguistic information had in virtue of competence with their language, and all non-linguistic information gained by *a priori* and *a posteriori* means. Now, suppose i is the information supplied in a truth-extensionalist theory: only that "true" applies to all and only truths, or that it has the truths as its extension at a world. Then, if condition (ii) holds, there is a possible language which differs from that of our target speakers only in conventions for the use of "true", but *consistently* with both i and I. That is, in the second language "true" also applies to all and only truths. If S has a different truth-value, then obviously I and i cannot settle that S is true (false). What the possibility of this alternate language shows is that I, i, and S's non-truth (non-falsehood) are consistent. So this would reveal that our hypothesis—that the semantic value of "true" determines merely an extension consisting of all and only truths, represented with

i—was incorrect. We need something besides i to model the aspect of speaker competence relevant to understanding the meaning of "true".

Note that these two conditions (i) and (ii) also specify the sense in which sentences generating compositional circularities for Truth-Extensionalism are 'special kinds' of semantic circularities. We saw that truth-tellers like (T) arguably do not generate such compositional circularities. If so, it is because (i) fails to hold of them: speakers don't judge truth-tellers true or false (whether by normal compositional means, or otherwise). We also saw that Kripke's interdependences, like (J) and (N), do not generate compositional circularities. This is because (ii) does not hold of them. In the relevant circumstances, there is no way of adopting a use for "true", applying to all and only truths, which assigns (J) and (N) truth-values other than those they actually have.

In this section, I'll use conditions (i) and (ii) to pinpoint a second class of claims that generate compositional circularities for truth-extensionalist theories. These claims are drawn from *semantic generalities*: sweeping statements about patterns of truth-value allotments, which those very statements may instantiate. Semantic generalities should sound familiar. These are the very kinds of statements we discussed in §1, which make providing a compositional semantics for words like "true" imperative. In §1, I effectively argued that conventionally-valued semantic generalities cannot be turned aside with an error theory, the way defaulters like (4) might be. Jettisoning defaulters stops Sid from talking about how good he is at keeping his promises. But jettisoning semantic generalities undercuts the expressive resources we need to do foundational work in the philosophy of language. When we tamper with such statements, we tamper with the theoretical goals in giving formal theories of truth quite generally. So if we find semantic generalities that create compositional circularities, the truth-extensionalist cannot relegate the judgments to an error theory. To do so would not merely add to their theoretical and expressive costs, but could well be self-defeating, by threatening to undermine the coherence and purpose of the very compositional theory they put forward.

As before, we need to be careful when looking for the relevant semantic generalities—this time, for two interrelated reasons. First, semantic generalities pose a well-known stumbling block for formal theories of truth. Second, the reasons for this likely have to do with getting the semantics of expressions *other* than "true" correct.

Consider, for example, that some of the simplest and most obvious generalities, like (5) and (6), elude expression by quite sophisticated formal theories of truth.

(5) Everything true is true.

(6) Nothing is both true and not true.

On many of the familiar fixed-point constructions developed in Kripke (1975), both (5) and (6) come out untrue. These include the least fixed-point, the largest intrinsic fixed-point, and any maximal fixed-point constructions for Weak and Strong Kleene schemes. Many other theories share similar aberrations.

Surely (5) and (6) (setting aside dialetheist views) should come out as truths of English. What precludes their truthful expression? A candidate culprit here is the presence of some projective truth-value other than mere truth or falsity. (By "projective" I mean the value is 'infectious', or tends to be inherited by logical compounds.) Such a value is present in both the Strong and Weak Kleene schemes. But there are other possible culprits worth considering.

For example, the languages used in standard fixed-point constructions render (5) using a unary quantifier from first-order logic, when a binary generalized quantifier more appropriately helps capture its semantics. Introducing generalized quantifiers could allow (5) to be true for intuitive reasons, even within a theory that otherwise admits projective truth values other than truth and falsity. After all, (5) would become a logical truth, even in three-valued logics, on such a construal.

Also, problems with (6), though doubtless connected with projective truth-values, may also concern the meanings of logical connectives. Bracketing paradox, (6) might come out true in a language admitting truth-value gaps as long as it also boasts a form of 'exclusion negation'—a use of negation that is true if its compliment is anything but true. That connective may be needed to capture the intuitive, true reading of (6).

I review these obstacles to the truthful expression of (5) and (6) because one may be tempted to think they could generate compositional circularities for truth-extensionalist theories. Ordinary speakers judge (5) to be true, ostensibly by normal compositional means, seemingly satisfying my semantic claim (i). And some formal theories of truth in which "true" accurately applies to all and only truths, like fixed-point constructions, render (5) untrue, seemingly satisfying my logical claim (ii).

But the appearance here is a little deceptive. Though the semantic claim (i) is probably satisfied for (5) and (6), whether the logical claim (ii) is also satisfied should be controversial. For the logical claim to hold, the theories in which (5) is untrue (e.g., various fixed-point theories) must give that sentence a semantics differing from the English use of (5) *only* in the conventions governing the behavior of "true". It was a stipulation of the logical claim that only the behavior of "true" be changed, and with good reason. After all, if we get a truth-value shift in a sentence S by shifting around not the use of "true", but instead the logical form of the sentence, or the meaning of separate words, then a plausible explanation of how speakers settle the truth-value of S is that they are aware of its logical form and the meanings of the relevant words. If this were to happen, then in my argument above the alternate language constructed would be in conflict with I (which subsumes all the information in the model of speakers' linguistic capacities vis-à-vis words other than "true"). And if so, the argument that S generates a compositional circularity wouldn't go through.

Just such worries cast doubt on whether (5) satisfies my logical claim (ii). If there are projective forms of semantic defect belonging to applications of "true", it is very plausible that speakers who judge (5) true are recognizing the use of a binary generalized quantifier. No truth-extensionalist language in which "every" is given the appropriate meaning will allow (5) to be untrue—it will become

a logical truth even on three-valued schemes.[29] So truth-extensionalists may have a possible legitimate explanation of how speakers recognize (5)'s truth. (5), though 'about itself', doesn't generate a compositional circularity. Similar remarks will apply to (6).

What this shows is that to uncover semantic generalities that create compositional circularities, we may have to look for somewhat more specialized cases. And arguing those cases generate circularities will require a great deal of care: we not only have to show how the truth-value of the sentence can shift owing to different conventions governing the truth predicate, but we also have to argue that other logical words aren't in any way contributing to the shift.

The work here is tricky, so I will focus on just one case. And in discussing it, I will be appealing to the following assumption.

(C) There is a projective form of truth-value other than simple truth or simple falsity (like the third value in the Strong or Weak Kleene schemes) which predications of "true" may bear owing to reflexive applications.

(C) can be motivated in several ways. It could perhaps be motivated just by consideration of odd semantic reflexivities like (T).[30] But the real reason for taking (C) on board is because in the model-theoretic tradition on which I'm focusing, devising a language with a truth-predicate whose extension consists in the true sentences of the language (essentially, a fixed-point in the sense of Kripke (1975)) is not obviously, or at least not easily, achievable, owing to the presence of liar-like circularity, without the assumption of something like (C). As such, I take it as satisfactory for my purposes here if I can argue against candidate compositional theories in the model-theoretic tradition that take (C), or something like it, on board as an assumption.[31]

[29] A caveat: I'm not saying it is easy, or even possible, to construct such a language. I'm provisionally conceding the possibility of such a language, for the sake of argument, on behalf of an objector.

[30] Perhaps when supplemented with empirical and theoretical grounds for the existence of projective values in other contexts (see Shaw (2014, forthcoming)).

[31] In the tradition of developing axiomatic theories of truth (see, e.g., Halbach (2011) for a survey), things become murkier. This is because it is not clear there is an uncontroversial sense that can be given to the claim that within an axiomatic theory the truth predicate has a particular extension (where we can also think of the truths of the theory to consist in some set of sentences, such that it becomes an open question whether this set, and the extension, line up). Some axiomatic theories can be shown consistent by model-theoretic arguments. And when the model providing the consistency proof is 'taken seriously' by the theorist, we may be willing to say that the extension of the truth-predicate in the axiomatic theory is just that given by the model, and the truths of the theory those assigned truth in the model. Then, in a sense, the information of interest to us, as compositional semanticists, is all in the model-theoretic setting, and what is distinctive about axiomatic investigation drops away. But a theorist needn't take the model theory to be, e.g., giving significance to the axiomatic theory. If so, there may be an open question as to what sense, if any, is given to an expression's 'having an extension' *for compositional purposes* within the theory (and even what the truths of the theory are). In part for this reason, these 'purely' axiomatic theories aren't subject to the criticisms I'm developing here. Indeed, such axiomatic theories may well provide us with one broad strategy for avoiding compositional circularities. Still, I believe there other important worries about the viability of pure axiomatic theories as contributions to specifically linguistic, compositional explanatory enterprises. I discuss both issues, briefly, in §5.

Now, on any reasonable construal of the semantics for disjunction, (7) should be simply false.

(7) Some disjunction which is not true is such that all of its disjuncts are true.

But, on the assumption of (C), there are formal theories of truth in which the extension of "true" consists of all and only the truths in which (7) is not false. For example, it is neither true nor false on the minimal Kripkean fixed-point for a suitably expressive language containing (7) on both the Weak and Strong Kleene schemes. As we've just seen, we have to be careful about what this shows. Perhaps (7)'s non-falsity derives primarily from the treatment of quantification, or the logical connectives, in these fixed points, as was argued for (5) and (6). But, in this case, there seem to be good arguments which show that we cannot trace trace the possibility of the truth-value shift to these other expressive resources. To see this, it is easiest to focus on a different problematic sentence: an equally false disjunction of (7) with itself.

(8) Some disjunction which is not true is such that all of its disjuncts are true or some disjunction which is not true is such that all of its disjuncts are true.

Supposing (8) can be false consistently with "true" applying to the truths, let us now ask whether it could also bear our hypothesized third projective value, call it "*u*", consistently with "true" applying to the truths. This does seem possible, regardless of how we construe the semantics of other terms in the sentence.

First, as effectively noted before, if the semantics for (8) makes use of unary quantifiers and is interpreted relative to either the Weak or Strong Kleene schemes, we have constructions which show how (8) and both its disjuncts can bear u consistently with the stipulation that "true" hold of all and only the truths. But unlike with (5) and (6), the absence of generalized quantifiers or exclusion negation can't be the source of the problem.

In particular, even if we accommodate both resources, we can show (in somewhat painful detail) that (8) can consistently bear the value u owing to semantic circularity. To do this, we start with three assumptions:

(a) "true" used in (8), as applied to (8), or either of its disjuncts, bears the projective value u (owing to circularity, as per (C)),

(b) "Some" and "all" in (8) are binary generalized quantifiers, and

(c) "not" in (8) expresses exclusion negation.

The goal is to show, on these assumptions, that (8) will compositionally evaluate to u consistently (especially, consistently with (a)). This would show that the reason (8) is consistently assigned the status u in certain fixed-point constructions has to do with a semantic circularity that can't be overcome by 'getting

the rest of the semantics and logical form of (8) right'—that is, allegedly, by the introduction of generalized quantification or exclusion negation.[32]

To see this, note first that on the assumption of (a) and (c), (8) satisfies the quantifier restrictor of "some" in its first disjunct. (8) after all is a disjunction. By (a), "true" in the quantifier restrictor of its first disjunct gets the value u when applied to (8). And since "not", in "not true", expresses exclusion negation by (c), "not true" as applied to (8) is therefore true. So (8) satisfies the quantifier restrictor of "some": (8) is a "disjunction which is not true" on the relevant hypotheses. Since (8) satisfies the quantifier's restrictor, we need to predicate the matrix "is such that all of its disjuncts are true" of it in evaluating the status of the existential quantified claim. In particular, "true" as used in the quantifier matrix of "some" must be applied to each of the disjuncts in (8), since "all" quantifies over both disjuncts. But "true" as applied to each of the disjuncts bears the value u, again by assumption (a). As a result "all of [(8)'s] disjuncts are true" itself bears the value u. So we have at least one object which satisfies the restrictor of (8)'s first disjunct, but evaluates to u in its matrix: (8) itself.

Note, as a logical matter, there will never be instances of the existentially quantified claim which satisfy its restrictor and also satisfy its matrix (no untrue disjunction will have only true disjuncts). Given this, the first disjunct of (8) itself will evaluate to u. Most or all other sentences besides (8) will either fail to satisfy its restrictor, or fail to satisfy its matrix. There will be no 'counterexamples' to the falsity of this generalization. But at least one case—(8) itself—will remain 'unresolved' (or whatever status u indicates). So the quantified claim itself will not become false, but u (whatever that value may be). Of course, the second disjunction will be u by parallel reasoning. So both of (8)'s disjuncts, and (8) itself, will have the status u.

This is what we've just shown: (8), and its disjuncts, can consistently bear a third projective truth-value in accordance with (C), regardless of whether it is interpreted to use generalized quantifiers, and regardless of whether it boasts a form of exclusion negation. So the reason (8) gets a third status in fixed-point constructions has nothing to do with the absence of these logical tools. It has to do with a special, ineliminable form of semantic circularity. What this means is that if it is possible for "true" to have an extension as its semantic value at all (and (C) holds), we should be able to find a language in which (8) gets a value (non-falsehood) other than the value it *actually* has (falsehood), consistently with that assumption. That is, not only does (8) satisfy condition (i) above, but condition (ii) as well. (8) generates a compositional circularity in truth-extensionalist frameworks. So those frameworks must be rejected.[33]

[32]The scare quotes are to flag that I'm not actually assuming that the logical form of (8) really does have a use of exclusion negation. I'm just framing a reply to someone who claims that it does, and that if a truth-extensionalist framework were changed to accommodate this fact, (8) could not longer consistently bear the value u, so that (ii) would fail for it.

[33]It is important to keep track of how the dialectic proceeds here. I am not claiming, nor am I beholden to claim, that we can devise a language with exclusion negation, generalized quantifiers, or both, that behaves as a fixed point in roughly Kripke' sense. Indeed, if we

Once we examine cases like (7) and (8), it becomes clearer how we can multiply examples of sentences satisfying (i) and (ii). That is, we can multiply sentences for which there is a discrepancy between how normal speakers evaluate them, and how they could have alternatively been evaluated if we altered only the compositional behavior of "true" consistently with its truthful application to all and only truths. These examples all involve a semantic generality meeting two conditions. First, the generality should report a general phenomenon, including itself as an instance. Second, crucially, the truth-values allotted within the sentence itself must be important to ascertaining whether the generality in fact holds. (8) (redundantly) reports a failed generality of which (8) itself is a critical instance: (8)'s status can compositionally affect (8)'s evaluation in a truth-extensionalist theory, given the truth-values allotted to other sentences. That's how (8) becomes perniciously circular.

To appreciate this second condition, contrast (9) and (10)—the disjunction of (9) with itself.

(9) Every disjunction which is not true is such that all of its disjuncts are true.

(10) Every disjunction which is not true is such that all of its disjuncts are true or every disjunction which is not true is such that all of its disjuncts are true.

(9), though similar to (7), will not generate a compositional circularity. Intuitively (9), like (7), should come out false. But this will in fact occur in any language which has the logical resources to capture what (9) would typically be used in English to say. In such a language, all we need to falsify (9) is to find one counter-instance, like (11).

(11) 2+2=5 or 2+2=6.

We can compositionally settle that (11) is false, and therewith that it satisfies the restrictor of (9) and falsifies its matrix. That's sufficient to show (9) is false (and (10) as well), *independently* of what statuses (9) or (10) have. In this respect (9)'s truth-value will never hinge on its own truth-value as does (7) or (8).

(9) doesn't 'significantly' concern its own status because its own status isn't necessarily relevant to its evaluation. But there are other ways in which a

integrate generalized quantifiers, a monotonicity constraint is no longer satisfied, and so the details of, e.g., a Kripkean fixed-point construction are problematized. And constructions are also problematized if we include exclusion negation. But at this stage of the dialectic it is the burden of proof of the hypothetical objector to show that when we add generalized quantifiers or exclusion negation to a coherent truth-extensionalist theory (e.g., a fixed point theory) the value of (8) will necessarily shift to falsehood. What I'm arguing here is that there is no grounds to think there would be such a shift, simply owing to the introduction of the relevant logical tools, provided we succeed in constructing something like a fixed-point. If there is nothing like the relevant fixed-point construction to begin with, that is a concern for the *objector* in the dialectic.

generality might not be significantly about itself. Sentence (5)—"Every true sentence is true"—is an example. If it uses a generalized quantifier, (5) will be true as long as "true" has any extension at all. (5)'s truth is secured on logical grounds by the meaning of its quantifier. So there is no danger that its own status could interfere with assessing its truth-value.

So, like we saw with defaulters, not just any semantic generality which 'concerns itself' generates compositional circularities. But this doesn't detract from the point that (presumably infinitely) many sentences do meet the conditions required to generate such circularities provided (C) holds.

Also, as already noted, an error theory concerning these judgments—for example that (7) and (8) are not actually false—is very hard to swallow, and comes with some quite striking theoretical consequences that I highlighted in §1. So we can't take the recourse we saw might be available for defaulters.

Indeed, it is worth flagging that examining semantic generalities also *strengthens* the case for the truth of defaulters like (4). Cases like (8) show we will need 'benign' reflexivity (that is, reflexivity that results in ordinary truth-evaluability) for certain semantic generalities anyway. This effectively undercuts the force of the objection to defaulting that makes use of an error theory, which relied essentially on an analogy between (4) and (T). If we think (7) or (8) is false, we concede that some reflexivity is benign, even if some, like that in (T), is not. There is then no reason to think (4) must be assimilated to the problem cases like (T) as opposed to the truth-evaluable cases like (7) or (8)—indeed, all evidence points to the opposite assimilation.

At least, all this holds if my arguments concerning semantic generalities hold. But one might worry there are newer, different objections to those arguments. Let me mention three.

First, one might try to revisit the option, considered in §3, that conceptual truths about truth will settle the status of (7). This option gains intuitive force when we consider semantic generalities that state general facts about linguistic or logical tools, as opposed to defaulting cases which often involve contingent facts. Nevertheless, the reply to this objection is the same. The structure of our argument required us to produce evidence for shifting assignments among logically possible languages in which "true" reported the presence of all and only the truths.[34] The very existence of such languages provides as clear a case as could be provided that a concept of truth, divorced from compositional considerations, won't of itself constrain how the presence of truths can be reported.

There is a second, related objection which has more clout. Can't we claim that the languages described which don't allow (7) to be false are extremely *unusual* or *unnatural*—perhaps especially unnatural if one had the kind of interests we do when we report truths? After all, as I've stressed, ensuring that sentences like (7) come out false is vital if we are to speak a language which does the kind of foundational work we need in the philosophy of language. Isn't that enough to explain why (7) *is* false, and known by speakers to be so?

[34] Again, contingent on the methodological assumption that any such language exists to begin with.

Though this reply starts from correct assumptions, it misfires as a defense of Truth-Extensionalism. Recall that the problem for Truth-Extensionalism is that it treats competence with "true" more or less as awareness that "true" applies to the truths, when there are multiple distinct, intra-linguistically incompatible ways of creating such a predicate. This shows us that the conventions governing our use of "true" aren't exhausted by the claim that "true" represents the truths, but by further conventions. Granted, an account of why we adopt those further conventions might be that they are extremely useful or natural. But the fact that a convention is useful or natural doesn't detract from its *being* a convention. Our compositional theories for "true" are supposed to exhibit the rule-governed conventions underlying our uses of words, however natural they might be. The objection to Truth-Extensionalism is that it forgoes the resources required to represent them.

Put another way, naturalness certainly may figure in an explanation of why (7) is simply false in our language. But it would do so by figuring in an explanation of why we speak English, with its conventions, as opposed to some other language—not as a way of explaining away the conventions of English as conventions. Questions of 'naturalness' are simply irrelevant to whether a particular theory generates compositional circularities.

A third set of concerns is about (C). Could it be denied in the context of something like a truth-extensionalist theory? I'm not sure whether compositional circularities could be generated for Truth-Extensionalism on different, possibly weaker, assumptions. The obstacle here, as mentioned before, is that paradox makes it is extremely hard to construct truth-extensionalist theories that accommodate the benign reflexivity required of semantic terms without assuming something like (C). In any event, blocking truth-extensionalist theories that accept (C) is already to do a great deal of work winnowing candidates for our compositional semantic theories of "true". As we've seen, it effectively rules out Kripkean fixed-point models and theories of truth based on that construction. In conjunction with the case from defaulting, which requires no assumption like (C), this warrants an investigation of what compositional theories for "true" should look like if we want to avoid compositional circularities.

My primary concerns here have been negative, to establish the problems with Truth-Extensionalism. So this concludes my main line of argument. But it is worth a discussion, though it will necessarily be brief and sketchy, of what a positive alternative to such a theory would look like. What resources do we need to add to our theories to accommodate defaulters and my chosen semantic generalities? I want to say just a few words on this issue, and why it matters.

5 Truth-Proceduralism

I've argued that compositional theories that merely assign "true" an extension at a context are inadequate. This is not because such theories fail to be compositional, nor because they assign incorrect truth-values. Nor is it necessarily because there is no extension for "true" in the sense of there being no set of

truths—I've neither assumed, nor argued that there isn't. It is because those theories generate compositional circularities inconsistent with their explanatory purpose: they fail to explain productive uses of compositional tools by natural language speakers. Truth-extensionalist theories encounter these problems because they lack the resources to represent special compositional conventions that govern the uses of terms like "true". We need a liberalized conception of the semantic value of "true" that is capable of encoding such conventions.

There is no hope of supplying such a semantic value here, even in a 'toy' model. Even a simple version of this model would have to be at least as complex, if not more so, than something like a Tarskian or Kripkean theory. So what follows is, of necessity, said merely by way of advertisement.

Even if we can't give the details of such a theory here, we can use the lessons of the preceding sections to draw out some general and instructive information about what a semantic value for "true" overcoming the problems with Truth-Extensionalism will have to look like. We've learned that uses of "true" are governed by special kinds of conventions. If we reflect just a little on what exactly those conventions are, they should give us insight into what work a semantic value that represents them has to do. In particular, when we reflect on the conventions, we see that they can instructively be divided into two separable components. Let me start with the first.

Consider again our defaulter (4), and our semantic generality (8).

(4) Everything I say today will be true.

(8) Some disjunction which is not true is such that all of its disjuncts are true or some disjunction which is not true is such that all of its disjuncts are true.

What we want to know is how speakers are positioned to arrive at the conclusion that Sid's utterance of (4) is true or that (8), and its disjuncts, are false. We need a better model of their semantic competence with "true" which gives us a picture of how the information they are able to acquire by other linguistic and cognitive means are sufficient, in conjunction with that competence, to settle the truth of (4) and (8).

Let's focus on (4). Clearly, to settle the utterance's truth a speaker must have the information that all other utterances Sid made the day he speaks (4) are true. We know how the speaker can get that information, so suppose she has it. Now, does she need any further information? No. Or, at least, she needs no further information about the truth-values of various utterances. The thing that needs to be registered by our model of her semantic competence with "true" is that *that information is 'enough'*, for compositional purposes, to establish the truth of Sid's (4). In other words, our model of her competence should encapsulate the information that *if* Sid's other utterances are true, his utterance of (4) is as well. Something roughly analogous is true of (8): our model of a speaker's competence needs to subsume the information that the possibility of (8)'s defectiveness is excluded from consideration, in settling whether the generality, of which (8) might be an instance, holds. We can, as it were, exclude

(8) and its disjuncts (and perhaps other claims as well) from consideration when looking for counterexamples to the disjunctive existential claim (8) makes.

As argued in §3, the aforementioned conditional pertaining to (4) ("if the rest of Sid's utterances are true, Sid's (4) is true") is no conceptual truth. It is a conventional truth—a linguistic convention. This is a piece of information that needs to be represented in modeling speakers' competence with English. As such, we need to be able to extract this conditional from the semantic value of "true" (in conjunction with other linguistic and non-linguistic information).[35] This is the information that was nowhere encoded in a mere function from an index-context pair to an extension, or to the property of truth.

We can begin to frame this conventional information in a slightly more technical way, appealing to the notion of *semantic dependence*. Consider sentences (12a)–(12c).

(12) (a) Roses are red.

(b) (12a) is true.

(c) (12b) is false.

To settle that what (12a) says is true, you need only know facts about roses. (12a)'s truth depends, if on anything, only on 'the facts'—that is, the non-semantic facts. To settle that what (12b) says is true, by contrast, you need to know some semantic facts: facts about what (12a) says and whether what it says is true. Likewise, the semantic status of (12c) depends on the semantic status of (12b).

Iterated ascriptions of semantic terms, like (12a)–(12c), create a natural kind of 'hierarchy' in which the semantic properties of ascriptions 'higher up' in the hierarchy intuitively depend on the semantic properties of those below. At the 'bottom' of the hierarchy (if all goes well) things depend on 'the facts'. Saying that one sentence 'semantically depends' on another is effectively to place it above the other in this hierarchy. Something like this notion of dependence is implicit in the notion of *groundedness* in Kripke (1975). Other versions of the concept are also treated more directly in Yablo (1982), Gaifman (1992), Maudlin (2004), and Leitgeb (2005).

The foregoing characterization of semantic dependence has obviously been informal and schematic. It can be fleshed out in many competing ways, as some of the above authors have done, depending on how one interprets the nature of the relation, and its formal properties. For example, is the relation metaphysical—a kind of grounding? Is it logico-semantic—merely a means of recording which items a sentence predicates "true" of? And as regards its structure: Is it transitive—does (12c) depend on (12a) because (12b) does? Does (12a) semantically depend on nothing? 'The facts'? All, or merely some?

[35] I grant that the conventions in question might operate for a 'language as a whole', rather than being encoded at the level of individual lexical items, especially for a case like English. I'll continue to talk as if the conventions are encoded in the word "true" for now, since it is at least conceptually possible that defaulting conventions differ from semantic word to semantic word.

Once we ask these questions, we should recognize that there is probably not just one notion of semantic dependence here, but a family, with different structures and theoretical purposes. What *we* should be interested in is a semantic and perhaps quasi-epistemic version of such a dependence relation—what I will call a *compositional semantic dependence relation*. Very, very roughly, the relation could be specified as follows.

> *An utterance u compositionally semantically depends on an utterance u'* just in case information about the truth-value of u' is compositionally prerequisite in settling the truth-value of u.

This is the dependence relation which interests us as compositional semanticists—the version of a semantic dependence relation that is picked out on the basis of its relationship with compositional operations.

Some examples: (12a) compositionally semantically depends on nothing (facts aren't utterances, so there is nothing for (12a) to depend on given my definition). (12b) compositionally semantically depends on (12a). A competent speaker who compositionally settles the truth-value of (12b) *needs* information about the truth of (12a) to do so. Similarly (12c) compositionally semantically depends on (12b). (12c) does not, however, compositionally semantically depend on (12a): if you had information about the truth of (12b) by non-compositional means (e.g., by being told), you could still compositionally settle the value of (12c) without knowing anything about (12a). So the compositional semantic dependence relation as I've characterized it is non-transitive. The truth-teller (T) arguably compositionally semantically depends on itself. That, arguably, is why it is defective in some way: it compositionally requires information about its own truth-value, where there is no 'antecedent' way of obtaining it. Finally, Sid's utterance of (4) compositionally semantically depends on all his *other* utterances spoken that day. That is, even though Sid's utterance *attributes* truth to itself, as I argued in §3, his utterance of (4) does *not* compositionally semantically depend on itself.

Why? If (4) is knowably true, this last claim is just a more technical way of saying that to settle (4)'s truth, one doesn't need antecedent non-compositional awareness of (4)'s truth. Otherwise one could never know it was true, and (4) would probably be defective (as (T) seems to be).

We can also also use this new terminology to restate our more recent lesson about the meaning of "true". I pointed out that the semantic value of "true" needs somehow to register the information that the truth of all Sid's other utterances is, by itself, sufficient to compositionally settle the truth-value of Sid's utterance. In our new technical terms: "true" needs to somehow encapsulate the information that (4) does not compositionally semantically depend on itself. Again, an extension assignment in standard model-theoretic frameworks simply won't model that information, even implicitly.

So: the first component of the conventions that contribute to the benign character of my reflexivities is represented by a compositional semantic dependence relation. In particular, these conventions dictate that sometimes compositional dependences of this kind are (more or less stipulatively) 'restricted' so

as to facilitate unambiguous compositional derivation in the presence of *some* kinds of virtuous semantic circularities. But I also mentioned that there was a second component of the conventions. To see this, note that there are hypothetical languages in which uses analogous to (4) purposefully and compositionally default to a third status, like a truth-value gap, or even to falsehood. Those languages will agree with our use of "true" that utterances like (4) only compositionally semantically depend on Sid's other utterances. So obviously the difference between "true" in our language and in these other languages depends on a further conventional feature: *how the truth-value of (4) is to be settled on the basis of its restricted compositional semantic dependences.* In our language, (4) is true if all its compositional semantic dependences are true, false if one of them is false. In the hypothetical language just envisioned, by contrast, an analogous utterance could be false even if all its compositional semantic dependences are true. The difference between the languages is just in these stipulated conventions for evaluating (4) on the basis of its compositional dependences.

Thus, compositional conventions for the use of "true" seem to encode two elements: relations of compositional semantic dependence, and a method for assigning truth-values to predications of "true" on the basis of those, perhaps conventionally restricted, compositional dependences.

Note that these two conventions are all that is needed. Indeed, the two kinds of conventions are individually necessary and jointly sufficient to cope with defaulters and semantic generalities. Dependence relations are necessary, since without the information that defaulters and generalities can be excluded from consideration when evaluating themselves—that is without the information that they're at least not compositionally semantically self-dependent—a model of speaker competence with "true" is sent on an unending loop in evaluating the statuses of the relevant utterances. Evaluation methods are also necessary, because even with information about which restricted set of semantic properties are needed for compositional purposes, owing to restricted dependences, there is still latitude in how that restricted information is to be applied.

Joint information about both dependence relations and evaluation methods is sufficient, because together they entail the conditionals that bridge the truth-values assigned to dependences and the truth-values of our problematic utterances. The joint information entails that if the rest of Sid's utterances are true, Sid's (4) is true. It entails that if there are no counterexamples to the generality in the first disjunct of (8) besides (8) (and perhaps some other interdependences), then that disjunct is false. Those are the very pieces of conditional information that we could not retrieve, but needed to retrieve, on the truth-extensionalist view. And with those pieces of information, the truth-values of the relevant utterances will be entailed by independent information that speakers can acquire—the information that the antecedents of the relevant conditionals are true.

So, as promised, reflecting on the conventions required by defaulters and generalities has clarified the work that a revamped semantic value for "true" needs to perform. The thesis that the semantic value for "true" does it is the following.

Truth-Proceduralism. The semantic value of "true" should

(a) contribute to the determination of a conventional compositional semantic dependence relation for uses of "true", and

(b) determine a conventional function which maps an utterance u containing predications of "true", and an allotment of truth-values to its conventional compositional semantic dependences, to a truth value.

A little metaphorically, the semantic value for "true", directly or indirectly, encodes two things by convention: paths through utterances along which truth-values can be compositionally assigned, and a procedure, or rule, for assigning truth-values based on assignments earlier on those paths.

The "proceduralism" in "Truth-Proceduralism" signals that the meaning of "true" is not just an extension, or intension, but a more liberalized semantic value—some special kind of method. This is a method in which conventionally determined partial information about the set of truths is used in a sequential, compositional assignment of truth-values to utterances. We've always known that speakers proceed in assigning truth-values to utterances in roughly specifiable patterns. What is unique about Truth-Proceduralism is that it constitutively links these patterns to the meaning of "true". Unlike on other theories, the source of the procedure's structure is partly semantic, not merely epistemic.

Before proceeding, I need to mention two caveats.

First, and perhaps most importantly, truth-proceduralist views do not simply treat the semantic value of "true" as determining a provisional extension in an utterance u, given by some subset of u's compositional semantic dependences. That would be to think of the word "true" as context-dependent and constantly shifting in meaning (or worse, it may be to think of the truths as constantly changing). This would eliminate compositional circularities, but at the cost of simply denying the basic data, and altogether eliminating our ability to employ genuinely reflexive applications of semantic terms, as our foundational theories seemingly need.

For example, if we think that in Sid's (4) "true" has an extension as its semantic value, which is (say contextually) restricted in a normal compositional derivation only to include truths other than Sid's utterance of (4), then speakers may be able to compositionally assess Sid's (4). But, importantly, Sid's (4) (if it is true at all) would now have to *express something different* from what Sid intended to report, and different from what we find speakers report he expresses. It would express a claim that involves something like a restricted quantifier. In particular, Sid would not be stating that the conditions of his promise to speak only truths will be fulfilled. Also, bizarrely, if Sid were to state (4) and someone else were use that sentence right afterwards, they would be *saying different things* (since another speaker uttering (4) would produce a new utterance that no longer required defaulting conventions, and so no longer required the relevant restrictions on quantifiers or extensions). This gets the data all wrong. Similar, and more pressing, remarks hold for (8): if in it "true" has a (again perhaps

contextually) restricted extension as its semantic value, it simply doesn't state the intuitive false generalization about *all* disjunctions—itself included.

Truth-Proceduralism is precisely meant to skirt these worries. Truth-proceduralists simply *give up* the use of extensions as the denotations of "true" in compositional derivations altogether, thereby giving up the relevant metaphysical and expressive commitments that come with their use. They can, and should, reject the metaphysical commitment that restricted compositional semantic dependences correspond to shifts in which property is being talked about. And they can, and should, reject the expressive commitment that a restricted set of such dependences changes what is 'talked about'—what is in a quantifier's domain, for example. Rather, to restrict the compositional semantic dependence relations is only to make a conventional, linguistic stipulation about which restricted information about the set of truths figures in certain compositional operations. To maintain this, crucially, any partial set of information about truths figuring in uses of "true" must not be thought of as an acting *pro tanto* extension. To do that would force us into uncomfortable views about truth, or what statements we are making, that actually undercut the very work Truth-Proceduralism is tailored to accomplish.

A second caveat, related to the first, is that although adopting Truth-Proceduralism involves a significant shift in our conception of semantic values of words, it needn't involve any significant shift in how we construe what is expressed by whole sentences. Though on this view the semantic value of "true" is highly irregular, this semantic value ultimately only gives us a new way of settling truth-conditions for whole sentences on the basis of their parts. The 'output' of compositional processes is still truth-conditions, so the view is compatible with views that treat propositional content as having such truth-conditions. Put another way, Truth-Proceduralism in no way requires us to adjust our conception of 'assertoric content'—it merely requires us to adjust our conception of how that content is compositionally generated.

Now, as I've said, the above conditions (a) and (b) in the definition of Truth-Proceduralism are merely constraints on a semantic value for "true". They specify the work that extension assignments were unable to do, and that procedural values must accomplish. This clearly doesn't tell us exactly what truth-proceduralist theories will look like. As I say, giving even a 'toy' implementation is beyond the scope of this paper. But it might be helpful to note that there are already two different kinds of highly developed classes of formal theories of truth that can accommodate Truth-Proceduralism.

The first such theories we can call "explicitly truth-proceduralist". These are theories which are constructed by explicitly developing the elements in (a) and (b) in giving a semantics for "true". Versions of the operational pointer semantics developed in Gaifman (1992, 2000) are like this. On (a slightly simplified version of) the theory of Gaifman (1992) for example, sentence tokens and their tokened parts (which form part of a larger class of 'pointers' for Gaifman, a notion we won't dwell on here) stand in 'calling' relations to each other. Truth-functionally complex tokens call their immediate constituents. Quantifiers call their substitutions instances. And any token ascription of truth or falsity to

another token calls that second token to which truth or falsity is ascribed. The calling relations create a directed graph, linking truth-bearing tokens to each other in a kind of hierarchy. The semantics, in addition to stipulating these calling relations, also stipulates a set of rules along which truth-values can be assigned 'upward' along the dependences in the graph (given more standard interpretational information for the non-semantic vocabulary). The semantics' unorthodox nature is appreciated by Gaifman who describes the theory as "a new kind of semantics in which truth values are assigned to pointers [e.g., tokens] and the usual recursive definition of truth is replaced by a set of rules for evaluating networks" (Gaifman (1992) p.227). Calling relations are, in effect, a version of what I called a semantic dependence relation. The semantics consists in stipulating these relations, along with rules for assigning truth-values along them. It is thus a truth-proceduralist semantics in the most immediate sense. I will have a little more to say about its structure in §6.[36]

But for now I want to note that my constraints above on semantic values allow for other theories, which we can call "implicitly truth-proceduralist". Such views don't explicitly detail compositional semantic dependence relations and rules for assignment in giving a semantics for "true". Instead, their semantics *induces* the elements (a) and (b). A broad class of theories which may do this include axiomatic theories. An axiomatic theory of truth gives rules for settling the values of utterances (as determined by derivability from the axioms along with empirical information, say). And those rules give an implicit characterization of a semantic dependence relation (as sentences one may yet have to prove, to establish the truth of the sentence one is presently trying to settle). Such a theory can explain defaulters and generalities in the way suggested above, perhaps simply treating the generalities as axioms, and directly stipulating separate axioms governing when defaulters may default. Such theories are structurally poised to overcome the particular obstacles for Truth-Extensionalism detailed here, precisely because they may enhance their model of the understanding of "true" arbitrarily with added axioms governing its behavior. I do want to flag that although the theories have this virtue, I think they face serious further challenges in satisfying the goals of a compositional theory if inferential relations are pervasively taken as explanatorily fundamental, instead of being derived from something more like a traditional recursive assignment of truth-conditions. Still, there is no space to discuss these issues here. It suffices for now to note that axiomatic theories are fully compatible with the particular arguments I've given

[36] As noted on Gaifman (2000) p.85, these kinds of formalisms have natural antecedents in 'operational semantics' for programming languages (see Winskel (1993) Ch.2). Gaifman often describes these kinds of semantics as "non-compositional" (e.g., Gaifman (2000) p.83), which may make them seem inappropriate for the explanatory purposes I am interested in. But Gaifman's sense of "non-compositional" does not, or does not obviously, mark something as being at odds with the working assumptions I gave in §2. For example, Gaifman acknowledges the importance of accommodating productivity facts of the kind I've claimed characteristically drive compositional investigation (Gaifman (1992) p.236). There are some terminological issues to iron out here in how to use the term "compositional". I am ambivalent about whether there is a fruitful sense in which a procedural semantics counts as being non-compositional. But I will not pursue this terminological question here.

so far.

As I conceded, the discussion in this section has proceeded at a high level of abstraction. This was necessary to gain an appreciation for the form shared by compositional theories that avoid the danger of generating compositional circularities. Still, even with only this abstract form in hand, we are poised to draw some important lessons that the transition to procedural semantics will involve. In particular, we learn a lesson about how to think not only about the virtuous circularities of foundational semantics, but the vicious circularity of the liar.

6 Procedural Accounts of Paradox and 'Token-Sensitivity'

So far the liar has played a secondary role in my discussion. The liar drives us to reflect on the nature of semantic circularity which, I've argued, must be accommodated if we are to legitimate foundational accounts of meaning that give significance to formal investigations of truth or other semantic notions. The liar is an obstacle that alerts us to both the importance, and the difficulty, of providing those theories. What I've been arguing is that if the foundational theories are possible to formulate at all, their statements will require treatment with a compositional theory entirely unlike those found in the model-theoretic tradition. Indeed, ordinary speakers' use of semantic terms provides evidence that we already speak a language that must be modeled in those non-standard ways. The resulting transition to a procedural semantics has implications for our understanding of the structure of compositional semantics as a linguistic enterprise, the character of our logical theories, and our understanding of expressive power. But rather than broaching these topics, I want to bring us back full circle with implications that procedural semantics may have for the liar.

Some authors have claimed that sentences containing semantic predicates exhibit some kind of sensitivity to the circumstances of their tokening. Consider what Gaifman (1992) calls the 'two line paradox'.

> *line 1*: What's written on line 1 is not true.
>
> *line 2*: What's written on line 1 is not true.

What's written on line 1 is paradoxical for familiar reasons. One might wish to call it untrue on those grounds. This is precisely what happens in line 2. If we can consistently pronounce paradoxical utterances untrue, one might expect that we are able to do so because new tokens of the same type are somehow immune to the defects of the original paradoxical utterances. Versions of something roughly like this idea have been noted and exploited in different ways by a number of authors, including Parsons (1974), Burge (1979), Gaifman (1992, 2000), Koons (1992), Simmons (1993), and Glanzberg (2001, 2004).

Any approach to paradox that treats what is uttered on line 1 as semantically problematic, and what is uttered on line 2 as simply true, obviously requires an

asymmetry in the semantics of lines 1 and 2—an asymmetry which traces somehow to the circumstances of their production, since they appear to be tokens of the same semantic type (perhaps up to hidden indexical elements). What is intriguing is that once one adopts a procedural semantics, as I've claimed we must, we will have independently motivated exploitable compositional asymmetries between lines 1 and 2 of a very special kind.

To understand why, it is helpful to return to our simple case of defaulting. Consider a more stilted sentence that Sid might have uttered.

(13) Everything Sid says throughout March 15th will be true.

Suppose Sid utters (13) on March 15th, a day on which he additionally produces only some respectable number of uncontroversially true statements. Then Sid's utterance, u_s, seems true for now familiar reasons. Suppose, at some point, Pia shares in Sid's assessment and utters (13) as well. Again, if the circumstances are as I mentioned, this new utterance u_p will count as true.

Sid's utterance u_s and Pia's u_p appear to be tokens of the same sentence type. They also have the same truth values. But what is interesting about them is that, from a compositional perspective, the truth of each utterance has different grounds. Sid's utterance requires a defaulting mechanism to engage. Were it not for the conventions of English, Sid's utterance might have been defective. But Pia's utterance requires no such mechanism. It does not ascribe a truth-value to itself in the way that Sid's utterance does in the circumstance I have described, and because of this, it can be compositionally sensitive to a broader distribution of truth-values than Sid's utterance. Sid's utterance u_s *must* not require information about itself while its truth-value is being settled if it is to be conventionally truth-valued—it must not be compositionally semantically self-dependent. However Pia's utterance may, should, and seemingly *is* compositionally responsive to that very same utterance u_s (just consider if Sid had said different things at that moment). So the operation of the truth-predicate's semantic value in these two utterances is different: it responds to different bodies of semantic information, and does so in different ways. Put in more technical terms: the two utterances have different compositional semantic dependences, despite being two tokens of the same type. And as such, a procedural semantics modeling their behavior 'makes room' for assignment rules for "true" that distinguish between the utterances and assign them different truth-values. Of course, a language which associated "true" with a different procedural semantic value on which u_s and u_i receive distinct truth-value assignments would be bizarre, and entirely unlike English. And it does not seem like it would have "true" express a viable truth-concept. What is important is that the shift to a procedural semantics independently opens up space for a difference in truth-value allotments to the two tokens of the same type, owing to the semantics of the word "true".

For related reasons, we have empirical evidence that English already associates the word "true" with a semantic value of a kind that could permit the tokens on lines 1 and 2 to bear different truth-values. These utterances, too,

are naturally treated as having a kind of asymmetry in the structure of their compositional semantic dependences: the first has reflexive dependences, while the second does not. A procedural theory may associate "true" with rules of evaluation along compositional dependences that end up treating utterances with unresolvable reflexive dependences like line 1 as less than true, with the non-reflexive utterance of line 2 as being true, and responsive to this defect.

But we need to be careful about the nature of the differential assignment that is opened up by a procedural theory. The potential sensitivity is, again, *not* a form of context sensitivity like that found in ordinary indexical pronouns or gradable adjectives. Up to the fact that they may mediate differential assignments of truth-values to tokens of the same type, the two kinds of sensitivity have little in common.

For compositional purposes, we can think of an indexical pronoun like "I", or a gradable adjective like "tall' as bearing an extension relative to a context-index pair: the speaker of a context, or the set of persons tall at the world of the index, relative to a reference class made salient in context. Shifting the context of utterance shifts the extension assignment, in turn shifting about the object or property one uses the relevant word to speak about. Many context-sensitive approaches to the liar, like Burge (1979), Koons (1992), and Simmons (1993) draw on a roughly analogous understanding in their hypothesized sensitivity of the truth-predicate. The approach seems to lead, by this analogy, to a fragmentation of the concept of truth, which many regard as a noteworthy cost of the approach.

We've already had occasion in §5 to appreciate that the essential work procedural semantic values accomplish should not—indeed cannot—be thought of on this model. Sensitivity in an extension to a context of utterance is of no help in avoiding compositional circularities. As noted in §5, no theory that allows for our benign reflexivities can forgo mechanisms that conventionally restrict the compositional semantic dependences of an utterance to exclude some items to which semantic properties in the utterance are significantly applied. Standard context-sensitive theories do not conventionally restrict dependences in that way. At best, they restrict the utterances to which the semantic properties are applied, or shift about the semantic property reported, thereby changing what is said so that the desired semantic reflexivity is simply no longer there.[37]

This distinction between familiar forms of context-sensitivity and what we may call the 'token-sensitivity' of procedural semantic values is connected with an important lesson that was stated at the end of §3: procedural values, and the data that motivates them, require us to separate out a property (or set of objects) we are talking about, from the compositional semantic value that forms part of the model of our linguistic competence in reporting that property. This separation, though unusual, is ultimately what allows us to maintain the ability to report the property, or the set of objects bearing that property, when an ex-

[37]Gaifman, who as noted in §5 operates within a truth-proceduralist framework, early on stressed differences like those I've been harping on between the 'token-sensitivity' of truth-proceduralist frameworks and ordinary forms of context sensitivity. See, for example, the extended discussions in Gaifman (1992) §3, §5.

tension does not adequately serve in a model of how speakers could productively speak about it.[38]

Data from virtuous semantic circularity thus intriguingly provides us with motivations for a special kind of semantic framework that accommodates vicious circularities like the liar, while retaining some of the needed expressive power to characterize their semantic status through a special kind of semantic sensitivity of types to their tokening. More specifically, examining the question of what relevance the liar has to compositional semantic theorizing surprisingly gives us *prima facie* motivation for a very special class of approaches to the semantic paradoxes advocated so far, to my knowledge, only by Haim Gaifman.

But I want to stress that, of course, the brief remarks I've made here hardly constitute anything like a proposed *resolution* of the liar paradox. I am not certain that even the highly developed truth-proceduralist frameworks of Gaifman (1992, 2000) provide us with a full account of this kind. Rather, these frameworks, along with the arguments I've given here, are the beginning of a conversation about how such resolutions might be developed. The main lesson we've learned is that a special class of approaches to paradox, with a highly unusual form of expressive flexibility, actually only make use of linguistic resources in understanding the semantic values of semantic terms like "true" that every adequate compositional theory must accommodate.

I have hopes that a broadly truth-proceduralist framework may provide us with the best resources to understand virtuous and vicious semantic circularity alike. But even if this hope is ill-founded, the goals of foundational and compositional semantic theorizing ensure that procedural semantic values will become an integral part of our final story about what it is to understand a language that is able to talk about its own semantic properties.

References

P. Anand & V. Hacquard (2013). 'Epistemics and Attitudes'. *Semantics and Pragmatics* **6**(8):1–59.

J. Beall (2009). *Spandrels of Truth*. Oxford University Press.

T. Burge (1979). 'Semantical Paradox'. *The Journal of Philosophy* **76**:169–198.

[38] In the most recent list of theorists whose views were contrasted with truth-proceduralist approaches, Glanzberg, and Parsons who he draws on, were omitted. This is because Glanzberg's theory and perhaps Parsons', unlike those others, may share one of the two virtues of the truth-proceduralist views I've just described. In particular, the 'extraordinary' kind of context-sensitivity Glanzberg posits in quantifier domains to account for shifts in truth-value allotments to utterances containing semantic predicates may allow Glanzberg to avoid fragmenting the concept of truth. Or, at least, it is much less clear that Glanzberg is somehow committed to fragmenting truth in his theory. Still, a sensitivity in quantifier domains, even of the special type Glanzberg posits, is unhelpful in accounting for the benign circularities I've made my focus. From that perspective, Glanzberg's theory, like the others, is not poised to account for the data without taking on the shift to a procedural semantics.

A. Burgess (2011). 'Mainstream Semantics + Deflationary Truth'. *Linguistics and Philosophy* **34**(5):397–410.

C. Chihara (1979). 'The Semantic Paradoxes: A Diagnostic Investigation'. *Philosophical Review* **88**(4):590–618.

N. Chomsky (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

N. Chomsky (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Praeger, New York.

J. Collins (2002). 'Truth or Meaning? A Question of Priority'. *Philosophy and Phenomenological Research* **65**(3):497–536.

C. Dorr & J. Hawthorne (2013). 'Embedding Epistemic Modals'. *Mind* **122**(488):867–913.

M. Dummett (1959). 'Truth'. *Proceedings of the Aristotelian Society* **59**(1):141–62.

M. Eklund (2002). 'Inconsistent Languages'. *Philosophy and Phenomenological Research* **64**(2):251–75.

H. Field (2008). *Saving Truth From Paradox*. Oxford University Press.

H. Gaifman (1992). 'Pointers to Truth'. *The Journal of Philosophy* **89**(5):223–261.

H. Gaifman (2000). 'Pointers to Propositions'. In A. Gupta & A. Chapuis (eds.), *Circularity, Definition and Truth*, pp. 79–121. Indian Council Of Philosophical Research.

M. Glanzberg (2001). 'The Liar in Context'. *Philosophical Studies* **103**:217–51.

M. Glanzberg (2004). 'A Contextual-Hierarchical Approach to Truth and the Liar Paradox'. *Journal of Philosophical Logic* **33**:27–88.

A. Gupta (1997). 'Definition and Revision: A Response to McGee and Martin'. *Philosophical Issues* **8**:419–443.

A. Gupta & N. Belnap (1993). *The Revision Theory of Truth*. MIT Press, Cambridge, MA.

V. Hacquard (2006). *Aspects of Modality*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

V. Hacquard (2010). 'On the event relativity of modal auxiliaries'. *Natural Language Semantics* **18**(1):79.

V. Hacquard & A. Wellwood (2012). 'Embedding Epistemic Modals in English: a Corpus-Based Study'. *Semantics and Pragmatics* **5**:1–29.

V. Halbach (2011). *Axiomatic Theories of Truth*. Cambridge University Press.

I. Heim & A. Kratzer (1998). *Semantics in Generative Grammar*. Blackwell.

C. Horisk, et al. (2000). 'Deflationism, Meaning and Truth-Conditions'. *Philosophical Studies* **101**(1):1–28.

R. C. Koons (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge University Press.

M. Kremer (1988). 'Kripke and the Logic of Truth'. *Journal of Philosophical Logic* **17**(3):225–278.

S. Kripke (1975). 'Outline of a Theory of Truth'. *Journal of Philosophy* **72**(19):690–716.

H. Leitgeb (2005). 'What Truth Depends on'. *Journal of Philosophical Logic* **34**(2):155–192.

E. Lepore & K. Ludwig (2007). *Donald Davidson's Truth-Theoretic Semantics*. Clarendon Press.

D. Lewis (1970). 'General Semantics'. *Synthese* **22**(1-2):18–67.

W. G. Lycan (2012). 'A Truth Predicate in the Object Language'. In G. Preyer (ed.), *Donald Davidson on Truth, Meaning, and the Mental*. Oxford University Press.

J. MacFarlane (2011). 'Epistemic Modals are Assessment Sensitive'. In A. Egan & B. Weatherson (eds.), *Epistemic Modality*, pp. 144–178. Oxford University Press, New York.

T. Maudlin (2004). *Truth and Paradox: Solving the Riddles*. Oxford University Press, Oxford.

V. McGee (1991). *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Hackett Publishing Co., Cambridge.

F. Moltmann (forthcoming). "Truth Predicates in Natural Language'. In D. Achourioti, H. Galinon, & J. Martinez (eds.), *Unifying Theories of Truth*. Springer.

S. Moss (forthcoming). 'On the Semantics and Pragmatics of Epistemic Vocabulary'. *Semantics and Pragmatics* .

C. Parsons (1974). 'The Liar Paradox'. *Journal of Philosophical Logic* **3**(4):381–412.

D. Patterson (2005). 'Deflationism and the Truth Conditional Theory of Meaning'. *Philosophical Studies* **124**(3):271–294.

D. Patterson (2007). 'Understanding the Liar'. In J. C. Beall (ed.), *Revenge of the Liar: New Essays on the Paradox*, p. 197. Oxford University Press.

D. Patterson (2009). 'Inconsistency Theories of Semantic Paradox'. *Philosophy and Phenomenological Research* **79**(2):387–422.

G. Priest (2006). *In Contradiction: A Study of the Transconsistent, 2nd ed.* Oxford University Press, Oxford.

K. Scharp (2013). 'Truth, the Liar, and Relativism'. *Philosophical Review* **122**(3):427–510.

J. R. Shaw (2014). 'What is a Truth-Value Gap?'. *Linguistics and Philosophy* **37**(6):50334.

J. R. Shaw (forthcoming). 'Anomaly and Quantification'. *Noûs* .

K. Simmons (1993). *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge University Press, Cambridge.

R. Stalnaker (1987). 'Semantics for Belief'. *Philosophical Topics* **15**(1):177–190.

K. von Fintel & A. Gillies (2007). 'An Opinionated Guide to Epistemic Modality'. In *Oxford Studies in Epistemology 2*. Oxford.

K. von Fintel & A. Gillies (2008). 'CIA Leaks'. *Philosophical Review* **177**(1):77–98.

M. Willer (2013). 'Dynamics of Epistemic Modality'. *Philosophical Review* **122**(1):45–92.

G. Winskel (1993). *The Formal Semantics of Programming Languages*. MIT Press, Cambridge, MA.

S. Yablo (1982). 'Grounding, Dependence, and Paradox'. *Journal of Philosophical Logic* **11**:117–137.

S. Yalcin (2007). 'Epistemic Modals'. *Mind* **116**(464):983–1026.

S. Yalcin (2010). 'Probability Operators'. *Philosophy Compass* **5/11**:916–937.

S. Yalcin (2011). 'Nonfactualism about Epistemic Modality'. In A. Egan & B. Weatherson (eds.), *Epistemic Modals*, pp. 295–332. Oxford University Press, New York.

S. Yalcin (2012). 'Bayesian Expressivism'. *Proceedings of the Aristotelian Society* **112**(2):123–160.

S. Yalcin (forthcoming). 'Semantics and Metasemantics in the Context of Generative Grammar'. In A. Burgess & B. Sherman (eds.), *Metasemantics: New Essays on the Foundations of Meaning*. Oxford University Press, Oxford.