A Procedural Semantics for Truth*

James R. Shaw

What contribution does a word like "true" make to the content of whole utterances containing it—what is its semantic value? I have tried to argue elsewhere that the best answer to this question is one which takes the semantic value of "true" and related semantic terms to be a kind of rule or procedure which pairs utterances with assertoric content. This supposition has the potential to ground fully general foundational work in the philosophy of language while explaining some very peculiar features of the truth predicate's compositional behavior.[1] I also believe this idea can be *part* of a larger strategy for diagnosing and resolving the semantic paradoxes. I don't want to discuss these points here. Instead, my goal is simply to give a clearer, and more systematic, account of the particular kind of procedure I think we should associate with the truth predicate as its meaning, and to explore the formal consequences of doing so.

Many aspects of the formalism are not entirely novel: the appeal to a special notion of 'semantic dependence', the use of supervaluations, and the accommodation of a kind of token-sensitivity for uses of the truth predicate, for example. What is novel about the semantics is the combination of these features, and the unique issues that their combination raises. I'll begin by laying out the core elements of the theory in §§1–3. These elements generate the need for a special analog of a consistency proof—a kind of coherence result. I'll prove this in §4 and explore the salient features of the resulting semantics in §§5–6.

1 Preliminaries

My main project is a familiar one: to take an interpreted formal language capable of self-reference and extend it to a new interpreted formal language capable of discussing its own semantic properties. One of my guiding principles is to mind special structural features of natural language use that may play an ancillary role in this task. Because of this, I'll build three special features into my 'base' interpreted languages, which I take to be relevant to semantic self-description in natural language.

First, for both empirical and theoretical reasons, I believe that utterances of

*In formulating the ideas of this paper, I've benefited tremendously from helpful discussions with Sharon Berry, Jon Litland, Bernhard Nickel, and especially Warren Goldfarb and Peter Koellner.

[1]See Shaw (forthcomingb) for a discussion.

syntactically well-formed sentences of a language, consisting of only meaningful terms, may nonetheless be semantically defective in the sense that they fail to be truth-evaluable.[2] In part because of the paradoxes, I take utterances of sentences containing the truth predicate to be no exception. Consequently, I'll be working with trivalent models for a language governed by the weak Kleene tables for connectives.[3] Failure of truth-evaluability marks a special kind of semantic *defect*, but the nature of this defect is largely irrelevant for present modeling purposes.

Second, in connection with the assumption of trivalence, I make use of binary generalized quantifiers.[4] Aside from the strong empirical grounds for this treatment of natural language quantification, generalized quantifiers with a restrictor are indispensable in properly interpreting quantified statements in the trivalent setting. As is well known, unrestricted quantifiers in trivalent models can be unnaturally prone to take on the third, defective, truth value because of their broad range of quantification. These considerations are relevant to the ability of a language to appropriately state *generalities* about its own semantic structure—a well-known stumbling block for trivalent theories of truth. A goal of the ensuing formalism is to skirt as many such difficulties as possible.

Third, I take it that utterances, and not sentences, are the best candidates for truth-evaluability.[5] Moreover, I believe that taking sentence types as bearers of truth not only cripples a proper representation of the compositional semantics for terms like "true", but must be avoided to get a semantics which skirts paradox while salvaging the greatest amount of semantic expressive power.

My motivations for this last view, which I can only give here in the barest outlines are as follows. As already alluded to, I hold a familiar view that paradoxical utterances bear a special kind of semantic defect, resulting in truth-valuelessness. Additionally, though, I believe that distinct utterances of the same sentence type without any normally context-sensitive expressions differ as to whether or not they witness paradox, and hence the relevant semantic defect. The basic idea here is familiar from both Gaifman (1992, 2000) and Glanzberg (2001) as regards what Gaifman calls the "two-line paradox". Here is the version of relevance to semantic defect:

> Jones at t_0: "What Jones utters at t_0 is false or defective."
>
> Jane at t_1: "What Jones utters at t_0 is false or defective."

Jones and Jane utter the same sentence, and Jones' utterance is paradoxical for familiar reasons. If one thinks that paradoxical utterances are defective, and

[2] I think a strong empirical case can be made for such defects from consideration of semantic anomaly. See Shaw (forthcominga) for a discussion. My foundational grounds for countenancing semantic defect arise from more controversial views about the nature of the speech act of assertion. See Shaw (forthcomingc).

[3] The choice of the weak Kleene tables is based partly on considerations of simplicity, and there will be no special bar in the discussion to follow in integrating other schemes for connectives and quantifiers.

[4] See Glanzberg (2006) for a helpful survey of the topic.

[5] This is compatible with the claim that utterances might bear their truth values derivatively, for example through their relation to more 'primary' truth-bearer such as a proposition.

therefore not truth-evaluable, then Jones' utterance will be truth-valueless. If we hold firm to this claim, there seems to be something correct about what Jane is saying, even though it is a repetition of Jones' uttered sentence. For certain special reasons, I think we can and should take Jane's utterance to be true, pure and simple, while coherently maintaining that Jones' utterance is not. The reasons for this have to do with the meanings I think semantic words like "true" and "false" have, and the nature of semantic defect and how it is reported. I won't be able to enter into most details of this story here, though a crucial component of the story about the meaning of semantic words is the idea I've already alluded to: that they are given by special kinds of procedures. Some of the evidence for, and benefits from, this idea are found elsewhere.[6] My purpose here, though, is to formalize the relevant procedures and show them to be coherent, *given* separate motivations for my preferred treatment of paradox.

It's worth mentioning that although I take a resolution to the paradoxes to involve sensitivities witnessed at the level of utterances in the way above, this sensitivity is *not* an instance of normal modes of context sensitivity.[7] For this reason, some features of a standard Kaplanian representation of the involvement of context will be unsuitable for my purposes. In particular, my system will have more lax constraints on how truth-values could be distributed among multiple tokenings of the same sentence type. The nature of this token-sensitive allotment of truth-values, and how it differs from some normal modes of context sensitivity, will hopefully become a little clearer as the system is developed.

With these preliminaries out of the way, let me now add to some familiar definitions that together will characterize the notion of a 'base' interpreted language. I'll take languages, terms, and formulas to be defined in the usual way for a first-order language containing negation and conjunction as its sole connectives, but adjusted to treat the syntax of the universal and existential quantifiers as binary generalized quantifiers, as per my second preparatory point above. The sentences of such a language can be used to model the properties of natural language sentence *types*. But, as per my third preparatory point above, we will also need a formal representation of sentence tokens, which will eventually be the proper bearers of truth-values. To accomplish this, I'll use an indexed sentence type to represent a tokening of that type, allowing multiple indices for the same sentence.

Definition 1.1. Let $c_1, c_2, c_3, \ldots$ be a countable set of *tokening parameters*.

Definition 1.2. An *utterance* U is a sentence ϕ such that it, and each of its sub-sentences, is indexed by a distinct tokening parameter.

When not subscripted, c, c', etc. are used metalinguistically to talk about utterances' tokening parameters schematically. So, for example, ϕ_c schematically

[6] Again in Shaw (forthcomingb).

[7] In this way I depart from Glanzberg's response to the two-line paradox. Glanzberg treats the sensitivity here as an 'extraordinary' kind, but one which still is modeled through shifting extension assignments to "true". Gaifman's approach to paradox and mine, by contrast, are very closely related, but there are important differences which are in part the occasion for this paper. See p.9 for a brief discussion.

denotes an utterance of the sentence ϕ indexed by some tokening parameter c.

We can model a series of tokens of a single sentence type with a set of utterances of the sentence, in my defined sense, each indexed with distinct tokening parameters. It may help to think of a tokening parameter as something like a spacio-temporal location where the sentence type is tokened. Tokening parameters can also be thought of as labeling truth-bearers. Not only do tokens of whole sentences bear truth-values, but so too do their constituent tokened sub-sentences. This is why we index these as well.

Example 1.1. The following are utterances of a language $\mathcal{L}$ containing a binary relation symbol R, a unary relation symbol P and constants including a and b.

$$Rab_{c_3}$$

$$(\forall v_1 : v_1 = a)((\exists v_2 : Rv_1v_2)(Rv_2a))_{c_5}$$

$$(\neg(Pa_{c_1} \wedge Pb_{c_2})_{c_3})_{c_4}$$

$$(\forall v_2 : Pv_2)((Pv_2 \wedge Rba_{c_2}))_{c_4}$$

$$((\exists v_1 : (v_1 = v_1))((Rv_1a)_{c_5} \wedge (\exists v_1 : (v_1 = v_1))(Rav_1)_{c_3})_{c_4}$$

If we model a set of actual sentence tokens with a set of utterances in the way I am suggesting, by ensuring one utterance per sentence token that receives a truth-value assignment, then the resulting set of utterances will have the following simple set of formal properties.

Definition 1.3. A set of *utterances* $\mathcal{U}$ is *natural* if

(i) no tokening parameter c indexes more than one sentence type among the utterances of $\mathcal{U}$,

(ii) if $\phi_c \in \mathcal{U}$, and $\psi_{c'}$ is an utterance that is a proper part of ϕ_c, then $\psi_{c'} \in \mathcal{U}$.

(iii) if ϕ_c occurs in two utterances $\psi_{c'} \neq \theta_{c''}$, then either $\psi_{c'}$ is a constituent utterance of $\theta_{c''}$ or $\theta_{c''}$ is a constituent utterance of $\psi_{c'}$.

(i) ensures every utterance is an utterance of exactly one sentence type, and therewith that a tokening parameter c always picks out at most one utterance in $\mathcal{U}$. (ii) ensures every uttered sub-sentence of an uttered sentence is taken into account. This will be important since these tokens are assigned truth values, and we will want truth-value assignments to parts and wholes to be appropriately related. Finally, (iii) ensures that no single utterance is a component part of two distinct utterances which would occur at different places and times.

Example 1.2. The following is an example of a natural set of utterances for the language recently considered.

$$
\begin{aligned}
\mathcal{U} = \{ & (Pa_{c_1} \wedge ((\neg Pb_{c_2})_{c_3} \wedge Rba_{c_4})_{c_5})_{c_6}, \\
& (Rba_{c_7} \wedge Raa_{c_8})_{c_9} \\
& Pa_{c_1}, \\
& Pb_{c_2}, \\
& (\neg Pb_{c_2})_{c_3}, \\
& Rba_{c_4}, \\
& ((\neg Pb_{c_2})_{c_3} \wedge Rba_{c_4})_{c_5} \\
& Rba_{c_7} \\
& Raa_{c_8} \}
\end{aligned}
$$

A natural set of utterances, in my sense, is designed to model concrete, actualized, uttered sentence tokens in some possible situation. Though utterances are the proper bearers of truth values in my system, I will assume that my 'base' interpreted language is one which exhibits no kind of sensitivity to the allotment of truth values at the level of sentence tokens, whether due to standard forms of context sensitivity or otherwise. Consequently, we can begin by characterizing more or less familiar trivalent models for the language.

As such a *base $\mathcal{L}$-model* $\mathcal{M}$ (often just a 'model') supplies a universe of discourse $M^{\mathcal{M}}$, and a partial interpretation function $I^{\mathcal{M}}$ mapping predicates symbols R to disjoint extension/anti-extension pairs $\langle R_t^{\mathcal{M}}, R_f^{\mathcal{M}} \rangle$. For simplicity I'll assume functions symbols are always assigned total functions as interpretations. A peculiarity of my base models is that I will stipulate that they contain information about 'which utterances exist', that is, which natural set of utterances $\mathcal{U}^{\mathcal{M}}$ of the language is available for the allotment of truth-values.

Note that since I allow base models to specify a partial interpretation function, the recursive clauses giving the interpretation of a sentence (or formula at a variable assignment) may leave it *undefined*. Being undefined is different from being assigned the third, defective, truth-value u. So base models may effect a *four-fold* division of sentences (into t, f, u and undefined).

As alluded to before, I'll adopt a weak Kleene scheme for connectives. We face a choice point here for how to treat generalized quantifiers in the trivalent setting. As I noted earlier, my motivations for using generalized quantifiers are to help avoid over-inheritance of defective status. Accordingly, I treat the truth-values of quantified statements as always ascertained relative to the objects satisfying the quantifier restrictor. Thus the restrictor never contributes to defective status—it only ever contributes a set of objects to which further predications are made by a quantifier matrix. How to set the contribution of quantifier matrices is something on which I have no particular commitments. Here I'll implement a scheme on which quantified statements are defective only when all objects satisfying the quantifier restrictor make the matrix defective though, as always, other schemes can be accommodated as well. So the clauses for quantifiers will be as follows:

$$[\![(\forall v:\phi)(\psi)]\!]^{\mathcal{M},g} = \begin{cases} t & \text{if } \{a \mid [\![\phi]\!]^{\mathcal{M},g[v\to a]} = t\} \subseteq \\ & \qquad \{a \mid [\![\psi]\!]^{\mathcal{M},g[v\to a]} = t\} \\ u & \text{if } \{a \mid [\![\phi]\!]^{\mathcal{M},g[v\to a]} = t\} \subseteq \\ & \qquad \{a \mid [\![\psi]\!]^{\mathcal{M},g[v\to a]} = u\} \\ f & \text{otherwise} \end{cases}$$

$$[\![(\exists v:\phi)(\psi)]\!]^{\mathcal{M},g} = \begin{cases} t & \text{if } \{a \mid [\![\phi]\!]^{\mathcal{M},g[v\to a]} = t\} \cap \\ & \qquad \{a \mid [\![\psi]\!]^{\mathcal{M},g[v\to a]} = t\} \neq \emptyset \\ u & \text{if } \{a \mid [\![\phi]\!]^{\mathcal{M},g[v\to a]} = t\} \subseteq \\ & \qquad \{a \mid [\![\psi]\!]^{\mathcal{M},g[v\to a]} = u\} \\ f & \text{otherwise} \end{cases}$$

These recursive clauses allow us to define the denotations of a sentence ϕ in a model, $[\![\phi]\!]^{\mathcal{M}}$, in the usual way. I have stressed that sentence tokens and not sentence types are the proper bearers of truth. But since our base interpreted languages were stipulated to exhibit no context sensitivity of any kind, we can consider the truth-value assignment to sentence tokens to be inherited from our provisional assignment of truth-values to types.

Definition 1.4. If $\phi_c \in \mathcal{U}$, then $[\![\phi_c]\!]^{\mathcal{M}} = [\![\phi]\!]^{\mathcal{M}}$.

2 Semantic Dependence

My task begins in the familiar way, with a base model of the sort just specified which contains uninterpreted semantic vocabulary. In particular, I'll work with a base model $\mathcal{M}$ for a language $\mathcal{L}$ containing T (for "truth") and U (for "defectiveness") as unary predicates. The model's interpretation function $\mathcal{I}^{\mathcal{M}}$ is undefined only at T and U. It will be helpful to have a name for the uninterpreted vocabulary.

Definition 2.1. T and U are *semantic predicates.*

By definition our base model $\mathcal{M}$ specifies a natural set of utterances $\mathcal{U}^{\mathcal{M}}$ of sentences from $\mathcal{L}$ in the universe of discourse of the model $M^{\mathcal{M}}$. We can assume that the model has some modes, perhaps stipulative ones, for referring to the utterances in $\mathcal{U}^{\mathcal{M}}$, and that these modes are sufficient for generating the kinds of self-reference required for paradox. To help represent this capacity, I'll make a slightly non-standard use of corner quotes: $\ulcorner\phi_c\urcorner$ will be used schematically to represent any term whose denotation in $\mathcal{M}$ is the utterance ϕ_c.

The goal of this paper is to show how truth-values are allotted among the utterances in $\mathcal{U}^{\mathcal{M}}$, where the difficulty in this task is to say how utterances containing semantic predicates like T and U get the truth values they have. As stressed before, in this task I take the 'meaning' of the terms T and U to be given by a kind of *procedure*. In this section, I will to lay out some of the technical apparatus needed to characterize the procedure, and how it generates a truth truth-value assignment to $\mathcal{U}^{\mathcal{M}}$.

But before I begin, I'd like to give an informal characterization of what it means to take the meaning of the truth predicate to be a kind of procedure, and what procedure I have in mind. Consider a very simple case of three utterances (u_a)–(u_c), two of which use the truth predicate.

> *Alice*: (u_a) I'm hungry.
>
> *Bert*: (u_b) What Alice just said is true.
>
> *Charles*: (u_c) What Bert just said is true.

If called on to assess the truth-value of Charles' (u_c), speakers will exhibit a natural and familiar pattern of reasoning: first they will find out 'what Bert just said'—minimally what utterance he produced. Then they will, on the basis of (u_b) find out what Alice said. They will then check Alice's utterance 'against the facts', use this information to establish the truth or falsity of (u_a), then use that to assess (u_b), and finally use *that* information to assess (u_c).

This simple and familiar example brings to light a vague notion that I would eventually like to render more precise—that of *semantic dependence*. Because (u_c) attributes a truth-value to (u_b), what truth value (u_c) bears seems to depend on what truth value (u_b) bears. Likewise for (u_b) as regards (u_a). This dependence relation is important for understanding how speakers assess the truth-values of utterances containing semantic terms because it seems that *speakers assess utterances along the chain of semantic dependences in reverse order*.

My motivating idea is this: that the meaning of the truth predicate *is* the procedure revealed by the patterns of reasoning speakers exhibit in cases like the simple one I just gave (although the details of the procedure obviously can become more complicated than the simple case shows). When speakers learn how to use the word 'true', they are learning a method for associating truth-values with utterances which are situated in a hierarchy of semantic dependences. And the method *determines* what truth values utterances actually bear (as opposed, for example, to determining an extension for the word "true" to have in a final model). Consequently, to say which truth values utterances in $\mathcal{U}^{\mathcal{M}}$ bear, my goal is twofold:

(I) to formalize the semantic dependence relation, and

(II) to formalize the method speakers use to assign truth values along chains of semantic dependences.

Let's begin with the first task.

Several different specifications of the vague notion of semantic dependence that can be found in the literature on truth. It is implicit in the notion of *groundedness* in Kripke (1975), and is treated more directly in Yablo (1982), Gaifman (1992), and Maudlin (2004) among others. There is plausibly a 'family' of equally legitimate ways of sharpening the informal notion of semantic dependence. Some previous characterizations of semantic dependence, though, will not work for my particular purposes. I am interested in a relation of semantic dependence for instrumental reasons: to model the order in which speakers are able to assign truth-values to utterances. Accordingly, a first informal gloss on the relation of semantic dependence that I'm after might be as follows:

> An utterance ϕ_c semantically depends on an utterance ψ_c if speakers require information about the truth-value of ψ_c to settle the truth-value of ϕ_c.

Some previous characterizations of semantic dependence won't be able to play the role glossed above because they are given in broadly syntactic terms (perhaps with information about the denotations of terms). This can often lead to a dependence relation which is too permissive. For example, it can force quantified statements like (1) to depend on far too many other truth-bearers—sometimes all of them.

(1) Everything Kate said is true.

Some of the problems here can be avoided by the incorporation of generalized quantifiers, and using information in quantifier restrictors to restrict dependences. But there are more complex cases which make executing this strategy more difficult than it might first seem. Consider for example (2).

(2) Every true statement of Kate's yesterday was believed by Tom.

To ascertain the truth of an utterance of (2) what information do speakers need to have? Information about *everything* Kate said? Not always. The answer seems to depends on how much of what Kate said is true, perhaps along with other facts. Consider a scenario in which we are able to ascertain that there is at least one truth Kate spoke—that Tom is unreliable—which Tom does not believe. It seems like at this point we know enough in this scenario to fix the value of an utterance of (2): it is false. This means that in this scenario *relative* to the information about which utterances are true, we need fix no more truth-values of utterances to settle the status of (2).

A lesson I think we should take from examples like (2) is that a semantic dependence relation defined for my purposes should be *dynamic*: what a given dependence relation looks like should depend on what truth-values speakers have already assigned using prior dependence relations. This means there should be a two-way interaction between relations of semantic dependence and allotments of truth-values. Dependence relations help fix the order in which speakers allot truth-values, but as more truth-values are allotted, dependence relations may shift in response to the new information given by the allotments. A notable discussion of semantic dependence which gives the resources to skirt worries like those given above is found in Leitgeb (2005). Readers of Leitgeb will recognize that the techniques he employs are very similar to those I will shortly adopt.

Before getting into the formal details, a quick philosophical point is in order. I want to stress that the meaning of the word "true" is *exhausted by the procedure of allotting truth-values along the hierarchies of semantic dependences.*

Thus, even though the procedure I characterize will ultimately deliver a set of truths, this set is not the meaning of the word, for example as its extension. Though this is an important point of departure from, say, Kripke (1975) and many of those building on his system, this idea too is not completely novel. Gupta & Belnap (1993) have also defended the view that the truth predicate is associated with a rule-of-revision. My application of the idea that the meaning of the truth predicate is a procedure is, as will become clear, significantly different from Gupta and Belnap's. Philosophically, my proposal is much more similar in spirit to that of Gaifman (1992, 2000), who likewise characterizes allotments of truth-values along a dependence relation. At a formal level, my proposal differs from Gaifman's primarily in my use of a more sophisticated semantic dependence relation. This difference is not incidental: the complications arising from such a dependence relation are required, I believe, to make sense of semantic generalities, and those complications raise a special problem that does not arise when the semantic dependence relation is simplified. It is these complications that will require us to produce a special coherence result in §4.[8]

With this preamble out of the way, let me now begin to get some of the technical machinery I'll be using on the table. Recall that our task is to say how to take an allotment of truth-values to utterances of $\mathcal{U}^{\mathcal{M}}$ given by the base model $\mathcal{M}$ and exhibit the procedure speakers use in allotting truth-values to utterances containing the semantic terms T and U. To help understand my formalism, it might be useful to think of it as representing the stepwise progress of an individual, idealized reasoner in following this procedure. The idealized reasoner has all information about the non-semantic facts (given by $\mathcal{M}$) and will gradually reason to the truth values that various utterances in $\mathcal{U}^{\mathcal{M}}$ have stepwise.

A important notion in this process is that of a truth-value assignment.

Definition 2.2. A *truth-value assignment* is a partial function $A : M^{\mathcal{M}} \to \{t, f, u\}$.

A truth-value assignment can be thought of as a partial allotment of truth values to the elements of the universe of $\mathcal{M}$ and, therewith, to $\mathcal{U}^{\mathcal{M}}$.[9] It plays three roles in the ensuing formalism. First, certain truth-value assignments can be used represent the provisional progress our ideal reasoner has made in allotting truth values to utterances in $\mathcal{U}^{\mathcal{M}}$. Second, a particular total truth-value assignment—the culmination of the partial ones just alluded to—will represent the total distribution of truth values among the utterances of $\mathcal{U}^{\mathcal{M}}$. And our original goal can be rethought of in these terms: characterizing the latter 'master' truth-value

[8]Gaifman does briefly consider the use of certain kinds of supervaluations in his system (like those which factor into my dependence relation), but not those constrained in the ways I will suggest on p.12. It is these special constraints on supervaluations which both give my system its special capacity to state semantic generalities, and require of it the coherence result I mention here.

[9]Is it really acceptable to think of non-linguistic entities as bearers of any truth value—even my third defective value u, as the truth-value assignments I consider will do? Probably not, but accommodating this point requires a needless complication of the formalism.

assignment and its properties. This master assignment will be a total extension of an initial assignment gleaned from the base model $\mathcal{M}$ in the following sense.

Definition 2.3. A *total extension of a truth-value assignment assignment A* (sometimes just 'an extension') is a truth-value assignment A' such that $A \subseteq A'$ and A' is total.

So truth-value assignments represent the partial, and final stages of the procedure our ideal reasoner executes. But in this section, truth-value assignments will play a different, third role. They will be used to help characterize a notion of semantic dependence—the relation that determines the order in which our ideal reasoner proceeds in allotting truth values. The way assignments do this is by also acting as *hypothetical* allotments of truth-values. Permutations of such hypothetical allotments can give us information about semantic dependences, an idea again familiar from Leitgeb (2005).

To see how this works, consider the influence the semantic dependences of an utterance of a sentence like (1) have on its truth-value.

(1) Everything Kate said is true.

If we suppose that Kate said at least one falsehood, (1) comes out false. If we suppose instead that Kate speaks nothing but the truth, (1) will be true. But suppose we hold the truth-values of Kate's utterances *fixed* and consider suppositional distributions of truth-values among other utterances. Then of course (1)'s truth-value will stay constant under those permutations. These simple reflections afford us an intuitive way of homing in on an utterance's semantic dependences. The semantic dependences of an utterance ϕ_c are those utterances *whose truth-values 'influence' the truth-value of ϕ_c* where the notion of 'influence' can be brought out counterfactually: changes among only the truth-values borne by the influencing utterances can alter the truth-value of the utterance influenced.

It is important to distinguish two ways that utterances can relate to truth-values in this process of hypothetical reasoning. On the one hand an utterance can *bear* a truth-value, by hypothetical stipulation. On the other hand, that utterance can 'evaluate to' a truth-value based on the truth-value distribution hypothesized. Certain utterances which self-ascribe semantic properties may evaluate to truth-values different than those they bear on hypothetical stipulation. The liar is characteristic in this regard.

(L) This very utterance is false.

If we suppose an utterance λ_c of (L) *bears* the value of truth, it will *evaluate to* falsehood. If we suppose it *bears* the value of being false, it will *evaluate to* truth. Liars exhibit the importance of distinguishing borne and evaluated truth-values in ascertaining a semantic dependence relation. We want to represent the liar utterances as semantically dependent on themselves. But this kind of semantic self-dependence would not be possible to ascertain, on the procedure I am outlining, if truth-values borne by utterances aren't distinguished from the truth-values to which those utterances evaluate.

So, roughly, the first step in determining the semantic dependences of an utterance ϕ_c is as follows: Permute the values borne by utterances in $\mathcal{U}^{\mathcal{M}}$. If permuting the values borne by only some particular set of utterances results in a change in evaluation of ϕ_c, then we know the truth-value of ϕ_c may be 'influenced' by those utterances. It should be clear how truth-value assignments are helpful in formalizing this idea. Truth-value assignments can represent the hypothetical allotments of *borne* truth-values. Moreover, a truth-value assignment also contains the information relevant to *evaluating* utterances as well, since a truth-value assignment implicitly contains the means for interpreting the uninterpreted semantic vocabulary T and U in $\mathcal{M}$. Here is one way of 'reading off' a complete model from $\mathcal{M}$ and a truth-value assignment A.

Definition 2.4. The *model engendered (for $\mathcal{M}$) by a total truth-value assignment A*, $\mathcal{M}_A$, is that given by the extending the interpretation function of $\mathcal{M}$ with

(i) $T_t^{\mathcal{M}_A} = \{x : A(x) = t\}$

(ii) $T_f^{\mathcal{M}_A} = \{x : A(x) \neq t\}$

(iii) $U_t^{\mathcal{M}_A} = \{x : A(x) = u\}$

(iv) $U_f^{\mathcal{M}_A} = \{x : A(x) \neq u\}$

Clauses (ii) and (iv) present one option for how to construe the utterances to which T and U truth-evaluably apply. These conditions can plausibly be altered in various ways—for example by setting $T_f^{\mathcal{M}_A} = \{x : A(x) = f\}$. I have nothing against alternative proposals, and I'll have occasion to briefly discuss the importance of my particular choices in §5.

Now, the complete model engendered by an assignment itself implicitly generates another total truth-value assignment of its own.

Definition 2.5. The *truth-value assignment read off of a model $\mathcal{M}$*, noted $A_{\mathcal{M}}$ is given as follows

$$A_{\mathcal{M}}(o) = \begin{cases} t & \text{if } [\![o]\!]^{\mathcal{M}} = t \\ f & \text{if } [\![o]\!]^{\mathcal{M}} = f \\ u & \text{if } [\![o]\!]^{\mathcal{M}} = u \text{ or } o \notin \mathcal{U}^{\mathcal{M}} \\ \text{undefined} & \text{if } o \in \mathcal{U}^{\mathcal{M}} \text{ and } [\![o]\!]^{\mathcal{M}} \text{ is undefined} \end{cases}$$

This gives us the formal tools needed to distinguish between borne and evaluated truth-values on a hypothetical allotment. A hypothetical set of *borne* truth-values can be represented by a stipulated truth-value assignment A. And the truth-values to which utterances *evaluate* on that hypothetical distribution are given by the truth-value assignment read off the engendered model, i.e., $A_{\mathcal{M}_A}$. As I mentioned before, it is not always the case that $A_{\mathcal{M}_A} = A$. The presence of liar utterances will ensure equality does not hold. For example, if $A(\lambda_c) = t$ then $A_{\mathcal{M}_A}(\lambda_c) = f$.

When we look at hypothetical distributions of truth-values borne by utterances to ascertain semantic dependences, we have a choice as to how permissive we will be in the kinds of distributions we observe. Obviously, some distributions of truth-values are ones which could never obtain—for example, those which simply ignore all compositional relations between the truth-values of utterances and the truth-values of their uttered parts. We will want the truth-value distributions which we permute in determining semantic dependences at least to be 'well-behaved' with respect to logical vocabulary and the behavior of the predicates T and U. The choice has 'downstream' implications for how truth-values are allotted: Which distributions we permute will affect what our semantic dependence relations ultimately look like. Since the semantic dependence relations give the paths along which truth-values are allotted, any effects on the semantic dependence relations will in turn affect which truth-values utterances bear. As we will soon see, constraining the range of truth-value distributions relevant to the determination of semantic dependences is integral to capturing semantic *generalities*—true claims about the total distribution of truth-values that utterances actually bear.

As a first step in this direction, I'll be focusing on what I'll call 'coherent' truth-value assignments, which capture some broad features of any reasonable truth-value distribution (given our assumption of a weak Kleene scheme).

Definition 2.6. A truth-value assignment A is *coherent* if it meets the following conditions for utterances in its domain.

(i) If $\phi_c = (\neg\psi_{c'})_c$, then $A(\phi_c) = t$ iff $A(\psi_{c'}) = f$, and $A(\phi_c) = f$ iff $A(\psi_{c'}) = t$.

(ii) If $\phi_c = (\psi_{c'} \wedge \theta_{c''})_c$, then $A(\phi_c) = t$ iff $A(\psi_{c'}) = A(\theta_{c''}) = t$, and $A(\phi_c) = f$ iff either $A(\psi_{c'}) = f$ and $A(\theta_{c''}) \in \{t, f\}$ or $A(\psi_{c'}) \in \{t, f\}$ and $A(\theta_{c''}) = f$.

(iii) If $\phi_c = (T\ulcorner\psi_{c'}\urcorner)_c$ and $A(\phi_c) \in \{t, f\}$, then $A(\phi_c) = t$ iff $A(\psi_{c'}) = t$.

(iv) If $\phi_c = (U\ulcorner\psi_{c'}\urcorner)_c$ and $A(\phi_c) \in \{t, f\}$, then $A(\phi_c) = t$ iff $A(\psi_{c'}) = u$.

(v) If $A(\phi_c), A(\phi_{c'}) \in \{t, f\}$, then $A(\phi_c) = A(\phi_{c'})$

The first two conditions state that the truth-value assignment function respects the behavior of negation and conjunction. Conditions (iii) and (iv) state that the truth-value assignment commutes with T and U *allowing* for the kind of token sensitivity required of the proposed resolution of the two-line paradox. Recall that this allows that token ascriptions of the same type may diverge in truth-value. Moreover, on the proposal, some versions of the 'strengthened liar' come out defective, and hence truth-valueless. Condition (v) is the only condition that relates tokens of the same type. It states that no two non-defective tokens of the same type can receive different truth-value assignments. This condition is reasonable since base models exhibit no token-sensitivity. Consequently, the only cases of token-sensitivity which arise in the subsequent system should owe in part to some instance of semantic defect arising from paradox, or circularity more generally.

The absence of conditions on the quantifiers is due to the absence of a satisfaction predicate in the object language and a truth-value assignment which additionally ranged over sub-sentential utterances and variable assignment pairs. Adding these would increasingly complicate the formalism. The present system should, however, give a suitable indication of how the satisfaction predicate and the conditions on quantifiers might be added. I'll discuss these issues further in §6.

So, to home in on the semantic dependence relation, *one* restriction on the kinds of assignments we permute will be that we only look at coherent assignments in the above sense. But there is a further restriction to enforce in keeping with our guiding metaphor of the allotment of truth-values as one done by a hypothetical reasoner proceeding in stepwise fashion. At any given moment, such a reasoner may have already *fixed* the truth-values of a substantial set of utterances, and the reasoner will have no reason to think those truth-values could change. So when this reasoner looks for the semantic dependences of new utterances, it stands to reason that they should only look at new potential truth-value assignments which *extend* the partial assignment they have made so far.

So a first step in learning the semantic dependences of an utterance *relative to* a partial allotment of truth-values is to permute coherent truth-value assignments extending that partial allotment. This will tell us which utterances 'influence' the truth-value of the utterance we are considering. Here is a pair of definitions which, together, formalize the relevant notion of influence.

Definition 2.7. Let $\Gamma \subseteq M^{\mathcal{M}}$, ϕ_c be an utterance in $\mathcal{U}^{\mathcal{M}}$, and A be a truth-value assignment. We say Γ *matters to* ϕ_c *relative to* A just in case there are two total coherent truth-value assignments A_1 and A_2 extending A, such that

(i) $A_1(\gamma) \neq A_2(\gamma)$ if and only if $\gamma \in \Gamma$.

(ii) $A_{\mathcal{M}_{A_1}}(\phi_c) \neq A_{\mathcal{M}_{A_2}}(\phi_c)$.

Definition 2.8. Γ *really matters to* ϕ_c *relative to* A if there are two total coherent extensions A_1 and A_2 which both witness the fact that Γ matters to ϕ and are such that there are no assignments A_1', A_2' and $\emptyset \neq \Gamma' \subseteq \Gamma$ meeting these conditions:

(i) $A_1 \restriction (M - \Gamma) = A_1' \restriction (M - \Gamma) = A_2 \restriction (M - \Gamma) = A_2' \restriction (M - \Gamma)$

(ii) $A_1' \restriction (\Gamma') = A_2' \restriction (\Gamma')$

(iii) $A_{\mathcal{M}_{A_1'}}(\phi_c) \neq A_{\mathcal{M}_{A_2'}}(\phi_c)$. [10]

[10] Compare *mattering* and *really mattering* with the notions of *dependence* and *essential*

Intuitively, a set of utterances Γ matters to ϕ_c if one can permute the borne values of all and only the utterances in Γ to alter the evaluated truth-value of ϕ_c. Γ really matters to ϕ_c when some permuted assignments that show Γ matters to ϕ_c don't contain 'extraneous' elements (elements in a non-empty Γ' which could be held constant and still result in a change in value to ϕ_c). When looking for the utterances which can 'influence' the evaluated value of an utterance ϕ_c, it is important to consider only utterances which really matter to ϕ_c, since by definition any superset of a set Γ of utterances that matter to ϕ_c will also matter to ϕ_c.

We can use this formalized notion of influence as the backbone of a semantic dependence relation. Simply identifying the semantic dependences of ϕ_c with the utterances that really matter to it, though, isn't well suited to my purposes for a number of reasons. First, making that identification doesn't ensure that when one utterance, ϕ_c, semantically depends on $\psi_{c'}$, that utterances with ϕ_c as a constituent also semantically depend on $\psi_{c'}$. Second, it does not ensure that an unassigned utterance of the form $T\ulcorner\phi_c\urcorner_{c'}$ or $U\ulcorner\phi_c\urcorner_{c'}$ depends on ϕ_c, since the conditions on coherence may cause these utterances to covary in their truth-values.[11] I'll need the semantic dependence relation to exhibit those dependences, if only to simplify upcoming proofs. Finally, the definition of really mattering has a bug: sometimes many sets of utterances can matter to ϕ_c, but *no* sets of utterances really matter to it. These problem cases merit a special name.

Definition 2.9. An utterance ϕ_c *has unspecifiable dependences relative to A* if some Γ matters to ϕ_c relative to A but no Γ' really matters to ϕ_c relative to A.

I'll give an example of a ϕ_c of this kind shortly. What's important for now is that the semantic dependences of an utterance will be those that really matter

dependence defined in Leitgeb (2005). There are some superficial differences between the definitions. For example, Leitgeb defines dependence for sentences, as opposed to utterances. But there are also a few more noteworthy differences. For Leitgeb, when ϕ depends on Γ, any change in the evaluated truth-value of ϕ *must* be traceable *only* to sentences in Γ. By contrast, when Γ matters to ϕ_c it may yet be possible to find ways of changing the evaluated truth-value of ϕ_c by permuting the borne values of utterances outside Γ. This leads to further differences. For Leitgeb, the set of essential dependences of ϕ, if it exists, is unique. But several sets of utterances may really matter to a given utterance ϕ_c. For example, if ϕ_c is of the form $(T\ulcorner\psi_{c'}\urcorner \vee T\ulcorner\theta_{c''}\urcorner)_c$, then as will become clearer shortly, it is entirely possible that two sets $\{\psi_{c'}\}$ and $\{\theta_{c''}\}$ really matter to ϕ_c. Also my definition of really mattering is somewhat more complex than Leitgeb's definition of essential dependence. The added complexity leads to a more permissive relation, which will be necessary to prove Proposition 2.1. One source of these differences is that my techniques for tracking dependence relations were originally developed independently of Leitgeb's work and for slightly different reasons. Each relation has its advantages. Leitgeb's relations have some regular mathematical properties. For example, the sets of dependences of ϕ in Leitgeb's sense form a filter, though the example just given shows that the sets of utterances that matter to an utterance ϕ_c need not. My relations make it slightly more difficult for certain anomalies to appear—what I call utterances with unspecifiable dependences on p.14. At bottom, though, the relations are very similar in spirit and structure, and are helpfully thought of as such.

[11] Let $u_1 = \phi_c$, $u_2 = T\ulcorner\phi_c\urcorner$, $u_3 = T\ulcorner T\ulcorner\phi_c\urcorner\urcorner$, and let $A(\phi_c) = u$ but be undefined elsewhere. Then u_2 'should' matter to u_3 (for my purposes), but it doesn't. Any coherent extension of A can't assign t to u_2. This means on every such assignment u_3 evaluates to false.

to it, with some added clauses to account for these three problem cases: direct ascriptions of semantic properties, unspecifiable dependences, and inheritance of semantic dependences by logically complex constructions.

Definition 2.10. $\psi_{c'}$ is a *semantic dependence of an utterance* ϕ_c *relative to a truth-value assignment* A if

(i) $\psi_{c'}$ is in some Γ that really matters to ϕ_c relative to A,

(ii) $\phi_c = S\ulcorner\psi_{c'}\urcorner$ for some semantic predicate S, and $\psi_{c'} \notin \text{dom}(A)$,

(iii) $\psi_{c'} \notin$dom(A) and ϕ_c has unspecifiable dependences relative to A, or

(iv) ϕ_c has a constituent utterance meeting conditions (i), (ii), or (iii) with respect to $\psi_{c'}$.

The set of semantic dependences of ϕ_c relative to a truth-value assignment A is noted $\text{SD}_A(\phi_c)$.[12]

Note that clause (iii) ensures that if ϕ_c has unspecifiable dependences relative to A, it is semantically dependent on every unassigned utterance.

The resulting definition of semantic dependence has two related features: it is *relative* and has a kind of *epistemic* character. It is relative in the sense that which semantic dependences an utterance has depends on which other utterances have had their truth-values fixed. It is epistemic in the sense that the semantic dependences of an utterance ϕ_c are *not* just the (unassigned) utterances that ϕ_c ascribes semantic properties to. Rather, to improve on our earlier informal characterization of semantic dependence

> The semantic dependences of an utterance ϕ_c relative to A are those utterances whose truth-values a reasoner, who already knows only the truth-values of the utterances in dom(A), might still have to figure out in order to determine the truth-value of ϕ_c.

Both features of the definition—its relativity and its epistemic character—are best brought out by looking at a few, somewhat idealized examples, focusing on how semantic dependences are generated in the 'standard case' by the relation of really mattering. The examples are idealized because I'll provisionally be assuming the availability of certain kinds of coherent extensions, like the following.

Definition 2.11. A truth-value assignment A *allows arbitrary permutations over a set* Γ if A has coherent extensions, and for any coherent extension A' of A, we can find other coherent extensions identical with A' over $M - \Gamma$, and arbitrarily permuting the truth values assigned over Γ.

[12] A main motivation for accommodating condition (ii) is that the stipulation will assist in upcoming proofs in §5. I'm not actually sure if the condition is required to ensure those proofs go through. Also, a more complete treatment of semantic dependence to model the intricacies of natural language use of semantic terms would have to be weakened through the accommodation what I have elsewhere called 'defaulting' conditions. See Shaw (forthcomingb).

I'll say a little bit more about the availability of such extensions shortly.

Example 2.1. ϕ_c is $(\exists v_1 : \theta(v_1))(T(v_1))_c$, and $\psi_{c'}$ is $(\exists v_1 : \theta(v_1))(\neg T(v_1))_{c'}$ where θ is a formula without semantic vocabulary which, in $\mathcal{M}$, defines a set of utterances U_1. Consider three cases:

(a) Let A be undefined on U_1, and allow arbitrary permutations over U_1. Then $SD(\phi_c) = SD(\psi_{c'}) = U_1$. To see this, consider coherent extensions A_1 and A_2 of A which map every utterance in U_1 to f except for one $u_1 \in U_1$ such that $A_1(u_1) = t$ and $A_2(u_1) = f$. Let A_1 and A_2 be identical elsewhere. Then $A_{\mathcal{M}_{A_1}}(\phi_c) = t \neq A_{\mathcal{M}_{A_2}}(\phi_c) = f$, showing that $\{u_1\}$ matters to ϕ_c relative to A. Any singleton that matters to an utterance relative to a truth-value assignment also *really* matters to it by definition. So $u_1 \in SD_A(\phi_c)$. But there was nothing special about u_1—we could show the same for any utterance in U_1. And, given the available extensions of A, no other utterances besides those in U_1 really matter to ϕ_c. Any set of utterances Γ that matters to ϕ_c must contain some element of U_1, revealing that if Γ is not a singleton from U_1, it does not really matter to ϕ_c. So the set of utterances that really matter to ϕ_c relative to A are those in U_1, hence $SD_A(\phi_c) = U_1$. Analogous reasoning gives the same semantic dependences to $\psi_{c'}$.

(b) Let A be such that $A(u_1) = t$ for some $u_1 \in U_1$ but is elsewhere undefined on U_1, and A allows arbitrary permutations over $U_1 - \{u_1\}$. Then $SD(\phi_c) = \emptyset$ and $SD(\psi_{c'}) = U_1 - \{u_1\}$. $SD(\phi_c) = \emptyset$ since every total coherent extension A' of A will be such that $A'(u_1) = t$, which will ensure that $A_{\mathcal{M}_{A'}}(\phi_c) = t$. So no utterances matter to ϕ_c relative to A. On the other hand, permuting the values of $U_1 - \{u_1\}$ will alter the evaluated truth-value of $\psi_{c'}$ just as in the previous example.

(c) Let A be such that $A(u_1) = f$ for some $u_1 \in U_1$ but is elsewhere undefined on U_1, and A allows arbitrary permutations over $U_1 - \{u_1\}$. Then $SD(\phi_c) = U_1 - \{u_1\}$ and $SD(\psi_c) = \emptyset$. The case is symmetric with (b).

This example brings out both of the features I claimed were characteristic of my semantic dependence relations. The relations are relative, as can be seen from the fact that the semantic dependences of ϕ_c and $\psi_{c'}$ relative to A change depending on how many truth-values have been fixed in A. Moreover, the uniquely epistemic character of the semantic dependence relations comes out when contrasting cases (b) and (c) above. If we think of semantic dependence in terms of which utterances are ascribed truth-values, then we would expect the dependences of ϕ_c and $\psi_{c'}$ to always be identical. But if we think of the dependences of an utterance as further utterances whose truth-values must yet be discovered to assess the original utterance for a truth-value, the dependences of ϕ_c and $\psi_{c'}$ can differ from each other dramatically, even relative to the same assignment. This is because sometimes learning the truth-value of one utterance in U_1 will be enough to settle the truth-value of one of ϕ_c and $\psi_{c'}$ but not the other. And as long as we can't ascertain the truth-value one of those utterances—ϕ_c say—we may potentially need information about all the unassigned utterances in U_1. This is nicely reflected in the determination of semantic dependences above.

The shifting character of the semantic dependence relation also affords us insight into the semantic dependences of utterances whose quantifiers are restricted using semantic terms, like our example (2) from before.

(2) Every true statement of Kate's yesterday was believed by Tom.

Example 2.2. ϕ_c is $(\forall v_1 : T(v_1) \wedge \theta(v_1))(\delta(v_1))_c$, where θ and δ are formulas of one free variable without semantic vocabulary which, in $\mathcal{M}$, define a set of utterances U_1 and U_2. Let $\bar{U_2}$ be the set of utterances defined by $\neg\delta(v)$. Again, consider three cases:

(a) Let A be undefined on $U_1 \cap \bar{U_2}$ and allow arbitrary permutations over $U_1 \cap \bar{U_2}$. Then $SD_A(\phi_c) = U_1 \cap \bar{U_2}$. Pick some $u_1 \in U_1 \cap \bar{U_2}$, and pick an extension A_1 of A such that $A_1(u_i) = f$ for all $u_i \in U_1 \cap \bar{U_2}$, and another extension A_2 differing from A_1 only in that $A_2(u_1) = t$. Then $A_{\mathcal{M}_{A_1}}(\phi_c) = t \neq A_{\mathcal{M}_{A_2}}(\phi_c) = f$, showing that $\{u_1\}$ matters to ϕ_c relative to A, hence it also really matters to ϕ_c relative to A. Again, since u_1 was arbitrary, any $u_i \in U_1 \cap \bar{U_2}$ is such that $\{u_i\}$ really matters to ϕ_c relative to A. It should be clear that no set of utterances not containing an element from $U_1 \cap \bar{U_2}$ will matter to ϕ_c. So, again given the available extensions of A, singletons among $U_1 \cap \bar{U_2}$ are the only ones that really matter to ϕ_c relative to A, and hence $SD_A(\phi_c) = U_1 \cap \bar{U_2}$.

Note that this means if $U_1 \cap \bar{U_2} = \emptyset$ then $SD_A(\phi_c) = \emptyset$. This is in keeping with my gloss on the semantic dependence relation. If nothing Kate said yesterday was not believed by Tom, then one needs no further information about truth-value distributions to ascertain the truth-value of an utterance of (2) above.

(b) Let A be such that $A(u_2) = f$ for some $u_2 \in U_1 \cap \bar{U_2}$ and is undefined elsewhere on $U_1 \cap \bar{U_2}$ and allows arbitrary permutations over that set. Then, by the same reasoning in (a), $SD_A(\phi_c) = (U_1 \cap \bar{U_2}) - \{u_1\}$.

(c) Let A be such that $A(u_1) = t$ for some $u_1 \in U_1 \cap \bar{U_2}$. Then as long as A has coherent total extensions, $SD_A(\phi_c) = \emptyset$. This is because for each total coherent extension A' of A, $A_{\mathcal{M}_{A'}}(\phi_c) = f$, so no utterances really matter to it.

So a quantified statement whose quantifier is restricted by utterances bearing certain truth-values has a complex set of semantic dependences determined by an interaction between *three* factors: the interpretation of the non-semantic vocabulary in the quantifier restrictor, the interpretation of the quantifier matrix, and the progress made in expanding a truth-value assignment. This seems true to the facts, given my gloss on the semantic dependence relation.

It is worth noting that it is easy for semantic self-dependence to arise on this picture, but also that such semantic self-dependence is often not vicious. ϕ_c in example 2.2 can easily semantically depend on itself if it is among $U_1 \cap \bar{U_2}$ relative to some assignment A. But such self-dependence might present no obstacles to ascertaining the truth-value of ϕ_c because relative to more information—an expanded assignment function—that self-dependence disappears. This would occur as soon as the conditions in (c) are satisfied, for example, so that we can pronounce on the truth-value of ϕ_c whether or not it intuitively ascribes a truth-value to itself.

Vicious self-reference is also not hard to come by.

Example 2.3. Let λ_c be of the form $(\neg T\ulcorner\lambda_c\urcorner_{c'})_c$, and suppose A is undefined on λ_c and $T\ulcorner\lambda_c\urcorner_{c'}$ with coherent extensions that arbitrarily permute only their truth-values (consistent with the conditions of coherence on negation). Then there are two ways to see that $SD_A(\lambda_c) = \{\lambda_c, T\ulcorner\lambda_c\urcorner_{c'}\}$. First, take extensions A_1 and A_2 of A, identical but that $A_1(\lambda_c) = f$ and $A_1(T\ulcorner\lambda_c\urcorner_{c'}) = t$ on the one hand, and $A_2(\lambda_c) = t$ and $A_2(T\ulcorner\lambda_c\urcorner_{c'}) = f$ on the other. Then $A_{\mathcal{M}_{A_1}}(\lambda_c) = t \neq A_{\mathcal{M}_{A_2}}(\lambda_c) = f$. So $\{\lambda_c, T\ulcorner\lambda_c\urcorner_{c'}\}$ matters to λ_c. Note that no two coherent assignments differ *only* on their assignment to λ_c and not $T\ulcorner\lambda_c\urcorner_{c'}$ as well. So $\{\lambda_c, T\ulcorner\lambda_c\urcorner_{c'}\}$ really matters to λ_c. No set devoid of λ_c matters to λ_c, so $SD_A(\lambda_c) = \{\lambda_c, T\ulcorner\lambda_c\urcorner_{c'}\}$. Another way to show that λ_c is semantically self-dependent relative to A is simply to appeal to clauses (ii) and (iii) of the definition of semantic dependence. These conditions, especially (ii), are a safeguard for when the coherent extensions of A required to show semantic dependence aren't around (or, as we'll see in the proofs to come, when their existence is tricky to prove).

The previous example shows one simple case where conditions on coherence may not allow arbitrary permutations of values assigned to utterances. This means that when the values of utterances must be permuted together, they tend to either both be in, or out, of the set of an utterance's semantic dependences relative to a truth-value assignment.

Note that the earlier point made in example 2.2—that ascription of a truth-value to oneself needn't produce semantic self-dependence—is highly relevant to determining the semantic dependences of semantic generalities. To see this consider the following utterance.

Example 2.4. Let ϕ_c be $(\forall v_1 : v_1 = v_1)(T(v_1) \vee \neg T(v_1))$. Then for any A, $SD_A(\phi_c) = \emptyset$. This is because for any total coherent assignment A' whatsoever, $A_{\mathcal{M}_{A'}}(\phi_c) = t$.

The supervaluational character of the determination of semantic dependence ensures that statements whose truth-values are easy to determine from global distributions of semantic properties will have few, if any, semantic dependences.

The foregoing examples should give give a flavor for the behavior of how semantic dependences are determined in the standard case: by appeal only to clause (i) of the definition of semantic dependence and hence to the relation of really mattering. Still, a small caveat is in order. I earlier called these cases 'idealized examples', and I did so because of the assumptions I made about the availability of total coherent extensions of certain kinds. Often the range of extensions that I appealed to will not, or could not, exist in the way I supposed. For example, sometimes the only coherent extensions of an assignment will have to jointly permute tuples of utterances, owing to the constraints on coherence. In some special cases, no coherent extensions of an assignment at all will be available (for example, if we start with a partial truth-value assignment which flouts the coherence constraints). How many, and what kinds of total coherent extensions a partial truth-value assignment has will obviously be an important question that we'll return to later. But hopefully the idealized examples can serve for now in getting a grip on how semantic dependences are typically assessed.

The application of clauses (ii) and (iii) of the definition of semantic dependence should be relatively straightforward, so let's turn to clause (iv) which deals with utterances with unspecifiable dependences. Here is an example of such an utterance which, as I noted before, points to a kind of 'bug' in the characterization of influence that really mattering affords.[13]

Example 2.5. Let D define in $\mathcal{M}$ a set of utterances $\mathcal{D} = \{\psi_{c_i} \mid i \in \omega\}$, and R define in $\mathcal{M}$ the relation $\{\langle \psi_{c_i}, \psi_{c_j} \rangle \mid i < j\}$. Let ϕ_c be

$$(\forall v : Dv)((\exists v' : Dv')(Rvv' \wedge Tv'))$$

Then, supposing A allows arbitrary permutations over $\mathcal{D}$, many sets of utterances matter to ϕ_c including, e.g., $\mathcal{D}$. But no sets of utterances really matter to ϕ_c since for any subset Γ that matters to ϕ_c with witnessing A_1 and A_2, we can always find suitable A_1' and A_2' which, e.g., additionally hold constant the elements in $\Gamma' = \{\psi_{c_i} \in \Gamma \mid i < k\}$ for some k large enough to make Γ' nonempty, while still ensuring $A_{\mathcal{M}_{A_1'}}(\phi_c) \neq A_{\mathcal{M}_{A_2'}}(\phi_c)$. Accordingly, by definition, $SD_A(\phi_c) = \mathcal{U}^{\mathcal{M}} - \mathrm{dom}(A)$.

An utterance ϕ_c with unspecifiable dependences relative to A is one which clearly 'depends' for its truth on the semantic properties of other utterances, but picking out any particular set of utterances as those which could be 'ultimately responsible' for its truth-value is unprincipled, since any set that one picks will always, of necessity, include utterances which can be viewed as extraneous. Utterances with unspecifiable dependences require finer adjustments to the definition of semantic dependence, or perhaps new tools altogether, to capture the sense in which these utterances depend for their truth on the semantic properties of other utterances. Since doing this would unduly complicate the formalism and, by and large, these are relatively outlying cases, I'll be content here with my strategy of stipulating these utterances are semantically dependent on every unassigned utterance.

[13] The example is adapted from Leitgeb (2005).

An important general fact about these semantic dependence relations is that (unspecifiable dependences aside) as truth-value assignments grow, semantic dependences relative to those assignments shrink. In more metaphorical terms: as our idealized reasoner learns more and more about which truth-values utterances bear, the less and less she will need to know to fix the truth-values of other utterances.

Proposition 2.1. *Let A, A', be truth-value assignments with $A \subseteq A'$. Then if ϕ_c and its constituent utterances don't have unspecifiable dependences relative to A', $SD_{A'}(\phi_c) \subseteq SD_A(\phi_c)$.*

Proof. Suppose $\psi_{c'} \in \mathrm{SD}_{A'}(\phi_c)$. If the grounds for the inclusion of $\psi_{c'}$ in $\mathrm{SD}_{A'}(\phi_c)$ are condition (ii)—that $\phi_c = S\ulcorner\psi_{c'}\urcorner$ for some semantic predicate S, and $\psi_{c'} \notin \mathrm{dom}(A')$—then trivially $\psi_{c'} \in \mathrm{SD}_A(\phi_c)$ as well. So suppose instead the grounds for inclusion are condition (i). Then $\psi_{c'} \in \Gamma$ for some Γ that really matters to ϕ_c relative to A', and there are total coherent extensions A'_1, A'_2 of A', differing only on Γ, such that $A_{\mathcal{M}_{A'_1}}(\phi_c) \neq A_{\mathcal{M}_{A'_2}}(\phi_c)$. Since $A \subseteq A'$, A'_1 and A'_2 extend A, witnessing that Γ matters to ϕ_c relative to A. Now, if there were A''_1, A''_2, and $\emptyset \neq \Gamma' \subseteq \Gamma$ which showed that Γ doesn't matter to ϕ_c relative to A, they would equally show Γ doesn't matter to ϕ_c relative to A'—a contradiction. So Γ really matters to ϕ_c relative to A and $\psi_{c'} \in \mathrm{SD}_A(\phi_c)$. Finally, if the grounds for the inclusion of $\psi_{c'}$ in $SD_{A'}(\phi_c)$ are that $\psi_{c'}$ has a constituent utterance $\gamma_{c''}$ meeting (i) or (ii), then $\gamma_{c''}$ meets (i) or (ii) for $\psi_{c'}$ relative to A as we have just shown, so again $\psi_{c'} \in SD_A(\phi_c)$. □

Recall that I had two main tasks in giving my semantics based on the idea that the meaning of the truth-predicate and other semantic terms are given by a procedure. They were:

(I) to formalize the semantic dependence relation, and

(II) to formalize the method speakers use to assign truth values along chains of semantic dependences.

The first of these tasks is now accomplished, and the second is not far off. In the rest of this section, I'd like to develop some formal tools that will help us with (II).

To begin, we should note that when I said earlier that speakers assign truth-values 'along a chain of semantic dependences', I was ignoring the complications that arise when utterances directly or indirectly semantically depend on themselves. The possibility of various kinds of semantic self-dependence and interdependence poses a slight problem for the idea that the semantic dependence relations provide us with an order in which speakers assign truth-values to utterances.

But the kinds of knots and tangles that arise in the dependence relations can be ironed out to create the relevant orderings. Note that a semantic dependence relation, relative to some truth-value assignment A, creates a directed graph among the utterances of $\mathcal{U}^{\mathcal{M}}$. And this directed graph can be transformed into a more revealing partial ordering by creating certain equivalence classes of utterances as follows.

Definition 2.12. Let A be a truth-value assignment and let SD^*_A be the transitive closure of SD_A (considered as a relation). A *semantic cluster relative to* A is either:

(1) A set Γ such that
   (a) For all $\phi_c, \psi_{c'} \in \Gamma$, $\phi_c \in SD^*_A(\psi_{c'})$, and
   (b) For all $\phi_c \in \Gamma$, if there is a $\psi_{c'}$ such that $\psi_{c'} \in SD^*_A(\phi_c)$ and $\phi_c \in SD^*_A(\psi_{c'})$, then $\psi_{c'} \in \Gamma$; or

(2) A singleton $\{\phi_c\}$ such that $\phi_c \notin \Gamma$ for any Γ satisfying the conditions of (1).

Each truth-value assignment A yields a semantic dependence relation SD_A which generates a set of semantic clusters that partition $\mathcal{U}^{\mathcal{M}}$

Definition 2.13. Let $\mathcal{C}_A = \{\Gamma \mid \Gamma \text{ is a semantic cluster relative to } A\}$.

Proposition 2.2. *For all truth-value assignments A, $\mathcal{C}_A$ partitions $\mathcal{U}^{\mathcal{M}}$.*

Proof. Singletons satisfying condition (2) of definition 2.12 clearly do not intersect each other, nor any of the sets satisfying condition (1) by stipulation. Suppose Γ and Γ' satisfy condition (1) and $\phi_c \in \Gamma \cap \Gamma'$. Pick any $\psi_{c'} \in \Gamma$ and $\theta_{c''} \in \Gamma'$. Then by applications of condition (1a) we have

(i) $\phi_c \in \mathrm{SD}^*_A(\psi_{c'})$ and $\psi_{c'} \in \mathrm{SD}^*_A(\phi_c)$

(ii) $\phi_c \in \mathrm{SD}^*_A(\theta_{c''})$ and $\theta_{c''} \in \mathrm{SD}^*_A(\phi_c)$.

But then by condition (1b) and the transitivity of SD^*_A we have $\psi_{c'} \in \Gamma'$, $\theta_{c''} \in \Gamma$. Since $\psi_{c'}$ and $\theta_{c''}$ were arbitrary members of Γ and Γ', we have $\Gamma = \Gamma'$.

$\mathcal{U}^{\mathcal{M}} \subseteq \bigcup \mathcal{C}_A$ since every utterance is in a cluster with properties (1) or (2) by stipulation. ☐

This legitimates the following definition

Definition 2.14. For truth-value assignment A and utterance ϕ_c, we define Γ_{A,ϕ_c} to be the semantic cluster relative to A containing ϕ_c.

Now, the partition generated by taking semantic clusters can be partially ordered with the help of the semantic dependence relation.

Definition 2.15. Let A be a truth-value assignment and $\mathcal{C}_A$ be the set of semantic clusters relative to A. Then let $\leq_A$ $(\subseteq \mathcal{C}_A \times \mathcal{C}_A)$ be given by

$$\Gamma \leq_A \Theta \text{ iff } \exists \phi_c \in \Gamma \text{ and } \exists \psi_{c'} \in \Theta \text{ such that } \phi_c \in SD^*_A(\psi_{c'}) \text{ or } \Gamma = \Theta$$

Proposition 2.3. *For all truth-value assignments A, $\leq_A$ partially orders $\mathcal{C}_A$*

Proof. Reflexivity follows from the definition, transitivity by the stipulated transitivity of SD^*_A, and antisymmetry by the 'maximality' condition (1b) from Definition 2.12. $\square$

This ordering will play an important role in the formalism to ensue. In particular, a certain type of semantic cluster will be of particular relevance: those that are, in a sense, 'up next' for assignment. These are the clusters which are least in the ordering.

Definition 2.16. Let A be a truth-value assignment. An utterance $\phi_c \notin \text{dom}(A)$ is A-*minimal* if

(i) there is no semantic cluster Γ such that $\Gamma <_A \Gamma_{A,\phi_c}$, or

(ii) ϕ_c has as a component utterance, or is a component utterance of, some unassigned $\psi_{c'}$ meeting condition (i).

Another important phenomenon that we will want to keep our eyes out for is sets of utterances whose clusters have *no* A-minimal element.

Definition 2.17. A set of utterances $\{\phi_{c_0}, \phi_{c_1}, \ldots \phi_{c_n} \ldots\}$ is a *semantic chain relative to a truth-value assignment* A if for $\phi_{c_{i+1}} \in \text{SD}^*_A(\phi_{c_i})$ for all $i \in \omega$, and for all $i \neq j$, $\Gamma_{A,\phi_{c_i}} \neq \Gamma_{A,\phi_{c_j}}$.

Part of the reason to keep track of semantic chains is that they are known to give rise to paradoxes.[14]

3 Procedural Evaluation Formalized

We have our relations of semantic dependence in hand, so it is now time to say how speakers assign truth-values to utterances along the orderings of semantic clusters they generate. Fortunately the work here is largely already done, since the definition of semantic dependence implicitly contains information about how this assignment is to proceed.

To see this, consider our idealized reasoner, who has progressed in allotting truth-values to utterances according to some truth-value assignment A. Those utterances which have no semantic dependences relative to that A have an important property: they *evaluate* to a single truth-value across the range of supervaluations that are used to ascertain semantic dependence. Recall our gloss on the semantic dependence relation: given a known partial allotment of truth-values to utterances, it is the relation which tells us which utterances' truth-values one might still need to fix in order to fix the truth-value of an utterance under consideration. And the way this relation was determined was by showing which utterances could influence the *evaluated* truth-value of a considered utterance by permuting them. So when an utterance has no semantic

[14] Yablo (1993).

dependences no utterances matter to it and it is 'ready to be assigned'. Moreover, it has a stable evaluated truth-value privileged by the procedure which determines it is ready to be assigned.

For any assignment A, we can group together the unassigned utterances ready to be assigned in this way according to the truth-values they should be allotted.

Definition 3.1. Let A be a truth-value assignment, ϕ_c be an utterance. We say

> ϕ_c is *stably true* on A if for all total coherent extensions A' of A, $A_{\mathcal{M}_{A'}}(\phi_c) = t$,
>
> ϕ_c is *stably false* on A if for all total coherent extensions A' of A, $A_{\mathcal{M}_{A'}}(\phi_c) = f$, and
>
> ϕ_c is *stably defective* on A if for all total coherent extensions A' of A, $A_{\mathcal{M}_{A'}}(\phi_c) = u$.

We define

$$t[A] = \{\phi_c \in \mathcal{U}^{\mathcal{M}} \mid \phi_c \notin \mathrm{dom}(A) \text{ is stably true on } A, \mathrm{SD}_A(\phi_c) = \emptyset\}$$

$$f[A] = \{\phi_c \in \mathcal{U}^{\mathcal{M}} \mid \phi_c \notin \mathrm{dom}(A) \text{ is stably false on } A, \mathrm{SD}_A(\phi_c) = \emptyset\}$$

$$u[A] = \{\phi_c \in \mathcal{U}^{\mathcal{M}} \mid \phi_c \notin \mathrm{dom}(A) \text{ is stably defective on } A, \mathrm{SD}_A(\phi_c) = \emptyset\}$$

So $t[A]$ represents those utterances that a reasoner who has arrived at the assignment A should *add* to that assignment as true. The reason for adding the condition that a $\mathrm{SD}_A(\phi_c) = \emptyset$ is to accommodate the additional conditions (ii) and (iv) on the semantic dependence relation: sometimes an utterance may be stably true, though it intuitively must 'wait' to be assigned because it has dependences secured by those added clauses.

I'll call the process by which our idealized reasoner expands their truth-value assignment with stable utterances *boosting.* As a reasoner progressively boosts to include relevant stable utterances, they approach a limit point before a number of steps determined as a function of $|\mathcal{U}^{\mathcal{M}}|$. Supposing for now that $|\mathcal{U}^{\mathcal{M}}| \leq \omega$, we know that the process of adding stable utterances will terminate before ω_1 iterations. But of course, at this stage, for many reasonable sets of utterances, the truth-value assignment function arrived at will still be partial. Many utterances left over will be utterances that are semantically self-dependent, or elements of semantic chains.

A central idea motivating the present formalism is that when the semantic value of a word does not unambiguously determine a truth-value for a particular utterance, then that utterance is defective in the sense marked off by the value u in the formalism. Certain utterances unassigned after boosting is exhausted are 'hopeless' in this sense: not only have they not yet been assigned a value in the procedural assignment but *if* they were ever to be so assigned, they *would* have been by now. These are the A-minimal utterances defined before, those least in the order of semantic clusters. These are groups of semantically interdependent and self-dependent utterances which depend on nothing further. So if any utterances deserve to be counted as defective, on the motivations alluded to, it is these.

This provides another method for assignment function expansion: taking an assignment function stable under boosting, and relegating the assignment-minimal utterances to those that are defective. I'll call this part of the procedural assignment *culling*.[15] It is important to cull only assignment minimal utterances, since as soon as this is done, other utterances that are less hopelessly situated in the hierarchy of semantic dependences open up again for normal assignment. Consider again the proposed treatment of our variant of the two-line paradox:

> Jones at t: "What Jones utters at t is false or defective."
>
> Jane at t_1: "What Jones utters at t is false or defective."

A model including the utterances of Jones and Jane would treat Jones' utterance as semantically self-dependent, and Jane's utterance as semantically dependent on just Jones' utterance (and perhaps its constituents). Thus Jones' utterance would never be assigned through the initial boosting process, nor would Jane's. At the end of the boosting process, Jones' utterance would be culled, but Jane's would not. This means if we begin boosting immediately after culling, utterances like Jane's, and those semantically dependent on Jane's utterance, are free to be boosted to truth or falsity as we like, in keeping with the kind of resolution to the two-line paradox that I endorsed earlier.

The process of alternating boosting until boosting is exhausted, then culling, then boosting again to exhaustion, and so on, is a process that can itself be iterated to exhaustion. Again, assuming $|\mathcal{U}^{\mathcal{M}}| \leq \omega$, ω_1 steps will suffice. At this point, if any utterances remain unassigned, I propose that they be relegated to be defective as well. I'll say a little more about these final utterances shortly. I'll call the process by which these final utterances are assigned u, the *discarding stage*. At this point, all utterances in $\mathcal{U}^{\mathcal{M}}$ have been procedurally assigned truth-values, in keeping with our goal (II) from above.

So to recaptulate: the process of assigning truth-values that I am proposing alternates boosting to exhaustion, and culling, until iterations of that pair of operations is itself exhausted. Any remaining utterances are discarded. The following definition formalizes this process. Subscripts up to ω_1 index boosting operations, and superscripts index culling operations. We start the process off with the partial assignment engendered by the base model $\mathcal{M}$—that is $A_{\mathcal{M}}$.

Definition 3.2. The *semantic extension of* $\mathcal{M}$, written $\mathcal{M}^*$ is the set of ordered pairs given as follows:

[15] Compare the *Closed Loop Rule* of Gaifman (1992, 2000).

$$
\begin{array}{rcll}
A^0_0 = & A_{\mathcal{M}} & & \\
A^\beta_{\alpha+1} = & A^\beta_\alpha & \cup \{\langle \phi_c, t\rangle \mid \phi_c \in t[A^\beta_\alpha]\} \\
& & \cup \{\langle \phi_c, f\rangle \mid \phi_c \in f[A^\beta_\alpha]\} \\
& & \cup \{\langle \phi_c, u\rangle \mid \phi_c \in u[A^\beta_\alpha]\} \\
& & \text{for } \alpha < \omega_1 \\
A^\beta_\lambda = & \bigcup_{\eta<\lambda} A^\beta_\eta & \text{for limit ordinals } \lambda \\
A^{\beta+1}_0 = & A^\beta_{\omega_1} & \cup \{\langle \phi_c, u\rangle \mid \phi \text{ is } A^\beta_{\omega_1}\text{-minimal}\} \\
A^\lambda_0 = & \bigcup_{\eta<\lambda} A^\eta_0 & \text{for limit ordinals } \lambda \\
\mathcal{M}^* = A^0_{\omega_1+1} = & A^{\omega_1}_0 & \cup \{\langle \phi_c, u\rangle \mid \langle \phi_c, v\rangle \notin A^{w_1}_0 \text{ for any } v\}
\end{array}
$$

Note that by construction $A^\beta_\alpha \subseteq A^{\beta'}_{\alpha'}$ for $\langle \beta, \alpha\rangle < \langle \beta', \alpha'\rangle$ (on the lexicographic ordering). I abuse a bit of notation by labeling our master assignment $\mathcal{M}^*$. $\mathcal{M}^*$ after all, is not a model in the traditional sense but only a set of ordered pairs. But it is designed to behave like one in a crucial respect: it partitions utterances according to their truth-value assignment.

I have already given loose justifications for the behavior of the boosting and culling operations, but what of the discarding operation? For example, what kinds of utterances are discarded? It turns out only utterances that are included in special kinds of semantic chains are assigned at this final stage.[16]

Proposition 3.1. *Provided $A^{w_1}_0$ is a truth-value assignment, if $\phi_{c_0} \notin$ dom($A^{w_1}_0$), then ϕ_c is in a semantic chain relative to $A^{w_1}_0$.*

Proof. Let $\phi_{c_0} \notin \mathrm{dom}(A^{w_1}_0)$ and consider the semantic cluster $\Gamma_{A^{\omega_1}_0, \phi_{c_0}}$. There must be a $\Gamma < \Gamma_{A^{\omega_1}_0, \phi_{c_0}}$, otherwise ϕ_{c_0} would be $A^{\omega_1}_0$-minimal and, if culled, would expand $A^{\omega_1}_0$. However, $A^{\omega_1}_0$ is stable under the operation of culling by construction. Since there is a $\Gamma < \Gamma_{A^{\omega_1}_0, \phi_{c_0}}$, we can pick some $\phi_{c_1} \in \Gamma$. We know that $\phi_{c_1} \notin \mathrm{dom}(A^{w_1}_0)$ (since semantic dependences relative to an assignment are always unassigned by definition). So by similar considerations, we can show that there is a $\Gamma < \Gamma_{A^{\omega_1}_0, \phi_{c_1}}$. In this way we can generate a series $\{\phi_c, \phi_{c_1}, \ldots \phi_{c_n} \ldots\}$. By construction $\phi_{c_{i+1}} \in \mathrm{SD}^*_A(\phi_{c_i})$ for all $i \in \omega$, and $\Gamma_{A^{\omega_1}_0, \phi_{c_i}} \neq \Gamma_{A^{\omega_1}_0, \phi_{c_j}}$ for $i \neq j$. □

It is important to note that not all utterances in semantic chains are 'hopeless'. In fact many utterances in semantic chains may get assigned truth-values at boosting stages in the process of assignments. This is because what are initially semantic chains relative to one assignment, may cease to be with respect to another, expanded assignment. But the semantic chains left over after boosting and culling stages are exhausted are as hopeless as can be. They are parts of infinite descending chains of semantic dependences, with no indication as to

[16]Thus, compare my discarding rule to the *Groundless Pointer Rule* of Gaifman (2000).

'where to begin' in assigning them truth-values. These are the utterances which are discarded.[17]

This concludes the second and final stage of my proposal: to formalize the method used to assign truth values along chains of semantic dependences.

4 Coherence and Harmony

It remains to explore the formal consequences of adopting this conception of the semantic value of semantic terms like "true". A first step in this process is to show that the construction avoids a special problem—a problem which arises uniquely for my system because of two features it has. First, the final system consists merely in a pairing of utterances with truth-values, where those truth-values are not paired through the familiar kind of compositional process that occurs in standard models (whether bivalent or trivalent). Second, the system makes those pairings roughly via a series of supervaluations over a class of 'well-behaved' assignments—assignments which exhibit some salient structure that standard compositional models would have. These were the coherent assignments.[18]

The special problem that arises for my system is that it is imperative to show that, at every stage of the pairing process, we have an assignment with total coherent extensions. If not, then the conditions on being stably true, false, and defective, will all be trivially satisfied by virtually every utterance at that stage, and as a consequence virtually every unassigned utterance will be associated *with all three truth values.* So we cannot yet be sure that $\mathcal{M}^*$ is a truth-value assignment: it might not be a function if the right kinds of coherent truth-value assignments do not exist.

The goal of this section is to show that this unfortunate outcome doesn't arise. This has another important beneficial consequence: if every assignment has coherent extensions, then the master assignment $\mathcal{M}^*$ is itself coherent. Thus there is a kind of 'harmony' that exists between the supervaluational techniques employed in boosting and the end result of assigning truth-values to utterances using those techniques. In particular, any utterance which is assigned t or f, is so assigned by a supervaluation over a class of truth-value distributions *including* the distribution which is the end product of such assignments. This kind of harmony in my proposed system is fairly straightforward consequence of the structure of the assignment process, but it will take a little work to show this.

How can we show that every assignment in the process of determining $\mathcal{M}^*$ has coherent extensions? Naturally by finding a 'standard way' of generating a total coherent truth-value assignment form a partial one that is already known

[17] This importantly means my rules can't be applied in an 'order independent' way—a distinction between my procedural assignment and that of Gaifman.

[18] The first feature marks an important distinction from Kripke (1975) and a large class of philosophers building on his system. The second feature distinguishes my proposal from the only other token-sensitive view I am aware of, namely Gaifman (1992, 2000).

to be coherent. I'll begin by characterizing this notion of a standard coherent extension, and then show how every partial assignment generated by the procedure can be extended in that standard way.

First, some perfunctory definitions.

Definition 4.1. t, f, and u are *truth values.*

Definition 4.2. A truth-value assignment A is *coherently extensible*, if A has some total coherent extension.

My method for generating a total coherent assignment from a partial assignment that is coherent is given as follows.

Proposition 4.1. *The* weak extension *of a truth-value assignment A is*

$$A \cup \{\langle \phi_c, u\rangle \mid \langle \phi_c, v\rangle \notin A \textit{ for any } v\}$$

The weak extension of a coherent assignment has good potential to be coherent itself, since coherence puts relatively few conditions on defective utterances. Coherence of the partial assignment, however, isn't enough to secure the coherence of its weak extension. We need the following additional constraint on assignments to special utterances using semantic vocabulary.

Definition 4.3. A truth-value assignment A is *weakly extensible* if A is coherent and if whenever $(S\ulcorner \phi_{c'} \urcorner)_c \in \text{dom}(A)$ for a semantic predicate S, $\phi_{c'} \in \text{dom}(A)$ as well.

Proposition 4.2. *If A is weakly extensible its weak extension is coherent, hence A is coherently extensible.*

Proof. Conditions (i), (ii), and (v) of Definition 2.6 are easily verified since the weak extension only adds pairs of the form $\langle \phi_c, u\rangle$. Conditions (iii) and (iv) are secured by the added constaint on weak extensibility. $\square$

My goal now is to show that A^β_α is weakly extensible for all α and β. To do this, I will need a short series of ancillary results that will make use of the following natural notion of utterance rank.

Definition 4.4. The *rank* of an utterance ϕ_c, rank(ϕ_c), is the pair $\langle \beta, \alpha\rangle$ such that $\langle \phi_c, v\rangle \in A^\beta_\alpha$ for some v, but for all $\langle \beta', \alpha'\rangle < \langle \beta, \alpha\rangle$ (where $<$ is the lexicographic ordering), $\langle \phi_c, v\rangle \notin A^{\beta'}_{\alpha'}$ for any v. We say that $\langle \beta, \alpha\rangle$ is

(i) ...the *initial stage* if $\alpha = \beta = 0$.

(ii) ...a *boosting stage* if $\alpha < \omega_1$ is a successor ordinal .

(iii) ...a *culling stage* if $\alpha = 0$ and β is a successor ordinal.

(iv) ...a *collecting stage* if α or β are limit ordinals.

(v) ...a *discarding stage* if $\alpha = \omega_1 + 1$, $\beta = 0$.

A first ancillary claim is that utterances assigned in the initial or boosting stages *evaluate* to the same truth-value they are assigned on all coherent extensions.

Proposition 4.3. *Suppose rank(ϕ_c) = $\langle \beta, \alpha \rangle$ is the initial stage or a boosting stage and $\langle \phi_c, v \rangle \in A^\beta_\alpha$. Then for all $\langle \beta', \alpha' \rangle \geq \langle \beta, \alpha \rangle$, and for all total coherent extensions A' of $A^{\beta'}_{\alpha'}$, $A_{\mathcal{M}_{A'}}(\phi_c) = v$. Since the initial and boosting stages are the only stages at which formulas are paired with the truth values t or f, it follows that for any $\langle \beta'', \alpha'' \rangle$, if $\langle \phi_c, v \rangle \in A^{\beta''}_{\alpha''}$ for $v \in \{t, f\}$, then any total coherent extension A' of $A^{\beta''}_{\alpha''}$ is such that $A_{\mathcal{M}_{A'}}(\phi_c) = v$.*

Proof. If $\langle \beta, \alpha \rangle$ is the initial stage, then ϕ_c evaluates to v in the models engendered by any assignments at all, so the result holds. So suppose $\langle \beta, \alpha \rangle$ is a boosting stage. If A^β_α is not a coherent truth-value assignment the claim is trivially true. So suppose further that A^β_α is coherent. By assumption ϕ_c has value v in any model engendered by a total complete extension of A^β_α. Then, if $A^\beta_\alpha \subseteq A^{\beta'}_{\alpha'}$, any total coherent extension of $A^{\beta'}_{\alpha'}$ is a total coherent extension of A^β_α, and so is one in whose engendered model ϕ_c has the value v. □

The rest of the ancillary claims needed are results about the ranks of utterances with the kind of structure relevant to the determination of coherence. The constraints on rank will constrain the possible permutations of truth-value assignments to the utterances mentioned in the coherence conditions. This in turn will help restrict what is required to show that the truth-value assignment function respects the requirements of coherence. For example, the following result concerns the relative rank of conjunctions and their conjuncts: conjunctions are never assigned before their conjuncts.

Proposition 4.4. *If $\phi_c = (\psi_{c'} \wedge \theta_{c''})_c$, then rank($\psi_{c'}$), rank($\theta_{c''}$) $\leq$ rank(ϕ_c)*

Proof. Note that by definition $\mathrm{SD}_A(\psi_{c'})$, $\mathrm{SD}_A(\theta_{c''}) \subseteq \mathrm{SD}_A(\phi_c)$, relative to any assignment A. Let rank(ϕ_c) = $\langle \beta, \alpha \rangle$. If $\langle \beta, \alpha \rangle$ is the initial stage, the result follows trivially. If $\langle \beta, \alpha \rangle$ is a boosting stage, then ϕ_c has no semantic dependences on $A^\beta_{pred(\alpha)}$, so neither do $\psi_{c'}$ and $\theta_{c''}$. As such, if they are not assigned already, they are stable and will be assigned along with ϕ_c at $\langle \beta, \alpha \rangle$. If $\langle \beta, \alpha \rangle$ is a culling stage, then ϕ_c is $A^{pred(\beta)}_{\omega_1}$-minimal*. By definition, if $\psi_{c'}$ or $\theta_{c''}$ are unassigned, they are $A^{pred(\beta)}_{\omega_1}$-minimal* as well, and are culled at $\langle \beta, \alpha \rangle$. If $\langle \beta, \alpha \rangle$ is a the discarding stage, the result follows trivially. □

And a final ancillary claim concerns the rank of ascriptions of truth-values more generally.

Proposition 4.5. *Suppose $\phi_c = S\ulcorner \psi_{c'} \urcorner$ where S is a semantic predicate. Then rank($\psi_{c'}$) $\leq$ rank(ϕ_c).*

Proof. Let rank($\psi_{c'}$) = $\langle \beta, \alpha \rangle$. If $\langle \beta, \alpha \rangle$ is the initial stage the result follows trivially. If $\langle \beta, \alpha \rangle$ is a boosting, culling, or discarding stage, then for all $\langle \beta', \alpha' \rangle <$

$\langle\beta,\alpha\rangle$, $\psi_{c'} \notin A^{\beta'}_{\alpha'}$, whence $\psi_{c'} \in \mathrm{SD}_{A^{\beta'}_{\alpha'}}(\phi_c)$ by condition (ii) of the definition of semantic dependence. So ϕ_c will not be assigned at the boosting or culling stages prior to $\langle\beta,\alpha\rangle$ and the result holds. □

With these ancillary claims established we are now ready to prove our coherence result.

Theorem 4.1. For all $\langle\beta,\alpha\rangle < \langle 0,\omega_1+1\rangle$, A^β_α is weakly extensible. It follows that $\mathcal{M}^*$ is a total coherent truth-value assignment.[19]

Proof. We show by induction that for all $\langle\beta,\alpha\rangle \leq \langle 0,\omega_1+1\rangle$

(A) all conditions (i)–(v) on coherence are satisfied by A^β_α, and

(B) if $\langle (S\ulcorner\psi_{c'}\urcorner)_c, v\rangle \in A^\beta_\alpha$ for some v and semantic predicate S, then $\langle\psi_{c'}, v\rangle \in A^\beta_\alpha$ for some v.

Let $\langle\beta,\alpha\rangle = \langle 0,0\rangle$ be the *initial stage.* Then:

A^0_0 satisfies conditions (i), (ii), and (v) of coherence since the allotment of truth values in $\mathcal{M}$ does. It also satisfies conditions (iii), (iv) and property (B) trivially.

Suppose for $\langle\beta',\alpha'\rangle < \langle\beta,\alpha\rangle$ (A) and (B) hold and suppose $\langle\beta,\alpha\rangle$ is...

...a *boosting stage.*

(A) (i) Let $\phi_c = (\neg\psi_{c'})_c$, rank($\phi_c$)=$\langle\beta,\alpha\rangle$, and $A^\beta_\alpha(\phi_c) \in \{t,f\}$. ϕ_c and $\psi_{c'}$ share their semantic dependences relative to any assignment, so rank(ψ_c)=$\langle\beta,\alpha\rangle$. Now, ϕ_c is stably false on $A^\beta_{pred(\alpha)}$ iff $\psi_{c'}$ is stably true on it. Likewise ϕ_c is stably true on $A^\beta_{pred(\alpha)}$ iff $\psi_{c'}$ is stably false over it. So condition (i) holds at A^β_α.

(ii) Let $\phi_c = (\psi_{c'} \wedge \theta_{c''})_c$. By proposition 4.4, rank($\psi_{c'}$), rank($\theta_{c''}$) $\leq$ rank(ϕ_c).
Suppose rank(ϕ_c) $= \langle\beta,\alpha\rangle$. Then $A^\beta_\alpha(\phi_c) = t$ iff ϕ_c is stably true on $A^\beta_{pred(\alpha)}$ iff $\psi_{c'}$ and $\theta_{c''}$ are stably true on $A^\beta_{pred(\alpha)}$ iff $A^\beta_\alpha(\psi_{c'}) = A^\beta_\alpha(\theta_{c''}) = t$ (by proposition 4.3 and the fact that if either $\psi_{c'}$ or $\theta_{c''}$ were culled, ϕ_c would have been as well). Likewise $A^\beta_\alpha(\phi_c) = f$ iff ϕ_c is stably false on $A^\beta_{pred(\alpha)}$ iff one of $\psi_{c'}$ and $\theta_{c''}$ is stably false, and the other stably true or false on $A^\beta_{pred(\alpha)}$ iff either $A^\beta_\alpha(\psi_{c'}) = f$ and $A^\beta_\alpha(\theta_{c''}) \in \{t,f\}$ or $A^\beta_\alpha(\psi_{c'}) \in \{t,f\}$ and $A^\beta_\alpha(\theta_{c''}) = f$.
Suppose instead rank(ϕ_c) $>$ rank($\psi_{c'}$) $= \langle\beta,\alpha\rangle$. Suppose, by way of contradiction, that rank($\theta_{c''}$) $\leq \langle\beta,\alpha\rangle$. Then if

[19] In what follows to simplify notation I'll use $A(\phi_c) = t$ instead of $\langle\phi_c, t\rangle \in A$—this abuse of notation is legitimated by the proof itself.

rank($\theta_{c''}$) is the initial stage, or a boosting stage, by proposition 4.3, $\theta_{c''}$ is stable on $A^{\beta}_{pred(\alpha)}$. So, by assumption, is $\psi_{c'}$. But then so is ϕ_c—a contradiction. Moreover, rank($\theta_{c''}$) cannot be a culling stage, since if $\theta_{c''}$ were culled at some stage, $\psi_{c'}$ and ϕ_c would have been culled at that same stage—another contradiction. So rank rank($\theta_{c''}$) $> \langle\beta,\alpha\rangle$. Thus $\theta_{c''} \notin \mathrm{dom}(A^{\beta}_{\alpha})$, and condition (ii) is trivially satisfied.

(iii) Let $\phi_c = (T\ulcorner\psi_{c'}\urcorner)_c$, and suppose $A^{\beta}_{\alpha}(\phi_c) \in \{t, f\}$. If rank($\phi_c$) $<$ $\langle\beta,\alpha\rangle$ the result follows from the induction hypothesis. So let rank(ϕ_c) $= \langle\beta,\alpha\rangle$. Then rank($\psi_{c'}$) $= \langle\beta',\alpha'\rangle < \langle\beta,\alpha\rangle$, otherwise by definition of semantic dependence $\psi_{c'} \in \mathrm{SD}_{A^{\beta}_{pred(\alpha)}}(\phi_c)$, a contradiction. So $A^{\beta}_{\alpha}(\phi_c) = t$ iff ϕ_c is stably true on $A^{\beta}_{pred(\alpha)}$ iff $A^{\beta}_{pred(\alpha)}(\psi_{c'}) = t$ iff $A^{\beta}_{\alpha}(\psi_{c'}) = t$.

(iv) The proof is analogous to that for (iii).

(v) When an utterance ϕ_c is boosted, it receives a constant value across a range of engendered models, which make no distinctions among tokens of sentence types. Given this, the condition follows from proposition 4.3.

(B) Follows directly from the claim that rank($\psi_{c'}$) $\leq$ rank(ϕ_c), which holds by Proposition 4.5.

. . . a *culling stage*:

(A)(i,ii) By the induction hypothesis $A^{pred(\beta)}_{\omega_1}$ is weakly extensible, hence coherent. Since at the culling stage we only add pairs of the form $\langle\phi_c, u\rangle$, properties (i) and (ii) on coherence will be preserved.

(iii) Let $\phi_c = (T\ulcorner\psi_{c'}\urcorner)_c$ and $A^{\beta}_{\alpha}(\phi_c) \in \{t, f\}$. So rank($\phi_c$) $=$ $\langle\beta',\alpha'\rangle < \langle\beta,\alpha\rangle$ where $\langle\beta',\alpha'\rangle$ is a boosting stage, and the result holds by the induction hypothesis.

(iv) As with (iii).

(v) Follows from the fact that only pairs of the form $\langle\phi_c, u\rangle$ are added.

(B) As before, follows from Proposition 4.5.

. . . a *gathering stage*:

(A)/(B) Both follow from the induction hypothesis.

□

Now that we know that $\mathcal{M}^*$ is a function, I'll follow through on my earlier abuse of notation and write $[\![\phi_c]\!]^{\mathcal{M}^*} = t$, $[\![\phi_c]\!]^{\mathcal{M}^*} = f$, and $[\![\phi_c]\!]^{\mathcal{M}^*} = u$ when $\langle\phi_c, t\rangle \in \mathcal{M}^*$, $\langle\phi_c, f\rangle \in \mathcal{M}^*$, and $\langle\phi_c, u\rangle \in \mathcal{M}^*$ respectively.

5 Properties of the Semantics

Now that we've established that $\mathcal{M}^*$ is an assignment, we can begin to explore its structural features. Right off the bat we can say something very informative about how familiar an assignment $\mathcal{M}^*$ is by comparison with more standard models. We can do this by comparing $\mathcal{M}^*$, which is a truth-value assignment, with the standard model that the assignment engenders, $\mathcal{M}_{\mathcal{M}^*}$ as follows.

Proposition 5.1. *For all $\phi_c \in \mathcal{U}^{\mathcal{M}}$*

$$\begin{array}{lcl} [\![\phi_c]\!]^{\mathcal{M}^*} = t & \Rightarrow & [\![\phi]\!]^{\mathcal{M}_{\mathcal{M}^*}} = t \\ [\![\phi_c]\!]^{\mathcal{M}^*} = f & \Rightarrow & [\![\phi]\!]^{\mathcal{M}_{\mathcal{M}^*}} = f \\ [\![\phi_c]\!]^{\mathcal{M}^*} = u & \Leftarrow & [\![\phi]\!]^{\mathcal{M}_{\mathcal{M}^*}} = u \end{array}$$

The converses of each claim may fail.

Proof. If $[\![\phi_c]\!]^{\mathcal{M}} = t$ then ϕ_c was decided at the initial stage or some boosting stage $\langle \beta, \alpha \rangle$. So ϕ evaluates to true on the models engendered by all total coherent extensions of A^{β}_{α}. But $\mathcal{M}_{\mathcal{M}^*}$ is just such a model. Likewise for $[\![\phi_c]\!]^{\mathcal{M}} = f$. The third claim follows from the first two and the fact that $\mathcal{M}^*$ is total.

Utterances which self-attribute U are counterexamples to the converse of the first claim. Sentences which self-attribute $\neg U$ are counterexamples to the converse of the second claim. Both are counterexamples to the third. □

This simple result merits several comments.

First, it is important to note that the failures of all converse entailments are not *limitations* of the system. On the contrary, any system which embodies a resolution of the two-line paradox like the one I endorse must witness these kinds of departures from standard models. Those failures are part of the *point* of developing the kind of truth-value allotment that semantic extensions represent.

Bearing this point in mind, the result shows that $\mathcal{M}^*$ is a truth-value assignment among utterances which is exactly that provided by a more standard trivalent model (without utterance sensitivity), *except that some additional utterances are assigned u*. That is, the model is a kind of minimal departure from standard compositional modes of truth-value assignment required to accommodate the token-sensitivity which motivates the system.

The result also gives us some resources to describe the logic of semantic extensions by tracing out its relationship to the logic of the base models of §1. There are a number of ways of extending a notion of logical consequence to the trivalent setting, but I'd like to focus here on the notion of what is sometimes called 'Strawson entailment'.

Definition 5.1. Γ *Strawson entails* ϕ, noted $\Gamma \models_{se} \phi$, if for all total base models $\mathcal{M}$, if $[\![\gamma]\!]^{\mathcal{M}} = t$ for all $\gamma \in \Gamma$ and $[\![\phi]\!]^{\mathcal{M}} \neq u$, then $[\![\phi]\!]^{\mathcal{M}} = t$.

Strawson entailment is of special interest since it effectively captures the contribution made by logical form to truth-preserving inference in trivalent models where only truth is treated as a designated value. There are also some grounds to think that a version of Strawson entailment is most effectively adopted as a logical consequence relation for natural language in which highly projective semantic defect, like that I've posited, is present.[20]

The Strawson entailment relation is a subset of the bivalent entailment relation $\models_b$ (for our language with our generalized quantifiers). After all, if $\Gamma \models_{se} \phi$ then $\Gamma \models_b \phi$, since every bivalent base model is a form of trivalent base model. Note also that as concerns propositional entailment, Strawson entailment simply coincides with the bivalent consequence relation, which is classical. If there is a some trivalent base model invalidated a classical propositional entailment from Γ to ϕ, since Γ and ϕ are truth-evaluable in the counter-model, all their propositional components are truth-evaluable, so this would be counterexample to the inference in the classical setting as a well—a *reductio*.[21]

Now, if we take Strawson entailment to be our logic for the base setting, we see that semantic extensions respect it.

Proposition 5.2. *Let $\Gamma' \subseteq \mathcal{U}^{\mathcal{M}}$, Γ be the set of sentences corresponding to the utterances appearing in Γ', and $\phi_{c'} \in \mathcal{U}^{\mathcal{M}}$. Then if for all $\gamma_c \in \Gamma'$, $[\![\gamma_c]\!]^{\mathcal{M}^*} = t$, $[\![\phi_{c'}]\!]^{\mathcal{M}^*} \neq u$, and $\Gamma \models_{se} \phi$, then $[\![\phi_{c'}]\!]^{\mathcal{M}^*} = t$.*

Proof. If $[\![\gamma_c]\!]^{\mathcal{M}^*} = t$, $[\![\gamma]\!]^{\mathcal{M}_{\mathcal{M}^*}} = t$ by Proposition 5.1. Since $[\![\phi_{c'}]\!]^{\mathcal{M}^*} \neq u$, by the same token $[\![\phi]\!]^{\mathcal{M}_{\mathcal{M}^*}} \neq u$. Since $\Gamma \models_{se} \phi$, $[\![\phi]\!]^{\mathcal{M}_{\mathcal{M}^*}} = t$, so $[\![\phi_{c'}]\!]^{\mathcal{M}^*} \neq f$ and $[\![\phi_{c'}]\!]^{\mathcal{M}^*} = t$. □

This is a welcome result, because it shows that although significant alterations to one's logic may come with the accommodation of trivalence or generalized quantification, aside from token-sensitivity no further alterations to logical consequence accrue through the accommodation of semantic vocabulary. Whatever departures from more familiar logics are required in the system come in with tools that can be independently motivated in the need to adequately describe the operations of natural language.

Let's note some further simple properties we would expect the assignment $\mathcal{M}^*$ to exhibit. It is (I) a complete allotment of truth values (II) extending the original allotment provided by our base model, which (III) commutes as expected with negation and (IV) conjunction, and (V) validates all non-defective utterances of Strawson validities (where a Strawson validity is a sentence true when non-defective in all total base models from §1).

Proposition 5.3.

[20] See Shaw (forthcominga) for a discussion.

[21] The same would be true on the strong Kleene logic. Though not all propositional components would be truth-evaluable, we could render their elementary propositional components truth-evaluable consistently with the same truth-value assignments to logical compounds, since on the strong Kleene scheme changing a constituent value from u to t or f never changes the truth-value of a composite from t to f or *vice versa*.

(I) For all $\phi_c \in \mathcal{U}$, $[\![\phi_c]\!]^{\mathcal{M}^*} \in \{t, f, u\}$

(II) For all $\phi_c \in \mathcal{U}^{\mathcal{M}}$

$$\begin{aligned}
[\![\phi]\!]^{\mathcal{M}} = t &\quad\Rightarrow\quad [\![\phi_c]\!]^{\mathcal{M}^*} = t \\
[\![\phi]\!]^{\mathcal{M}} = f &\quad\Rightarrow\quad [\![\phi_c]\!]^{\mathcal{M}^*} = f \\
[\![\phi]\!]^{\mathcal{M}} = u &\quad\Rightarrow\quad [\![\phi_c]\!]^{\mathcal{M}^*} = u
\end{aligned}$$

(III) If $\phi_c = (\neg\psi_{c'})_c$ *then*

$$[\![\phi_c]\!]^{\mathcal{M}^*} = t \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = f$$
$$[\![\phi_c]\!]^{\mathcal{M}^*} = f \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = t$$

(IV) If $\phi_c = (\psi_{c'} \wedge \theta_{c''})_c$

$$[\![\phi_c]\!]^{\mathcal{M}^*} = t \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = [\![\theta_{c''}]\!]^{\mathcal{M}^*} = t$$
$$[\![\phi_c]\!]^{\mathcal{M}^*} = u \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = u \text{ or } [\![\theta_{c''}]\!]^{\mathcal{M}^*} = u$$

(V) If ϕ *is a Strawson validity, then if* $[\![\phi_c]\!]^{\mathcal{M}^*} \neq u$, $[\![\phi_c]\!]^{\mathcal{M}^*} = t$.

Proof. (I) and (II) are consequences of the fact that $\mathcal{M}^*$ is a total assignment extending $A_{\mathcal{M}}$. (III) and (IV) follow from the fact that $\mathcal{M}^*$ is coherent. (V) follows from Proposition 5.1. □

It is more difficult to state the results one would like for quantifiers, given that the assignment function does not apply to utterances that are not tokens of full sentence types. Nonetheless, we can get a basic intuitive result on quantifiers by using the connections between $\mathcal{M}^*$ and the model it engenders $\mathcal{M}^*_{\mathcal{M}}$ to exploit the satisfaction relation that we get from the latter engendered model. Once we do this, we see that non-defective quantified utterances have the truth-values one would expect.

Proposition 5.4.

(VI) If $\phi_c = (\exists v_1 : \psi(v_1))(\theta(v_1))_c$, *then provided* $[\![\phi_c]\!]^{\mathcal{M}^*} \in \{t, f\}$

$$\begin{aligned}
[\![\phi_c]\!]^{\mathcal{M}^*} = t \quad\Leftrightarrow\quad & \{a \in M^{\mathcal{M}} \mid [\![\psi(v_1)]\!]^{\mathcal{M}_{\mathcal{M}^*}, g[v_1 \to a]} = t\} \cap \\
& \{a \in M^{\mathcal{M}} \mid [\![\theta(v_1)]\!]^{\mathcal{M}_{\mathcal{M}^*}, g[v_1 \to a]} = t\} \neq \emptyset
\end{aligned}$$

(VII) If $\phi_c = (\forall v_1 : \psi(v_1))(\theta(v_1))_c$ *then provided* $[\![\phi_c]\!]^{\mathcal{M}^*} \in \{t, f\}$

$$\begin{aligned}
[\![\phi_c]\!]^{\mathcal{M}^*} = t \quad\Leftrightarrow\quad & \{a \in M^{\mathcal{M}} \mid [\![\psi(v_1)]\!]^{\mathcal{M}_{\mathcal{M}^*}, g[v_1 \to a]} = t\} \subseteq \\
& \{a \in M^{\mathcal{M}} \mid [\![\theta(v_1)]\!]^{\mathcal{M}_{\mathcal{M}^*}, g[v_1 \to a]} = t\}
\end{aligned}$$

Proof. The left-to-right direction of both (VI) and (VII) follow from proposition 5.1. The right-to-left directions follow once we see that if the right condition holds, then for all $\langle\beta,\alpha\rangle$, there is always a total coherent extension of A^{β}_{α} (namely $\mathcal{M}^*$) such that ϕ evaluates to true on the model engendered by that assignment. So if ϕ_c is ever boosted, it must boost to t. □

Of course, we don't merely want to note global results about the distribution of truth-values generally, but those relevant to the system as a theory of truth. Here are the two important results: (VIII) the truth value of a first utterance ascribing truth to a second utterance is, when non-defective, true just in case the second utterance is, and (IX) when ascribing semantic defect is non-defective, the ascription is true just in case the utterance to which defectiveness is ascribed is defective.

Proposition 5.5.

(VIII) If $\phi_c = (T\ulcorner\psi_{c'}\urcorner)_c$, *and* $[\![\phi_c]\!]^{\mathcal{M}^*} \neq u$, *then*

$$[\![\phi_c]\!]^{\mathcal{M}^*} = t \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = t$$

(IX) If $\phi_c = (U\ulcorner\psi_{c'}\urcorner)_c$, *and* $[\![\phi_c]\!]^{\mathcal{M}^*} \neq u$, *then*

$$[\![\phi_c]\!]^{\mathcal{M}^*} = t \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = u$$

Proof. Follows from the coherence of $\mathcal{M}^*$.

□

Some authors may have an interest in preserving something broadly like what Field (2008) calls 'the intersubstitutivity principle': that the result of substituting $T\ulcorner\phi\urcorner$ for ϕ, and vice versa, leads to logical equivalents. For the present system the closest analog would be:

(VIII*) If $\phi_c = (T\ulcorner\psi_{c'}\urcorner)_c$, then for all v

$$[\![\phi_c]\!]^{\mathcal{M}^*} = v \Leftrightarrow [\![\psi_{c'}]\!]^{\mathcal{M}^*} = v$$

This property can be achieved, I believe, by simply changing condition (ii) of the definition of engendered models (Definition 2.4) as follows.

(i) $T_t^{\mathcal{M}_A} = \{x : A(x) = t\}$

(ii$''$) $T_f^{\mathcal{M}_A} = \{x : A(x) = f\}$

This ensures that at ascriptions of T are boosted they always pick up the values of their ascribed utterances. If an ascription of T is culled or discarded, this still preserves (VIII*), since both the utterance ascribing truth and the utterance to which truth is ascribed will be simultaneously culled or discarded.[22]

I prefer the formulation on which ascriptions of truth to defective utterances count (when non-defective) as false. This allows, as we will shortly see, for the appropriate statement of semantic generalities. But there is no bar to retooling the truth-predicate as just noted, or even to including two truth-predicates, or a single truth-predicate whose attributions are context sensitive and may shift between the two interpretations.

Let's turn then to the next kind of statement which semantic extensions are particularly suited to accommodate, namely semantic generalities. Using $F\alpha$ as an abbreviation for $\neg T\alpha \wedge \neg U\alpha$, in order to talk of plain falsehood, we see that we can truly state the metalanguage analogs of (I)–(V), and (VIII)–(IX) above in the object language. That is, the system can do this provided the base language has the resources to express utterance-hood (S), being a truth, falsehood, or defective utterance of the base model $\mathcal{M}$ ($T_{\mathcal{M}}, F_{\mathcal{M}}, U_{\mathcal{M}}$), negation ($neg$), conjunction ($conj$), application of the truth and defectiveness predicates ($tapp$, $uapp$), and the form of Strawson validity ($Valid$).[23] The hindrance to object language statement of the analogs of (VI)–(VII) is the acknowledged lack of conditions on quantification for coherence, something I'll comment on further in §6.

Proposition 5.6. *Utterances ϕ_c of the following form are such that $[\![\phi_c]\!]^{\mathcal{M}^*} = t$ provided $\phi_c \in \mathcal{U}^{\mathcal{M}}$.*

(I′) $(\forall v_1 : Sv_1)(Tv_1 \vee Fv_1 \vee Uv_1)$

(II′) $(\forall v_1 : Sv_1 \wedge T_{\mathcal{M}}v_1)(Tv_1)$

$(\forall v_1 : Sv_1 \wedge F_{\mathcal{M}}v_1)(Fv_1)$

$(\forall v_1 : Sv_1 \wedge U_{\mathcal{M}}v_1)(Uv_1)$

(III′) $(\forall v_1, v_2 : Sv_1 \wedge Sv_2 \wedge v_1 = neg(v_2))((Tv_1 \leftrightarrow Fv_2) \wedge (Fv_1 \leftrightarrow Tv_2))$[24]

(IV′) $(\forall v_1, v_2, v_3 : Sv_1 \wedge Sv_2 \wedge Sv_3 \wedge v_1 = conj(v_2, v_3))(Tv_1 \leftrightarrow Tv_2 \wedge Tv_3)$

$(\forall v_1, v_2, v_3 : Sv_1 \wedge Sv_2 \wedge Sv_3 \wedge v_1 = conj(v_2, v_3))(Uv_1 \leftrightarrow Uv_2 \vee Uv_3)$

[22] The result is, as I say 'broadly like' an intersubstitutivity principle. It is often unobvious how to translate alleged desirable conditions from the type-based setting to the token-based setting, and Field's condition is no exception. I take my formulation here to capture the rough spirit of the condition. This should be fine, since trying to take it too seriously in the token-based setting leads to incoherent or false principles. For example, consider that it is not equally true to write, on the chalkboard C, an instance of the sentence type "The string "true" only appears once in the sentence on chalkboard C" and the result of 'substituting' in a truth-attribution ""The string "true" only appears once in the sentence on chalkboard C" is true". Also, regardless of whether we go for a context, or token-sensitive system, it is generally acknowledged that something closer to sentence tokens, rather than sentence types, are the proper bearers of truth, so in general the problem of transposing the intersubstitutivity criterion is a problem for the principle, not a problem with the token-based setting.

[23] Note: $tapp(v_1, v_2)$ reads: v_1 is the result of applying the truth-predicate to a term denoting v_2. Likewise for *uapp*.

[24] I'm here using the obvious abbreviations to simplify the notion of generalized quantification.

(V′) $(\forall v_1 : Sv_1 \wedge \neg Uv_1 \wedge Valid(v_1))(Tv_1)$

(VIII′) $(\forall v_1, v_2 : Sv_1 \wedge Sv_2 \wedge tapp(v_1, v_2) \wedge \neg Uv_1)(Tv_1 \leftrightarrow Tv_2)$

(IX′) $(\forall v_1, v_2 : Sv_1 \wedge Sv_2 \wedge uapp(v_1, v_2) \wedge \neg Uv_1)(Tv_1 \leftrightarrow Uv_2)$

Proof. All are easily seen to be assigned at A_1^0 by the definition of a total coherent assignment. □

As regards expressing object language equivalents of metalanguage statements about the logic of the system, expression is again hindered by the fact that coherent truth-value assignments have no conditions on quantification built into them. If we provisionally ignore this complication, focusing on the propositional logic of $\mathcal{M}^*$, we've seen that the Strawson valid propositional inferences are just the normal valid inferences of bivalent propositional logic. Thanks to the conditions on coherence, semantic extensions can 'see' this aspect of their consequence relation, as long as the base language can define the a bivalent propositional consequence relation, $PropCqn$ (though defined over utterances).

Proposition 5.7. *If ϕ_c is of the following form*

$$(\forall v_1, v_2 : Sv_1 \wedge Sv_2 \wedge PropCqn(v_1, v_2))(Tv_1 \wedge \neg Uv_2 \rightarrow Tv_2)$$

$[\![\phi_c]\!]^{\mathcal{M}^*} = t$*, provided* $\phi_c \in \mathcal{U}^{\mathcal{M}}$*.*

Proof. Again, this is seen to be assigned t at A_1^0 due to the conditions on coherence. □

The system would be able to 'see' and express its full logic, in the object language, if conditions for quantification were successfully built into the definition of coherence (though doing this successfully may of course be a non-trivial task).

Because of the shifting nature of the semantic dependence relations, the system is also able to capture somewhat more complex semantic generalities whose truth needn't turn on conditions of coherence. Here are two simple examples.

Proposition 5.8. *Let D be a formula of $\mathcal{L}$ with one free variable and without semantic vocabulary which, in $\mathcal{M}$, defines a set $\mathcal{D}$ such that*

$$\mathcal{D} \cap \mathcal{U}^{\mathcal{M}} \subseteq \{\psi_{c'} \mid [\![\psi_{c'}]\!]^{\mathcal{M}^*} = t\}$$

and let ϕ_c be

$$(\forall v_1 : Sv_1 \wedge Dv_1)(Tv_1)$$

Then $[\![\phi_c]\!]^{\mathcal{M}^*} = t$ *provided* $\phi_c \in \mathcal{U}^{\mathcal{M}}$*.*

Proof. If ϕ_c is assigned at a boosting stage the result follows, so we need only show that ϕ_c is not culled or discarded. If ϕ_c were culled at $\langle \beta, \alpha \rangle$, it would have semantic dependences on $A_{\omega_1}^{pred(\beta)}$ including unassigned utterances among the utterances in $\mathcal{D}$ (otherwise ϕ_c would have boosted to truth at an earlier stage). But then these utterances would be culled as well—contradicting our assumption. If ϕ_c were discarded it would await $A_0^{\omega_1}$ before assignment. But we know the utterances from $\mathcal{D}$ are all assigned before $A_0^{\omega_1}$. But then ϕ_c would have boosted to truth after all those utterances were assigned, and so couldn't be discarded. $\square$

Proposition 5.9. *Let D be a formula of $\mathcal{L}$ with one free variable and without semantic vocabulary which, in $\mathcal{M}$, defines a set $\mathcal{D}$ such that there is some $\psi_{c'}$ is such that $[\![\psi_{c'}]\!]^{\mathcal{M}^*} = t$ and*

$$\psi_{c'} \in \mathcal{D} \cap \mathcal{U}^{\mathcal{M}}$$

Then if ϕ_c is

$$(\exists v_1 : Sv_1 \wedge Dv_1)(Tv_1)$$

$[\![\phi_c]\!]^{\mathcal{M}^*} = t$ provided $\phi_c \in \mathcal{U}^{\mathcal{M}}$.

Proof. Let $\gamma_{c''}$ be the true utterance in $\mathcal{D}$ which is of lowest rank, and consider any $\langle \beta, \alpha \rangle < \text{rank}(\gamma_{c''})$. (If there are no such $\langle \beta, \alpha \rangle$, $\text{rank}(\gamma_{c''})$ is the initial stage and the claim follows immediately.) Consider the non-empty set

$$K = \{\kappa_c \in \mathcal{U}^{\mathcal{M}} | \kappa_c \in \mathcal{D},\ [\![\kappa_c]\!]^{\mathcal{M}^*} \in \{t, f\} \text{ and } \kappa_c \notin \text{dom}(A_\alpha^\beta)\}$$

By considering $\mathcal{M}^*$ and the weak extension of A_α^β we know that some superset K' of K matters to ϕ_c relative to A_α^β. Note that K' does not contain elements from

$$\{\kappa_c \in \mathcal{U}^{\mathcal{M}} | \kappa_c \in \mathcal{D} \text{ and } [\![\kappa_c]\!]^{\mathcal{M}^*} = u\}$$

Now, unless ϕ_c has unspecifiable dependences, some subset of K' really matters to ϕ_c, and whatever that subset is, it must contain elements from K. So whether or not ϕ_c has unspecifiable dependences it is semantically dependent, relative to A_α^β, on unassigned utterances that will be boosted to truth or falsity. This in turn means that ϕ_c is not boosted or culled at any of those stages. So, where $\text{rank}(\gamma_{c''}) = \langle \beta', \alpha' \rangle$, $\phi_c \notin \text{dom}(A_{\alpha'}^{\beta'})$. This means ϕ_c will then immediately boost to truth at the next stage. $\square$

6 Limitations and Imperfections of the Model

There are two main areas one might take issue with the system given: its general approach to the paradoxes, and its ability to express enough aspects of its own semantic structure.

As regards the handling of paradox, I will have little to say here. Whether paradoxical utterances embody a form of semantic defect, what this would imply, and how it should be modeled are complex questions which no formalism, no matter how desirable its features, will settle on its own. We need need an independent understanding of semantic defect and its source that engages questions about the nature and purpose of assertion. Such issues go well beyond the scope of this paper. But the formalism here does, as I have alluded to, embody ideas relevant to paradox developed and defended by myself and others elsewhere—for example the idea that we should adopt a kind of token-sensitive resolution to versions of two-line paradox. If those ideas and arguments are sound, then the present formalism would go a ways to showing that a conception of the meaning of the semantic vocabulary as procedurally governed, and coupled with a conception of paradox as a form of semantic defect, can be supported with a rigorous formalism with some intuitively desirable features.

But just how desirable are the features of the formalism, especially in areas of relevance to gauging the value of a theory of truth? There are three main such areas, as I see the project of giving a theory of truth: the logic of the system, the behavior of the truth predicate and other semantic vocabulary, and the expressive power of the system especially as regards its own semantic structure. The first two features of the system we have already seen: its logic is a form of Strawson entailment; applications of the truth-predicate to an utterance, when non-defective, track whether that utterance is true;[25] and attributions of semantic defect to an utterance, where themselves non-defective, are true just in case the utterance is defective.

But the final feature of relevance to truth—expressive power—may not be so obvious from the system's construction, and merits a few special remarks. The question of whether a theory of truth expresses everything we would like it to express can be divided into two sub-questions: does the system express every *concept* we would like it to express? And does the system express every *proposition* or *fact* we would like it to express.[26]

Take the question about concepts first. What kinds of conceptual expressive resources the system can in principle accommodate turns on whether the meaning of semantic terms is indeed given by a kind of procedure, in the sense elucidated by the semantics I have been elaborating. If the proper way of talking about truth, falsity, defectiveness, and other semantic properties is through a predicate whose use is procedurally governed, then there is good reason to think that the system (or some reasonable variant of it) does express every concept relevant to the system's operation, or at least could accommodate every concept of interest 'all at once'. After all, there are only three truth-values assigned in the process of generating the semantics, each of which has a procedurally governed predicate which is tailored to track the presence of those properties as well as could be done. The properties represented by the truth-values allotted in the semantics are, of course, not the only semantic properties there are. But there seems to be no in principle bar to accommodating as many semantic terms, with procedural rules, as one would like. Properties like *reference*, and *satisfaction* can be given such a treatment as well. Moreover *non-procedural* concepts, ex-

[25]As I noted, however, the rules governing the truth-predicate's semantics can easily be changed so that it always shares a value with its compliment utterance if desired.

[26]When we ask if a system does, or can, express every *concept* we want it to we must read the term "concept" broadly, to include not only the ability to talk about certain properties, but also to include logical concepts like negation. At least, we must do so if the first question is to combine with the second to be exhaustive. I won't be able to delve into the details of the latter kinds of logical concepts here.

pressed by predicates associated with standard extensions can be integrated as desired, subject of course to cardinality limitations. Talk of any particular set of utterances can be integrated in this non-procedural way into the system, though interpreting a predicate with a particular set of utterances will ensure that this set is not the set of truths, falsehoods, or defective utterances of the system. If the procedural conception of semantic terms is right this is no accident—in fact, this is just a manifestation of the fact that extensions as semantic values are not properly suited for the representation of semantic properties.[27]

When we turn from questions about the expressibility of concepts to the expressibility of *facts* or *propositions*, on the other hand, there are uncontroversial cases of expressive limitation present in the system. Some of these expressive limitations I am happy to concede are genuine imperfections, but others I have tried elsewhere to defend as essential limitations, and it is worth flagging the difference between them.

First off, though the system is geared towards the proper expression of semantic generalities, and the system goes a fair distance to that end as seen, for example, in Proposition 5.6, there are many generalities which elude proper assignment in the system I have presented. Examples include formalizations, in the system, of certain metalanguage statements: for example, "If a universally quantified statement is true, then every set of elements satisfying the quantifier restrictor also satisfies the quantifier matrix." These statements are true statements about the system, and there is every reason to hope or expect the system to be able to capture their truth 'from within'. As such the examples call for various kinds of emendations to the system. Statements about truth-value distributions require new, tighter requirements on the conditions of coherence, and statements about quantification require adopting truth-value assignments that range over sequence-utterance pairs, along with a careful formulation of the relevant added conditions on coherence. The added conditions on coherence must be such to preserve the truth of the coherence result of §4. Though many of these emendations are ones I have not ventured myself, I hope the formalism here can supply the sense that conditions tight enough to capture suitable 'desirable' formal constraints on a system can be found, which are also lax enough to be part of a coherence proof like the one I have given. I admit though, that in this regard the formalism I've given here is but a first step, and that without the amendments the system remains incomplete.[28]

[27] Again, see Shaw (forthcomingb) for a discussion of these issues.

[28] Another, perhaps more promising line is to divide up two jobs I have given to the truth-value assignments over which supervaluations are performed during the course of the construction. On the one hand these truth-value assignments are used to determine semantic dependence relations. On the other, they are used to help ensure the system captures appropriate sets of semantic generalities. Though it would distort the semantic dependence relations, at least given my gloss on them, it might be more helpful to allow supervaluations to range more broadly—perhaps as broadly as possible—over truth-value assignments extending an assignment one has so far in determining semantic dependences, and to relegate the task of assigning truth-values to semantic generalities to a separate rule in the formalism, akin to boosting or culling. This would still operate broadly in the spirit of the procedural conception of truth, but might ease the kinds of burdens that arise in proving a 'harmony' result as conditions on coherence becoming increasingly stringent. Of course the right kinds

But there are other kinds of facts about the allotment of truth-values that the system does not capture, which no simple amendments to the system would fix, and which I think the system *should not be expected* to 'fix'. These arise exclusively for certain kinds of defective utterances which object language utterances of the system cannot truthfully classify as defective. What kinds of statements might fit this description will depend on the range of utterances available in $\mathcal{U}^{\mathcal{M}}$, but let me quickly give two simple examples. One kind of defective statement that eludes classification as defective are those which 'artificially' situate themselves as high in the ranking as any other utterance ascribing them semantic properties. For example a formal equivalent of "anyone who says of what I'm now saying that it is false or defective, are saying something false or defective" could likely be defective, though no other object language utterance could truthfully report this.[29]

Another kind of utterance which eludes classification as defective are those in semantic chains which are discarded. Consider an infinite array of utterances, each one saying of the next that it is defective. My system claims each of these utterances is defective in some sense I haven't fully described here. We can say this in the metalanguage in which we are describing the features of the truth truth-value assignment given by $\mathcal{M}^*$, but it is clear that any object language attempt to say of an utterance in such a semantic chain that it is defective will simply become part of a semantic chain itself, dooming itself to defectiveness. This seems to point to an essential expressive limitation of the system—one which could only be avoided, if at all, by a wholesale rejection of large portions of the system and starting afresh, or giving up my aims of modeling a natural language like English, which has no essentially stronger metalanguage in terms of elementary semantic resources.

I acknowledge these expressive limitations, but I do not count them as a defect mandating rejection of my system or my aims. This is, I think, helpfully seen through consideration of discarded semantic chains as an example. It is an artifact of our abstract modeling situation that there is a position from which we can make a principled division between object language and metalanguage, and hence a position from which we can say where a lengthy semantic chain stretching between object and metalanguage is 'cut off', and reopened for the simple kinds of assignment witnessed in the boosting phases of my construction. In natural language use of semantic vocabulary, there is no principled separation between object language and metalanguage from which such a distinction can be drawn—no privileging of my use of "true" over yours, or privileging semantic talk in English over semantic talk in Italian or Igbo. There is no strategy I know of for a coherent and satisfying treatment of such semantic chains, in natural languages, as simply and intelligibly representational. There seems to be nowhere to begin.

So I see no hope of avoiding some such expressive limitations, but I cannot

of conditions for the new rule capturing generalities would have to be found. I leave the task of finding such conditions to future research.

[29]Compare some of the examples discussed by Gaifman (1992, 2000) under the heading of "holes" and "black holes".

hide the drastic consequence of this way of thinking about the relevant problematic utterances. To concede their existence in any model for natural language, and to conjointly make the claim that natural languages contain all the basic semantic expressive resources there are (say, tools for talking of truth-values), entails that there are, in some sense, ineffable truths about semantics—true pieces of information which cannot be truthfully reported, at least with the standard kinds of compositional mechanisms for the determination of content that I have been appealing to in my formalism. I myself am willing to face this as a consequence, and I have independently argued for its inevitability provided that we allow the truth predicate is governed by coherent norms.[30] No few words will do justice to this aspect of my program here, though, so I will have to be content to flag the unavailability of certain reports about semantic defect as essential expressive limitations of the system whose status as *imperfections* in the system is, I believe, up for dispute.

References

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. Chapuis & A. Gupta (eds.), *Circularity, Definition and Truth*, pp. 79–121. Indian Council of Philosophical Research.

M. Glanzberg (2001). 'The Liar in Context'. *Philosophical Studies* **103**(3):217–251.

M. Glanzberg (2004). 'A Contextual-Hierarchical Approach to Truth and the Liar Paradox'. *Journal of Philosophical Logic* **33**(1):27–88.

M. Glanzberg (2005). 'Truth, Reflection, and Hierarchies'. *Synthese* **142**(3):289–315.

M. Glanzberg (2006). 'Quantifiers'. *The Oxford Handbook of Philosophy of Language* pp. 794–821.

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

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.

[30]Shaw (2013).

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

J. R. Shaw (2013). 'Truth, Paradox, and Ineffable Propositions'. *Philosophy and Phenomenological Research* **86**(1):64–104.

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

J. R. Shaw (forthcomingb). 'Semantics for Semantics'. In B. Amour-Garb (ed.), *The Relevance of the Liar*. Oxford University Press.

J. R. Shaw (forthcomingc). 'What is a Truth-Value Gap?'. *Linguistics and Philosophy* .

A. Tarski (1956). 'The Concept of Truth in Formalized Languages'. *Logic, Semantics, Metamathematics* pp. 152–278.

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

S. Yablo (1993). 'Paradox without Self-Reference'. *Analysis* **53**(4):251–252.