THE REVIEW OF SYMBOLIC LOGIC
Volume 3, Number 3, September 2010

SUPER LIARS

PHILIPPE SCHLENKER
Institut Jean-Nicod and New York University

Abstract. Kripke's theory of truth offered a trivalent semantics for a language which, like English, contains a truth predicate and means of self-reference; but it did so by severely restricting the expressive power of the logic. In Kripke's analysis, the Liar (e.g., *This very sentence is not true*) receives the indeterminate truth value, but this fact cannot be expressed in the language; by contrast, it is straightforward to say in English that *the Liar is something other than true*. Kripke's theory also fails to handle the Strengthened Liar, which can be expressed in English as: *This very sentence is something other than true*. We develop a theory which seeks to overcome these difficulties, and is based on a detailed analysis of some of the linguistic means by which the Strengthened Liar can be expressed in English. In particular, we propose to take *literally* the quantificational form of the negative expression *something other than true.* Like other quantifiers, it may have different implicit domain restrictions, which give rise to a variety of negations of different strengths (e.g., *something other than true among the values* {*0, 1*}, or *among* {*0, 1, 2*}, etc). This analysis naturally leads to a logic with as many truth values as there are ordinals—a conclusion reached independently by Cook (2008a). We develop the theory within a generalization of the Strong Kleene Logic, augmented with negations that each have a nonmonotonic semantics. We show that fixed points can be constructed for our logic, and that it enjoys a limited form of 'expressive completeness'. Finally, we discuss the relation between our theory and various alternatives, including one in which the word *true* (rather than negation) is semantically ambiguous, and gives rise to a hierarchy of truth predicates of increasing strength.

§1. Introduction. Kripke (1975) succeeded in giving a semantics for a language that contains its own truth predicate and means of self-reference. But he remarked at the end of his article that the success was by no means complete, and that in the end 'the ghost of the Tarski hierarchy is still with us'. The heart of the matter is that Kripke's logic is successful only to the extent that it is expressively incomplete. Some of Kripke's sentences are true, others are false, and others are indeterminate (= neither true nor false). There is an obvious sense in which the latter are *not* true, but *this fact cannot be expressed in the language itself*—despite the fact that it is straightforward to say in English that, for instance, *the Liar is something other than true*. Kripke's theory also fails to handle the Strengthened Liar, which can be expressed as: *This very sentence is something other than true*. If the goal is to provide a semantics for *natural language*, Kripke's analysis falls short.

We will develop a theory which (i) seeks to overcome these expressive limitations, and which (ii) is based on a detailed analysis of some of the linguistic means by which the Strengthened Liar can be expressed in English. To keep the task manageable, we study fragments that reflect in simplified form some properties of English. In our main fragment, we take as our starting point the quantificational form of the negative expression *something other than true.* Like other quantifiers, it may have different implicit domain restrictions,

Received: May 31, 2009

© Association for Symbolic Logic, 2010

doi:10.1017/S1755020310000067

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

which give rise to a variety of negations of different strengths (e.g., *something other than true among the values* {*0, 1*}, or *among* {*0, 1, 2*}, etc). This analysis naturally leads to a logic with as many truth values (and negations) as there are ordinals—a conclusion reached independently by Cook (2008a). In our secondary fragment, we consider (somewhat in the spirit of Burge, 1979) a possible ambiguity in the meaning of the word *true* itself, due to pragmatically determined selectional restrictions; this yields an analysis with a hierarchy of truth predicates, which stands in a systematic relation to our main fragment.

To be concrete, let us consider the simple Liar, which can be seen as a sentence $\neg Tr(\lambda)$ named by a term λ. To simplify the exposition, we will write $<\lambda, \neg \mathrm{Tr}(\lambda) >$ to describe this situation: the first member of the pair, λ, is a term of the object language which, in the classical interpretation under study (call it i), denotes the second member, $\neg Tr(\lambda)$. We will also import the sentence-denoting terms into the meta-language by underlining them—so that $\underline{\lambda}$ is now *our* name for the object language sentence $\neg Tr(\lambda)$ (in other words, $\underline{\lambda} = i(\lambda) = \neg Tr(\lambda)$). It is clear, then, that $\underline{\lambda}$ must be indeterminate (or, as we will say, that it must have the value #). For Kripke, this causes no contradiction because his logic guarantees that $\neg F$ is indeterminate whenever F is. This is another way of saying that Kripke's negation is *strong*: it returns the value # when it applies to a formula with value #. But we can certainly define a weak negation $\neg^*$, which returns *true* just in case it applies to a formula which is false *or indeterminate*. However, weak negation cannot be defined in the Strong Kleene logic used by Kripke,[1] nor can it be added to it. This is because it would give rise to a 'Super Liar' $\underline{\lambda^*}$, which would force the theory to assert a contradiction.[2] Given the pair $<\lambda^*, \neg^*\mathrm{Tr}(\lambda^*)>$, it is clear that $\underline{\lambda^*}$ should be true if and only if it is false or indeterminate, and that it should be false if and only if it is true. But this immediately shows that $\underline{\lambda^*}$ cannot be assigned any of the three truth values if *Tr* is indeed interpreted as the truth predicate.

Kripke's logic is thus expressively incomplete. By itself, this need not be a problem. But it becomes one if the analysis is supposed to provide a semantics *for English*. Granted, the sentence in (1) can plausibly be assigned the indeterminate truth value:

(1) This very sentence is false.

The same applies to (2), though only on one of its readings, on which *not true* means the same thing as *false* (a point to which we return below):

(2) This very sentence is not true.

But there are definitely ways to express weak negation in English. First, the sentences in (3) are intuitively true, but they are inexpressible in Kripke's system:

(3) a. The sentence in (1) is something other than true.
b. The sentence in (1) has a value different from true.

Second, the same quantificational expressions—*something other than true*, *a value different from true*—also make it possible to express Super Liars which cannot be assigned any of Kripke's three truth values:

(4) a. This very sentence is something other than true.
b. This very sentence has a value different from true.

[1] This result follows because weak negation has a nonmonotonic semantics: $\neg^* F$ has the value 1 if F has the value # but it has the value 0 if F has the value 1. However, all the Strong Kleene connectives and operators are monotonic, and so is any formula defined from them.

[2] Super Liars are more commonly called 'Strengthened' or 'Revenge' Liars.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Kripke's theory is only successful for a very small fragment of English, one that includes sentences like (1) and (some readings of) (2), but not (3) or (4). Our goal is to provide an analysis that can handle a larger fragment of English. How should we proceed?

1. Part of the methodology is standard: as in all linguistic endeavors, our goal should be to derive as much as possible of speakers' truth-conditional intuitions about well-formed sentences of their language. It should go without saying that one must specify *which particular fragment* of the language one is trying to account for. Thus the sentences in (3) and (4) all involve the form of existential quantification over truth values (rather than standard negation), and the analysis of natural language quantification should accordingly play a role in the theory.
2. Still, part of the methodology is special to paradoxes. In other cases, the semanticist obtains his data from unreflective or 'intuitive' judgments volunteered by his informants. But this method is not applicable when it comes to paradoxes: it would seem that a speaker of English has to become a bit of a logician to come to terms with sentences like (1)–(4), and some conscious reasoning is certainly needed to assess them.[3] There are two sides to this observation:
   —First, it seems that the cognitive system that underlies a speaker's semantic intuitions is not 'designed' to deal with paradoxes. This should come as no surprise: neither in the linguistic development of the child nor in the biological evolution of the species are paradoxes likely to have played much of a role. Thus neither the innate nor the learned component of language is likely to have been shaped to treat paradoxes with anything like the speed with which other semantic phenomena are processed.
   —Second, the speaker can still come to *some* semantic judgments about sentences that involve paradoxes. And these judgments are undoubtedly constrained by the grammar of his language, though it might well be that the speaker must revise some of his *assumptions* to treat the problematic sentences—for instance he may have to relinquish the belief that there are just two truth values. He may also have to *extend* some of the semantic mechanisms of his language. One can view the Strong Kleene system as one such extension, whose treatment of indeterminacy is parasitic on that of classical values. Our main claim is that *once the semantic machinery of natural language has been extended to handle a third truth value, there is no way to block the introduction of many other truth values as well*; in fact, the linguistic means with which we characterize the semantic status of paradoxes seem to call for such an extension.
3. Still, it cannot be determined *a priori* whether we can give a semantics for *all* of English; nor whether the speaker's semantic intuitions are fully coherent. But we can certainly try to improve on Kripke's theory by accounting for a larger fragment of the language, one that includes sentences (3) and (4), and others like them—but the enterprise may break down somewhere.
4. Finally, the *data* to account for already indicate that we must depart from one of Kripke's main assumptions: we just cannot stick to the view that the logic is *monotonic*, in the sense that whatever classical values are obtained in an interpretation

[3] By contrast, speakers who hear the sentence *John knows that it is raining* when it is not raining can assess the sentence as being 'weird' (neither true nor false) without giving it much thought—and this observation extends to other standard cases of presupposition failure.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

> I are preserved when I is made 'more classical' by assigning classical values to atomic propositions which heretofore were indeterminate. For suppose that we wish to say about an atomic sentence *S* which is indeterminate that *it is something other than true*. This statement is intuitively true; however if the value of *S* is changed to 'true', the italicized statement should stop being true, and hence *something other than true* is a non-monotonic operator. Because monotonicity was crucial for Kripke to obtain his technical results, we will have to refine his methods in order to obtain comparable results.

In the rest of this paper, our main goal is to account for truth value judgments that are plausibly obtained by a speaker who reflects on a fragment of English that includes the expression *is something other than true* (or *has a value different from true*), as well as more standard connectives. We do not claim, however, that *is something other than true* is the only expression that yields weak negations and strengthened paradoxes. In fact, the simple statement *x is not true* can produce the same semantic effect on one of its readings, and we will explain how a relatively ordinary adjectival semantics for the word *true* can yield a semantic account of sentences such as (5), one in which negation is unitary but the truth predicate is ramified:

(5) This very sentence is not true.

The rest of this paper is organized as follows. In Section 1, we motivate our analysis of the negative expression *is something other than true* and define a fragment L with ordinal-many truth values and negations. In Section 2, we offer a rather different analysis for the expression *is not true*, and we argue that on one of its readings it can serve to define Super Liars as well; we show that a (simplified) version of the corresponding fragment, L*, can be translated back into L. In Section 3, we discuss some of the basic properties of the first fragment (L), and show that it achieves a limited form of expressive completeness (in this respect, it differs significantly from Kripke's logic). In Section 4, we prove that if L forms a set, a variety of fixed points can be found (which entails that fixed points for L* can be found as well). In Section 5, we discuss some extensions of the analysis to quantificational languages and to languages that do not form a set. We compare our approach to several contemporary analyses of the Liar in Section 6—in particular to Cook (2008a,b), who argued on independent grounds that there should be as many truth values as there are ordinals. The paper ends with some concluding remarks in Section 7.

§2. A fragment with is something other than true. The expressions *F is something other than true* or *F has a value different from true* offer particularly clear expressions of weak negation in English. In both cases, the negative expression includes a quantifier over truth values.[4] A proper semantic analysis must take this fact into account, and derive the meaning of these expressions from independent properties of natural language quantification. At the same time, however, we must recognize that the paradoxes force speakers to reconsider the assumptions they may normally make about the domain of possible truth values. One way to interpret our theory is to assume that a speaker who

[4] The quantification is explicitly over truth values in *has a value different from true.* For the expression *is something other than true*, we need to assume an implicit restriction to truth values to guarantee that the quantifier is only over truth values. We henceforth assume that such a restriction is present.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

reflects on paradoxical statements will *gradually expand the domain of truth values that he considers*. In this section, this fact is accommodated by positing that the implicit domain restriction on the quantifier *something other than true* (which for our purposes is equivalent to *has a value different from true*) will change as the speaker reflects on paradoxes (there might be other expressions of weak negation in English, and we will consider in Section 2 another way in which a related meaning can be obtained).

2.1. Something other than true.

2.1.1. Linguistic motivations If $\underline{F}$ is a sentence denoted by a name F, we consider the negation of $\underline{F}$ obtained by stating that *F is something other than true* or *F has a value different from true*. This ploy has one drawback: a sentence can be negated in this way only if it has a name in the object language. However this worry can be circumvented by resorting to *that*-clauses:

(6) a. That the US won the war in Irak is something other than true.
b. That the Liar is true is something other than true.
c. That this very sentence is true is something other than true.

Thus *is something other than true* or *has a value different from true* offer excellent—and fully general—expressions of weak negation.

As is well known, natural language quantifiers typically carry an implicit domain restriction, as shown by (7):

(7) Every Italian voted for the same applicant.

Clearly, (7) does not mean that every Italian *in the world* voted for the same candidate, but rather that every Italian within a given salient domain (for instance every Italian *on the committee*) voted for the same applicant. While it has sometimes been thought that this restriction can be handled by choosing a small enough universe for an entire sentence or discourse, this is not so. Different occurrences of the same quantifier can appear in one and the same sentence with different domain restrictions, as is suggested by the following example (it was pointed out to me by A. Szabolcsi and is originally due to D. Westerstahl):

(8) [*Situation:* A committee must select some applicants. Some of the applicants are Italian, and there are also Italians on the committee, though of course they are not the same.]
Every Italian voted for every Italian.

Here the first occurrence of the quantifier is naturally understood to refer to Italians *on the committee*, while the second occurrence refers to Italians *among the applicants*. One convenient way to implement this analysis is to include a domain-denoting expression on the determiner—which immediately makes it possible to impose different restrictions on various occurrences of the same quantifier:

(9) a. Every$_D$ Italian voted for every$_{D'}$ Italian.
b. (a) is true if and only if every Italian who is in the denotation of D voted for every Italian who is in the denotation of D'.

It is left to the context—that is, to the intentions of the speaker—to specify the values of D and D'.

When we apply this analysis to the expression *is something other than true* or *has a value different from true*, we obtain the following representations and truth conditions:

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(10) a. F is something$_D$ other than true.
b. F has a$_D$ value different from true.
c. (a) and (b) are true if and only if the denotation of F has a value in the denotation of D which is different from true.

It remains to specify what the denotation of D is. We believe that a speaker who is innocent of paradoxes will typically take the denotation of D to be {0, 1}, where 0 represents falsity and 1 represents truth.[5] This makes it straightforward to express a simple Liar:

(11) This very sentence is something$_D$ other than true.

But when our speaker is exposed to (11) in a context in which D denotes {0, 1}, he might conclude that there are more values than he initially included in D. If so, he will be led to 'add' to his universe a third truth value—call it 2—and to take (11), uttered in a context in which D denotes {0, 1}, to have value 2. He can then go on to say *truly* that (11) is something other than true by assigning to his domain variable D' the enlarged domain {0, 1, 2} rather than the original domain {0, 1}:

(12) The sentence in (11) is something$_{D'}$ other than true.

Needless to say, our speaker can now construct a strengthened Liar by uttering (11) in a context in which the intended denotation of D is the new domain {0, 1, 2}. As he reflects on the value of his utterance, he may conclude once again that there are more values than he previously thought, and he will add to his domain of discourse a fourth truth value, which we call 3—so that his utterance can have a value which is neither 0, nor 1, nor 2. The process can clearly be iterated.

Given the flexibility of domain restrictions (as seen in (9)), different occurrences of *is something other than true* may carry different domain variables, which should give considerable expressive power to the resulting logic, as we will see shortly. We will assume, however, that the choice of domain restrictions is not entirely unconstrained, and that if a speaker intends to include, say, the value 3, he must include all lower values as well (this is rather natural: why would one consider a domain restriction that includes 3 but not 0 or 2?).

2.1.2. Formal treatment and simplifications We could develop our analysis with respect to a realistic fragment of natural language, obtained with the words *and*, *or*, and all the negative expressions of the form *[That F] is something$_D$ other than true*, for various domain variables D (using *that*-clauses to guarantee that each sentence of the language, including those that have no name, can be negated). We would then have to specify in each case which value the context assigns to the domain variable. In order to keep the analysis manageable, however, we will make some simplifying assumptions.

(i) Instead of leaving it to the context to specify the value of the domain variables, we explicitly represent domain restrictions as constants in the object language.
(ii) Instead of carrying along the entire expression *is something$_D$ other than true*, we will represent it in our formal language as $\neg^{\alpha}$, where α is an explicit representation of the value of D. Thus we will have a series of negations $\neg^2$ (where 2 indicates

[5] This is only correct if we ignore the possibility of presupposition failure. When we do take it into account, further possibilities become available, as is discussed in Section 2.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

that the domain restriction is {0, 1}), $\neg^3$ (with domain restriction {0, 1, 2}), $\neg^4$ (with domain restriction {0, 1, 2, 3}), and so on.[6]

Let us immediately turn to some examples. $\neg^2 F$ represents the natural language sentence *[That F] is something*$_D$ *other than true*, where D denotes {0, 1}. Its semantics of $\neg^2$ is defined as follows:

(13) $I(\neg^2 F) = 1$ iff $I(F) = 0$
$I(\neg^2 F) = 0$ iff $I(F) = 1$
$I(\neg^2 F) = I(F)$ otherwise.

These truth conditions can be paraphrased by saying that *among the values 0 and 1, F has a value different from 1*. There is no doubt that the sentence should be true if F is false and that it should be false if F is true; and it is certainly plausible that if F has the indeterminate value 2, so does $\neg^2 F$. But what if F has, say, the value 3? Certainly F should have an indeterminate value—and crucially, one that does *not* lie within the implicit domain restriction. In the rule stated in (13), we chose to give $\neg^2 F$ the same (indeterminate) value as F when F has a value greater than 1. I do not think that this particular choice is justified by any linguistic data, and I also don't think that speakers have any semantic intuitions on this matter. It would be just as natural to decide that in this case the entire formula has value 2 whenever F has a value of at least 2; but we will see in Section 3.4.1 that this choice makes the logic inconsistent.

If 2 were the only nonclassical value, our semantics would yield a notational variant of Kleene's Strong logic (where the third truth value is called 2 rather than, say, #). But more interesting results are obtained when additional truth values are considered. To illustrate, let us a start with a simple Liar defined as $< \lambda^2, \neg^2 \mathrm{Tr}(\lambda^2) >$. It is clear that $\underline{\lambda}^2$ cannot coherently be assigned the values 0 or 1. On the other hand, it can coherently be assigned any value greater than 2, at least if $Tr(\lambda^2)$ has the same value as $\underline{\lambda}^2$ (which must be the case if the interpretation is a 'fixed point' in Kripke's sense, i.e., if it assigns to each formula *Tr(s)* the same value as to $\underline{s}$; we return below to the empirical adequacy of this condition). Given the last clause of our semantics for $\neg^2$, $\neg^2 Tr(\lambda^2)$ must have value $\beta \geq 2$ as soon as $Tr(\lambda^2)$ (and thus $\underline{\lambda}^2$) does, as desired.

Let us now define the semantics of a negation $\neg^3$, whose domain restriction is {0, 1, 2} rather than {0, 1}. Unlike $\neg^2$, $\neg^3$ returns the value 1 when it applies to a formula with value 2:

(14) $I(\neg^3 F) = 1$ iff for some $v \in \{0, 2\}$, $I(F) = v$
$I(\neg^3 F) = 0$ iff $I(F) = 1$
$I(\neg^3 F) = I(F)$ otherwise.

With our initial negation $\neg^2$, we could not state the fact that $\underline{\lambda}^2$ has a value other than 1, that is, that (in a sense) it is not true. We could have tried to state that $\neg^2 Tr(\lambda^2)$, but this formula is identical to $\underline{\lambda}^2$ itself, and thus it cannot have the value 1, as shown above. By contrast, we can state truly that $\neg^3 Tr(\lambda^2)$: as is desired, this formula comes out true if $\underline{\lambda}^2$

[6] The result is in agreement with a proposal sketched on different grounds in Fitch (1964), who suggested that a 'universal language' for philosophy should contain a hierarchy of increasingly stronger negations.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

has the value 2.[7] We can also define a Super Liar as $<\lambda^3, \neg^3\mathrm{Tr}(\lambda^3)>$, and it is immediate that $\underline{\lambda}^3$ cannot be assigned any of the values 0, 1, 2.

In the general case, then, we define the following semantics for $\neg^\alpha$ (for ordinal α):

(15) $I(\neg^\alpha F) = 1$ iff for some $v < \alpha$ with $v \neq 1$, $I(F) = v$
$I(\neg^\alpha F) = 0$ iff $I(F) = 1$
$I(\neg^\alpha F) = I(F)$ otherwise.

We note for future reference that for any $\alpha \geq 2$ we can define an α-Liar as $<\lambda^\alpha, \neg^\alpha\mathrm{Tr}(\lambda^\alpha)>$ (we just call it $\underline{\lambda}^\alpha$). And it is immediate that an α-Liar cannot be assigned any value smaller than α.

2.2. The fragment L. Let us now define our fragment. For simplicity, we take it to include propositional connectives, propositional letters, sentence names, and a truth predicate, but no quantifiers (it is designed to be as simple as possible while still allowing for a study of truth and self-reference). As in Kripke's construction, we start from a language L^- which does not contain a truth predicate. We add to L^- a truth predicate *Tr*, thus obtaining an expanded language L. J and J' are (possibly proper) proper class of ordinals that index the propositional letters and the sentence names. K is a (possibly proper) class of ordinals that index the many negations of the language (it is only if J, J', and K are all sets that L is one as well).

(16) Syntax of L^- and L
L^- is formed by rules a-c; L is formed by rules a-d.
a. Propositional constants: $p := p_i$ (for $i \in J$)
b. Sentence-denoting constants: $s := s_i$ (for $i \in J'$)
c. Formulas: $F := p \mid \neg^\alpha F \mid (F \wedge F) \mid (F \vee F)$
(for α any ordinal such that $\alpha \in K$ and $\alpha \geq 2$)
d. Formulas: Add to c. the rule: $F := \mathrm{Tr}(s)$.

As is customary, we will often omit the outermost parentheses of a formula. If a negation has an index i, we call i the *rank* of the negation. And we extend this terminology to formulas: if *F* is a formula, we say that it has rank 1 if it does not contain any negation; and for $i \geq 2$ we say that it has rank i if i the *highest index* found on the negations of F (in this case we write $\mathrm{rk}(F) = i$).

The semantics is defined in two steps. We start from a ground interpretation i for L^-. i is fully classical: it assigns to each propositional constant a *classical* value i(p) which belongs to {0, 1}, and it assigns to each sentence-denoting constant *s* a value i(s) which belongs to the class of sentences of L defined by (16). Since our semantics has as many truth values as there are ordinals, we extend i to an interpretation I for L by defining for each ordinal α an α-extension for *Tr*, which we write as $I^\alpha(\mathrm{Tr})$ (in some cases this may be a proper class rather than a set). Thus $I^0(\mathrm{Tr})$ is the class of objects (i.e., of sentences) that make *Tr* false; $I^1(\mathrm{Tr})$ is the class of objects that make *Tr* true; and for every ordinal $\alpha > 1$, $I^\alpha(\mathrm{Tr})$ is the class of objects that give *Tr* the indeterminate value α. For terminological simplicity,

[7] It should be noted that $\underline{\lambda}^2$ can coherently be assigned any value α greater than 2, in which case $\neg^3 Tr(\lambda^2)$ will have the value α; a negation $\neg^{\alpha'}$ of higher rank (with $\alpha' > \alpha$) would then be needed to express the desired claim.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

we call $I^0(Tr)$, $I^1(Tr)$, $I^2(Tr)$, and so on the *extensions* of *Tr*[8]. We further require that all extensions of *Tr* be mutually disjoint and exhaustive of the language L.

The semantics is thus defined as follows:

(17) Semantics for L
Let i assign a classical value (0 or 1) to every propositional letter, and for every ordinal α, let $I^\alpha(Tr)$ be a class of sentences of L, with the condition that for every sentence S of L, there is exactly ordinal α' such that S belongs to $I^{\alpha'}(Tr)$.
$I(p) = i(p)$
$I(Tr(s)) = \alpha$, where α is the (unique) ordinal for which I(s) belongs to $I^\alpha(Tr)$.
$I(F \wedge G) = 1$ iff $I(F) = I(G) = 1$; $I(F \wedge G) = 0$ iff $I(F) = 0$ or $I(G) = 0$; otherwise, $I(F \wedge G) = Max(I(F), I(G))$.
$I(F \vee G) = 1$ iff $I(F) = 1$ or $I(G) = 1$; $I(F \vee G) = 0$ iff $I(F) = I(F) = 0$; otherwise, $I(F \vee G) = Max(I(F), I(G))$.
$I(\neg^\alpha F) = 1$ iff for some ordinal $\beta < \alpha$ with $\beta \neq 1$, $I(F) = \beta$; $I(\neg^\alpha F) = 0$ iff $I(F) = 1$; otherwise, $I(\neg^\alpha F) = I(F)$.

We have already justified our semantics for negation. But something should be said about the semantics for *(F ∧ G)* and *(F ∨ G)*. In case at most one subformula has an indeterminate value β, our rules agree with the Strong Kleene scheme in which β is taken as the name of the third truth value. When both formulas have an indeterminate value, we assume that their disjunction and their conjunction each has the *greater* of the two values. Here too, there is no linguistic motivation for this decision, and we will show in Section 3.4.2 that other natural choices lead to inconsistency. As a terminological point, it should be observed that for us the indeterminate value 2 is *greater* than the classical values, whereas one usually says that in Kripke's logic the indeterminate value # is *smaller* than the classical values.

Since we remain within the general framework of Kripke's theory of truth, we will want to find interpretations that are 'fixed points' relative to the interpretation of the truth predicate:

(18) Fixed-Point Requirement
An interpretation I is a *fixed point* for L if for every sentence S of L, for every ordinal α, $S \in Tr^\alpha$ if and only if $I(S) = \alpha$.

We will prove the existence of fixed points in Section 4.

§3. A fragment with *is not true*.[9] Before we study the formal properties of the language L we just defined, it is worth considering the fragment (which we call L*) obtained when we focus on a different negative expression, namely *is not true*. In accordance with the methodology we espoused in the previous section, we base our analysis on the ordinary meaning that adjectives have in English. We will thus explore a different fragment, one that *also* makes it possible to define strengthened Liars, but does so with a unitary negation and

[8] In standard terminology, $I^1(Tr)$ is the extension of *Tr*, and $I^0(Tr)$ is its antiextension. As a referee suggested, for $\alpha \geq 2$, $I^\alpha(Tr)$ is a kind of 'pseudoextension' of *Tr*.

[9] This section was greatly influenced by discussions with Galinon. In Galinon (2006), he argued for a bivalent semantics with a plurality of truth predicates. By contrast, the first logic developed in this section has ordinal-many truth values, and also ordinal-many truth predicates. But we show in Section 3.4. how indeterminate values can be 'collapsed', which leads to a trivalent system with a hierarchy of truth predicates.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

a hierarchy of truth predicates (somewhat in the spirit of Burge, 1979). We will then show that there is a systematic correspondence between the resulting analysis and the hierarchy of negations obtained in Section 1 (as a result, fixed points for L can be transformed into fixed points for L*).

3.1. True.

3.1.1. Linguistic motivations Consider the adjective *tasty.* It has two properties that matter for our purposes.

1. First, *tasty* comes with a 'selectional restriction', that is, a presupposition that its subject is the kind of thing that one can taste, and in particular that it is edible (the presuppositional status of this inference can be established by considering questions—e.g., *Is this tasty?*—and embedding under negative expressions, e.g., *This isn't tasty*, *None of these things is tasty*; in all cases we obtain the same patterns of inference e.g., the inference that *this is edible* or that *each of these things is edible*—as for standard presupposition triggers, e.g., *stop smoking*[10]).

2. Second, *x is not tasty* is often understood to mean that *x has a bad taste ('contrary reading')*, even though the *form* of the sentence would lead one to expect the weaker meaning: *it is not the case that x is tasty* ('contradictory reading'). As reported in Horn (1989), this pattern is quite common: *John isn't happy* generally means that John is unhappy; and *I don't like Mary* is usually understood as *I dislike Mary*. In each case, we obtain a contrary reading where a compositional analysis would lead one to expect a contradictory reading. The *explanation* of these patterns has remained somewhat elusive; but several authors have explored a pragmatic analysis, whereby contrary readings arise just in case the speech act participants share a presupposition that, say, *Either John is happy or John is unhappy*.[11] Given this additional assumption, the assertion that *John is not happy* is indeed equivalent to *John is unhappy;* and the contrary reading of *this isn't tasty* can be obtained in a similar way.

Our suggestion is that the predicate *true* behaves somewhat like *tasty*: 1. It triggers a presupposition that its argument denotes an object with certain properties—namely a sentence that is true or false; and 2. It yields contrary readings where one would expect contradictory ones, which explains why *This sentence is not true* is often understood as synonymous with *This sentence is false*, *even* when one is convinced that there are more than two truth values. We will assume that the presupposition that the argument of *true* is a sentence with a classical truth value can be revised, which will give rise to a variety of (pragmatic) interpretations of *true*.[12] Specifically, *x is true* might originally presuppose that *x has the values 0 or 1*, then that *x has the values 0, 1, or 2*, then that *x has the values 0, 1, 2, or 3*, and so on. We will thus obtain as many semantic analyses as there are cycles in the learning process.

Importantly, the predicate *true* as we have characterized it does not satisfy Kripke's condition on 'fixed points'. Consider for instance a use of *true* with a selectional restriction to the values {0, 1, 2}—which we represent as $true_3$. Suppose further that we decided that an earlier use of the word *true* resulted in a sentence with value 2, for instance because it

[10] See Kadmon (2001) for a discussion of presupposition tests, and Chemla (2009) for experimental data.

[11] See Horn (1989) and Gajewski (2005) for discussion.

[12] Similarly, someone who learned that watches, tables, and chairs *can* in fact be eaten would probably revise *his* assumptions about the selectional restrictions of *tasty*.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

involved a simple Liar (i.e., a Liar with a use of *true* with selection restrictions to {0, 1}, written as $true_2$), as in (19):

(19) This very sentence is not $true_2$.

Now if we say that *(19) is* $true_3$, our claim should have the value 0 rather than 2. But this means that $true_3$ does *not* yield the value 2 when its argument denotes a sentence with value 2; in this respect it crucially differs from the Kripkean truth predicate we considered in Section 1.

3.1.2. Formal treatment and simplifications As in our earlier discussions, we will adopt some conventions to facilitate the technical analysis. Instead of letting the context specify which interpretation of *true* is intended (i.e., which selectional restrictions it has), we write Tr_α when the truth predicate carries a presupposition that its subject denotes a sentence whose value is smaller than α.[13] In order to obtain the intended semantics, the interpretation of Tr_α should satisfy a 'modified fixed-point requirement' as stated in (20) for a language L* whose truth predicates are indexed by K (a formal definition of L* is given below).

(20) Modified Fixed-Point Condition (Initial Version)
An interpretation I is a modified fixed point for L* if for every sentence S of L*, for every ordinal $\alpha \in$ K,
$S \in (\mathrm{Tr}_\alpha)^1$ if and only if $I(S) = 1$
$S \in (\mathrm{Tr}_\alpha)^0$ if and only if $I(S) = 0$ or $1 < I(S) < \alpha$
$S \in (\mathrm{Tr}_\alpha)^{\#}$ if and only if for $\beta \geq \alpha$, $I(S) = \beta$.

As before, if P is a predicate, P^β denotes the things that lie in the β-extension of P. Thus the notation $S \in (Tr_\alpha)^\beta$ means that S lies in the β-extension of the truth predicate Tr_α. This notation also applies to the value #, which encodes presupposition failure: $S \in (Tr_\alpha)^{\#}$ indicates that a presupposition failure is obtained when the argument of Tr_α denotes S.

For ease of comparison with the fragment L of Section 1, we will now make an important simplification. Instead of introducing yet another truth value in the system, namely #, we require that Tr_α return the value β if its argument has value β and $\beta \geq \alpha$. In effect, we treat in this case β as the 'error signal' produced by a presupposition failure, thus conflating two different sources of multivalence. There is no linguistic motivation for this move, but it has the advantage of simplifying the theory and the exposition. With this assumption, we must adapt the fixed-point requirement, as in (21):

(21) Modified Fixed-Point Condition (Simplified Version)
An interpretation I is a modified fixed point for L* if for every sentence S of L*, for every $\alpha \in$ K, for every ordinal β,
$S \in (\mathrm{Tr}_\alpha)^\beta$ if and only (i) $\beta = 1$ and $I(S) = 1$, or
(ii) $\beta = 0$ and ($I(S) = 0$ or $1 < I(S) < \alpha$), or
(iii) $\beta > 1$ and $\beta \geq \alpha$ and $I(S) = \beta$.

It follows from (21) that if I is a modified fixed point for L*, the set of sentences that are declared false by Tr_α is increasing in α, the set of sentences that are declared true is stable, and the set of sentences that are declared neither true nor false is decreasing in α:

[13] It must be granted, however, that it is not particularly natural to allow distinct occurrences of *true* that appear in the same sentence to carry different indices. The reader can take this to be a simplification, aimed at facilitating the comparison with the logic based on *is something other than true*.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(22) In any modified fixed point for L*, for $\alpha \in$ K, $(\mathrm{Tr}_\alpha)^0$ is increasing in α, $(\mathrm{Tr}_\alpha)^1$ is stable, and $\cup_{\gamma>1}(\mathrm{Tr}_\alpha)^\gamma$ is decreasing in α.

3.2. The fragment L*. We can now define our fragment L^*, derived from a classical language L^{*-} which has a classical interpretation i. J, J', and K are (possibly proper) classes of ordinals that index the propositional letters, the sentence names, and the truth predicates respectively. We note for future reference that the nonlogical vocabulary of L^{*-} is exactly the same as that of L^- (discussed in Section 1.2 in (16)).

(23) Syntax of L^{*-} and L*
L^{*-} is formed by rules a-c; L* is formed by rules a-d.
a. Propositional constants: $p := p_i$ (for $i \in J$)
b. Sentence-denoting constants: $s := s_i$ (for $i \in J$)
c. Formulas: $F := p \mid \neg F \mid (F \wedge F) \mid (F \vee F)$
d. Formulas: add to c. the rule: $F := \mathrm{Tr}_\alpha(s)$ (for α any ordinal such that $\alpha \in$ K and $\alpha \geq 2$).

(24) Semantics for L*
Let i assign a classical value (0 or 1) to every propositional letter, and for every $\alpha \in$ K and for every ordinal β, let $I^\beta(\mathrm{Tr}_\alpha)$ be a class of sentences of L, with the condition that for every sentence S of L*, there is exactly one ordinal β' such that S belongs to $I^{\beta'}(\mathrm{Tr}_\alpha)$.
$I(p) = i(p)$
$I(\mathrm{Tr}_\alpha(s)) = \beta$, where β is the (unique) ordinal for which I(s) belongs to $I^\beta(\mathrm{Tr}_\alpha)$.
$I(F \wedge G) = 1$ iff $I(F) = I(G) = 1$; $I(F \wedge G) = 0$ iff $I(F) = 0$ or $I(G) = 0$; otherwise, $I(F \wedge G) = \mathrm{Max}(I(F), I(G))$.
$I(F \vee G) = 1$ iff $I(F) = 1$ or $I(G) = 1$; $I(F \vee G) = 0$ iff $I(F) = I(F) = 0$; otherwise, $I(F \vee G) = \mathrm{Max}(I(F), I(G))$.
$I(\neg F) = 1$ iff $I(F) = 0$; $I(\neg F) = 0$ iff $I(F) = 1$; otherwise, $I(\neg F) = I(F)$.

It can be seen that conjunction and disjunction have the same semantics in L* as they did in the language L of Section 1. By contrast with L, however, L* has a unique negation, which can be seen to be a notational variant of the negation $\neg^2$ of L (it returns the value of its argument when the latter has an indeterminate truth value). The interpretive rule for $Tr_\alpha(s)$ is the same as that for *Tr* in the earlier system. The crucial difference between the two predicates lies in the fixed-point requirement, which now takes an entirely different form—and which depends on the subscript α carried by the truth predicate.

The Modified Fixed-Point Condition was stated in (21). But do modified fixed points exist for L*? The answer is positive, but we will obtain it indirectly. We show in the next paragraph that fixed points for L can be transformed into fixed points for L*. And we show in Section 4 that fixed points for L do indeed exist.

3.3. Translation from L* into L. Let us now see how fixed points for L can be turned into (modified) fixed points for L*. Since $\neg$ is a notational variant of $\neg^2$, we only need to worry about the various truth predicates Tr_α of L*.

Now if *Tr* in L has the intended interpretation, we can translate $Tr_\alpha(c)$ as $\neg^\alpha\neg^\alpha Tr(c)$.[14] Let us consider the latter expression in an interpretation I which is a fixed point for L.

[14] The same result would have been obtained with $\neg^2\neg^\alpha Tr(c)$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

- If $\underline{c}$ has a classical value, it is clear that the translation has the intended semantics.
- If $\underline{c}$ has value β with $1 < \beta < \alpha$, $\neg^{\alpha} Tr(c)$ has value 1 and $\neg^{\alpha}\neg^{\alpha} Tr(c)$ has value 0, as desired.
- If $\underline{c}$ has value β with $\beta \geq \alpha$, $\neg^{\alpha} Tr(c)$ has value β, and so does $\neg^{\alpha}\neg^{\alpha} Tr(c)$—which is also what we wanted.

We can thus define a translation $< . >$ from L* into L, and use it to transform fixed points for L into fixed points for L*.

(25) Translation from L* into L
$<p> = p$
$<Tr_{\alpha}(s)> = \neg^{\alpha}\neg^{\alpha} Tr(s)$
$<F \wedge G> = <F> \wedge <G>$
$<F \vee G> = <F> \vee <G>$
$<\neg F> = \neg^{2} <F>$.

(26) Let i be an interpretation of atomic expressions of L^{*-}. We define an interpretation j for L^{-} by specifying that j agrees with i on the nonlogical vocabulary (which is common to L^{*-} and to L^{-}), except for the interpretation of sentence names:
$j(s) = <F>$ iff $i(s) = F$.
Now we consider an interpretation J which extends j over L, and we use it to define an interpretation I* which extends i over L*. The interpretation of the truth predicates Tr_{α} in I* is defined by:
$S \in I^{*\beta}(Tr_{\alpha})$ iff $J(\neg^{\alpha}\neg^{\alpha} <S>) = \beta$.

If s is a sentence name, we write $I^{*}(s)$ as $\underline{s}$—and thus we have that $I^{*}(Tr_{\alpha}(s)) = J(\neg^{\alpha}\neg^{\alpha} <\underline{s}>)$. Furthermore, since $J(s) = <\underline{s}>$, if J is a fixed point for L, $J(\neg^{\alpha}\neg^{\alpha} <\underline{s}>) = J(\neg^{\alpha}\neg^{\alpha} Tr(s))$.

Claim: If J is a fixed point for L, then:

(i) for each sentence S of L*, $I^{*}(S) = J(<S>)$, and
(ii) I* is a modified fixed point for L*.

Proof.

(i) The proof is by induction on the construction of formulas of L*. The only nontrivial case is that of atomic formulas of the form $Tr_{\alpha}'(s)$. By (26) and the assumption that J is a fixed point, we have $I^{*}(Tr_{\alpha}'(s)) = J(\neg^{\alpha'}\neg^{\alpha'} <\underline{s}>) = J(\neg^{\alpha'}\neg^{\alpha'} Tr(s)) = J(<Tr_{\alpha}'(s)>)$. (The other cases are straightforward because the logical symbols on the left-hand side of the equalities in (25) have the same semantics as the corresponding symbols of the right-hand side).

(ii) For each sentence S of L*, $S \in I^{*\beta}(Tr_{\alpha})$ iff $J(\neg^{\alpha}\neg^{\alpha}<S>) = \beta$. By (i), $J(<S>) = I^{*}(S)$.

— If $\beta = 0$, $S \in I^{*\beta}(Tr_{\alpha})$ iff $J(\neg^{\alpha}\neg^{\alpha}<S>) = 0$, iff $J(\neg^{\alpha}<S>) = 1$, iff $J(<S>) = 0$ or $1 < J(<S>) < \alpha$, iff $1 < I^{*}(S) < \alpha$.
— If $\beta = 1$, $S \in I^{*\beta}(Tr_{\alpha})$ iff $J(\neg^{\alpha}\neg^{\alpha}<S>) = 1$, iff $J(\neg^{\alpha} <S>) = 0$ or $1 < J(\neg^{\alpha}<S>) < \alpha$, iff $J(<S>) = 1$ (since the second case cannot arise), iff $I^{*}(S) = 1$.
— If $\beta > 1$, $S \in I^{*\beta}(Tr_{\alpha})$ iff $J(\neg^{\alpha}\neg^{\alpha} <S>) = \beta$, iff $J(<S>) = \beta$ and $\beta \geq \alpha$ (by the semantics of $\neg^{\alpha}$), iff $I^{*}(S) = \beta$ and $\beta \geq \alpha$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

We leave a deeper investigation of L* for future research.[15] Starting in Section 3, we will concentrate on L, with the guarantee that some fixed points for L* can be obtained from fixed points for L.

3.4. Collapsing indeterminate values. For purposes of comparison (in particular with Burge's theory), it will be useful to note that we can reinterpret the logic of Section 2.2 by taking all the indeterminate values 2, 3, ... to be varieties of presupposition failure, which we can uniformly encode as #. We obtain in this way a trivalent logic with a hierarchy of truth predicates.

Specifically, let us define from an interpretation I for L* a trivalent interpretation $I^{\#}$ as in (27), with the convention that we write as $(\mathrm{Tr}_\alpha)^{\beta}_{\#}$ and $(\mathrm{Tr}_\alpha)^{\beta}$ the β-extensions of Tr_α according to $I^{\#}$ and I respectively.

(27) a. If s is a sentence name, $I^{\#}(s) = I(s) = \underline{s}$.
b. $I^{\#}(p) = I(p)$ iff $I(p) \in \{0, 1\}$; $I^{\#}(p) = \#$ iff $I(p) \notin \{0, 1\}$.
c. $(\mathrm{Tr}_\alpha)^{\beta}_{\#} = (\mathrm{Tr}_\alpha)^{\beta}$ if $\beta \in \{0, 1\}$; $(\mathrm{Tr}_\alpha)^{\beta}_{\#} = \cup_{\gamma > 1} (\mathrm{Tr}_\alpha)^{\gamma}$ if $\beta = \#$.
d. $I^{\#}(F \wedge G) = 1$ iff $I^{\#}(F) = I^{\#}(G) = 1$; $I^{\#}(F \wedge G) = 0$ iff $I^{\#}(F) = 0$ or $I^{\#}(G) = 0$; otherwise, $I^{\#}(F \wedge G) = \#$.
e. $I^{\#}(F \vee G) = 1$ iff $I^{\#}(F) = 1$ or $I^{\#}(G) = 1$; $I^{\#}(F \vee G) = 0$ iff $I^{\#}(F) = I^{\#}(F) = 0$; otherwise, $I^{\#}(F \vee G) = \#$.
f. $I^{\#}(\neg F) = 1$ iff $I^{\#}(F) = 0$; $I^{\#}(\neg F) = 0$ iff $I^{\#}(F) = 1$; otherwise, $I^{\#}(\neg F) = \#$.

It can be checked that $I^{\#}$ is indeed obtained from I by collapsing its indeterminate values into a single one, as is shown by the following property, which can be proven by induction:

(28) Let I be an interpretation for L* and let $I^{\#}$ be defined from I as in (27). Then for each sentence S of L*, $I^{\#}(S) = I(S)$ iff $I(S) \in \{0, 1\}$; $I^{\#}(S) = \#$ iff $I(S) \notin \{0, 1\}$.

It follows from (27)c that if the interpretation I for L* satisfies the Modified Fixed-Point Condition in (21), $I^{\#}$ will satisfy the condition in (29)[16]:

(29) a. $S \in (\mathrm{Tr}_\alpha)^{1}_{\#} = 1$ if and only $I^{\#}(S) = 1$.
b. If $S \in (\mathrm{Tr}_\alpha)^{0}_{\#}$, then $I^{\#}(S) = 0$ or $I^{\#}(S) = \#$.
c. If $\in (\mathrm{Tr}_\alpha)^{\#}_{\#}$, then $I^{\#}(S) = \#$.

Furthermore, from (27)c and (22) it follows that $(\mathrm{Tr}_\alpha)^{0}_{\#}$ is increasing in α, $(\mathrm{Tr}_\alpha)^{1}_{\#}$ is stable, and $(\mathrm{Tr}_\alpha)^{\#}_{\#}$ is decreasing in α. In other words, the predicates Tr_α are increasingly successful at declaring sentences to be nontrue as α increases.

---

[15] It would be interesting to find ways to translate L into L*. The natural suggestion would be to translate the Kripkean truth predicate Tr as Tr_2 (since in a modified fixed point for L*, $Tr_2(s)$must have the same value as $\underline{s}$ as soon as $\underline{s}$ has a value ≥ 2; and of course the condition is also satisfied when $\underline{s}$ has value 0 or 1). It would also be natural to translate $\neg^{\alpha} F$ as $\neg Tr_\alpha(c_F)$, where c_F denotes the translation of F; but this requires that each formula F of L be denoted by some constant c_F. Thus one could attempt to translate L into an expanded language L* ∪ {c_F: F is a formula of L}. We leave this enterprise for future research.

[16] By (27)c, $S \in (\mathrm{Tr}_\alpha)^{1}_{\#}$ iff $S \in (\mathrm{Tr}_\alpha)^{1}$, iff $I(S) = 1$ (because I satisfies the modified fixed-point condition in (21)), iff $I^{\#}(S) = 1$ (by (28)). By (27)c, if $S \in (\mathrm{Tr}_\alpha)^{0}_{\#}$, then $S \in (\mathrm{Tr}_\alpha)^{0}$, hence $I(S) = 0$ or $1 < I(S) < \alpha$—from which it follows that $I^{\#}(S) = 0$ or $I^{\#}(S) = \#$. Finally, by (27)c, if $S \in (\mathrm{Tr}_\alpha)^{\#}_{\#}$, $S \notin (\mathrm{Tr}_\alpha)^{0}_{\#}$, and $S \notin (\mathrm{Tr}_\alpha)^{1}_{\#}$. Since I satisfies the modified fixed point condition in (21), $I(S) > \alpha$, and by (28), $I^{\#}(S) = \#$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

§4. Basic properties of L. In this section we discuss some basic properties of the language L and some conceivable alternatives to the semantics we gave. We start with the connectives and then discuss general properties of the logic, in particular monotonicity (which fails) and expressive completeness (which holds in weakened form). We end the section with a discussion of an alternative semantics for the connectives, which leads to catastrophic results.

4.1. Properties of connectives: De Morgan's laws and iterations of negation.

4.1.1. De Morgan's laws Our semantics preserves certain features of the interdefinability of conjunction and disjunction in terms of negation. But the negation that must be used is $\neg^2$:

(30) De Morgan's laws
In any interpretation I,
a. $I(F \wedge G) = I(\neg^2(\neg^2 F \vee \neg^2 G))$.
b. $I(F \vee G) = I(\neg^2(\neg^2 F \wedge \neg^2 G))$.
In general, the identities fail if $\neg^2$ is replaced by some other negation.

Proof. In both cases, the result is immediate if the left-hand side has a classical value. If the left-hand side has an indeterminate value, it must be (in both cases) Max(I(F), I(G)). But a routine inspection shows that the right-hand side also has the value Max(I(F), I(G)) (this is because no matter what the indeterminate value of I(H) is, $I(\neg^2 H) = I(H)$).

Interdefinability fails, however, if we replace $\neg^2$ with a negation of higher rank. Suppose for instance that the outer most negations were replaced with $\neg^3$. In case both *F* and *G* have value 2, the right-hand sides of (30)a-b will have value 1, while the left-hand sides will have value 2.

In what follows we take $\wedge$ to be primitive and $\vee$ to be defined in terms of $\{\wedge, \neg^2\}$—which has the advantage of shortening some of our proofs.

4.1.2. Iterations of negation Some iterations of negation can be simplified according to the following rule:

(31) Iterations of negation
For any interpretation I, for any formula F and for any ordinals α and β:
a. $I(\neg^2\neg^2 F) = I(F)$ (though in general if $\alpha > 2$ $I(\neg^\alpha\neg^\alpha F) \neq I(F)$).
b. $I(\neg^\alpha\neg^\alpha\neg^\alpha F) = I(\neg^\alpha F)$.
c. if $\beta \leq \alpha$, $I(\neg^\beta\neg^\alpha F) = I(\neg^\alpha\neg^\alpha F)$.

Proof.
a. The result is obvious if $I(F) \leq 1$. Otherwise, $I(\neg^2\neg^2 F) = I(\neg^2 F) = I(F)$.
Note that in general the result fails for negations of higher rank. Thus if $I(F) = 2$, $I(\neg^3\neg^3 F) = 0$.
b. The result is obvious if $I(F) \leq 1$. If $2 \leq I(F) \leq \alpha$, $I(\neg^\alpha F) = 1$ and hence $I(\neg^\alpha\neg^\alpha\neg^\alpha F) = I(\neg^\alpha F)$. If $I(F) \geq \alpha$, $I(\neg^\alpha F) = I(F)$ and $I(\neg^\alpha\neg^\alpha\neg^\alpha F) = I(\neg^\alpha F)$.
c. We start by noting that for any formula F, $I(\neg^\alpha F) \geq \alpha$ or $I(\neg^\alpha F) \leq 1$. In either case $I(\neg^\beta\neg^\alpha F) = I(\neg^\alpha\neg^\alpha F)$. But in general, if $\beta < \alpha$, it is *not* the case that $I(\neg^\alpha\neg^\beta F) = I(\neg^\alpha\neg^\alpha F)$. Thus if $I(F) = 2$, $I(\neg^3\neg^2 F) = 1$ but $I(\neg^3\neg^3 F) = 0$.

4.2. Properties of the logic.

4.2.1. Semantic Restraint The logic handles indeterminate values with restraint. Specifically, the value of a formula is either (i) a classical value, or (ii) the indeterminate value of one of its atomic parts:

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(32) Semantic Restraint
Let V be a set of truth values. If all the atomic formulas of F have values in V, F has values in V ∪ {0, 1}.

Proof. The proof is by induction on the construction of formulas. The atomic case is trivial. And given De Morgan's laws, we only need to consider negation and conjunction. But it is immediate that if $I(\neg^{\alpha}G) \geq 2$ then $I(\neg^{\alpha}F) = I(F)$ and similarly that if $I(G \wedge H) \geq 2$ then $I(G \wedge H) = I(G)$ or $I(H)$.

4.2.2. Monotonicity

□ *Monotonicity in the Strong Kleene logic*

A property of the Strong Kleene logic which turned out to be crucial for Kripke's construction was its *monotonicity*. In essence, the idea is that when one changes the value of some indeterminates for classical values, one need not revise any classical values that one had heretofore assigned. To put it more formally, if I and I' are two expansions of the classical interpretation i obtained by assigning an extension and an antiextension to the truth predicate, then:

(33) Monotonicity
If $<I^0(Tr), I^1(Tr)> \subseteq <I'^0(Tr), I'^1(Tr)>$, if $I(F) \in \{0, 1\}$, then $I'(F) = I(F)$.[17]

□ *Failure of monotonicity*

Importantly, Monotonicity *fails* in the present logic. The culprit is negation. Suppose that $I(Tr(s)) = \beta$ with $1 < \beta < \alpha$ (thus $\underline{s}$ belongs to $I^{\beta}(Tr)$). Then $I(\neg^{\alpha}Tr(s)) = 1$. But now consider an interpretation I' which is identical to I, except that $\underline{s}$ is moved from $I^{\beta}(Tr)$ to $I^1(Tr)$. It is immediate that $I'(\neg^{\alpha}Tr(s)) = 0$. Thus although $<I^0(Tr), I^1(Tr)> \subseteq <I'^0(Tr), I'^1(Tr)>$, it may happen that $I(F) \in \{0, 1\}$ but $I'(F) \neq I(F)$, which refutes Monotonicity as standardly stated. We will see below that a weakened form of Monotonicity holds, which will be enough to construct some fixed points for our logic despite the failure of the 'usual' notion of monotonicity.[18]

4.2.3. Lack of a satisfactory notion of logical implication As noted in Galinon (2006), our system lacks a logically satisfactory notion of implication (a weakness it shares with the Strong Kleene scheme). If we define $p \rightarrow^{\alpha} q$ as $(\neg^{\alpha} p \vee q)$, we obtain the following truth conditions:

[17] The proof is by induction on the construction of formulas. (i) The property is obviously true of atomic formulas. (ii) If it is true of G and H, it is also true of $(G \wedge H)$: if $I(G \wedge H) = 1$, $I(G) = I(H) = 1$ and the property is vacuously satisfied; if $I(G \wedge H) = 0$, $I(G)$ or $I(H) = 0$ and the property is again verified. (iii) Finally, if the property holds of G, it also holds of $\neg G$ since the $I(\neg G) \in \{0, 1\}$ only if $I(G) \in \{0, 1\}$.

[18] This does not mean that a different ordering of the truth values couldn't give rise to a monotonic semantics. The ordering that we assume in this paragraph is one for which any classical value lies 'higher' than any indeterminate value. In Section 4.1, we define a different ordering, one in which *all* the values under a certain indeterminate β are 'higher' than β but unordered with respect to each other. We show that under certain conditions this gives rise to a monotonic semantics (which we call 'weak monotonicity') for certains parts of the language. Interestingly, this ploy is almost identical to that used by Cook (2008a) to construct fixed points for his own logic, which also includes ordinal-many truth values (instead of talking of 'weak monotonicity', Cook simply defines a notion of monotonicity with respect to a nonstandard ordering of truth values; nothing hinges on this terminological difference).

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(34) a. $I(F \rightarrow^{\alpha} G) = 1$ iff $I(\neg^{\alpha} F) = 1$ or $I(G) = 1$, iff $I(F) = 0$ or $1 < I(F) < \alpha$ or $I(G) = 1$.
b. $I(F \rightarrow^{\alpha} G) = 0$ iff $I(\neg^{\alpha} F) = I(G) = 0$, iff $I(F) = 1$ and $I(G) = 0$.
c. Otherwise, $I(F \rightarrow^{\alpha} G) = \text{Max}(I(\neg^{\alpha} F), I(G)) = \text{Max}(I(F), I(G))$.[19]

We obtain in this way a hierarchy of conditionals, which can be used to produce a hierarchy of Curry-style paradoxes.[20] We could try to motivate such a semantics on linguistic grounds.[21] But as an approximation of material implication in logic, $\rightarrow^{\alpha}$ yields some unintuitive results—for instance that p 'entails' q as soon as p has an indeterminate value less than α. And if we add that $p \leftrightarrow^{\alpha} q$ is defined $(p \rightarrow^{\alpha} q) \wedge (q \rightarrow^{\alpha} p)$, we get the amusing result that p and q are deemed materially equivalent as soon as both have a value different from 1 and less than α (for instance, $p \leftrightarrow^{6} q$ is true if p has value 5 and q has value 0).

Of course one could hope that a better notion of implication can be defined with the help of more complex formulas. But in the general case this is not so. As Galinon points out, if $\beta > \alpha$, any formula of rank α is bound to have value β whenever each of its atomic formulas has value β.[22] It follows that on no definition of $\rightarrow$ (no matter how complex) is it the case that $p \rightarrow p$ is valid (i.e., true in every model). This fact, which also holds in the Strong Kleene logic, is an important reason for claiming that 'nothing like sustained ordinary reasoning' can be carried out in the latter (Feferman, 1984; Field, 2003, 2008). The criticism would appear to apply just as well to the system developed here.

Two points should be borne in mind, however.

1. We started this paper with a linguistic project, which was to account for some *semantic intuitions* that speakers of English might have about various forms of the Liar; we primarily focused on the expression 'is something other than true', and secondarily on 'true'. We could have made it our goal to give a semantics for the word *if* as it is used in paradox-prone sentences—which would have meant yet another fragment of English. But there are no *a priori* grounds for thinking that conditionals involving paradoxical statements in

[19] The conditions for truth and falsity in (34)a-b are immediate. For (34)c, it cannot be that $I(F) = 0$ nor that $1 < I(F) < \alpha$, for if so we would be in case (ii)a; so we only have to derive the result in (ii)c for $I(F) = 1$ and $I(F) = 2$. If $I(F) = 1$, $I(\neg^{\alpha} F) = 0$ and $I(F \rightarrow^{\alpha} G) = I(G)$; since in case (34)c $I(F \rightarrow^{\alpha} G) = 2$, $I(G) = 2$ and $I(F \rightarrow^{\alpha} G) = I(G) = \text{Max}(I(F), I(G))$. If $I(F) = \alpha$, $I(\neg^{\alpha} F) = I(F)$ and thus $I(F \rightarrow^{\alpha} G) = \text{Max}(I(F), I(G))$.

[20] Let us define, for every $\alpha \geq 2$, $< \gamma_{\alpha}, \text{Tr}(\gamma_{\alpha}) \rightarrow^{\alpha} F>$, where F is a contradiction. It is immediate that $Tr(\gamma_{\alpha})$, and hence $\underline{\gamma}_{\alpha}$, cannot be given a value β with $\beta = 1$ (as this would entail the truth of F), nor with $1 < \beta < \overline{\alpha}$ (as this would make $\underline{\gamma}_{\alpha}$ true, contradicting $\beta \neq 1$). On the other hand, no contradiction arises for $\beta = \alpha$.

[21] The truth conditions in (34) have some initial plausibility if we consider that *if F, G* is typically treated by linguists as a quantificational structure, with the meaning: *every situation that satisfies F (within a given domain) also satisfies G* (but see Kratzer, 1991; von Fintel, 2001; and Schlenker, 2004, for various refinements). In the spirit of our analysis of negation, we could posit that *if F, G* comes with is an implicit restriction to situations *in which F has a value* $< \alpha$. When this condition is met, the conditional is true just in case whenever $I(F) = 1$, $I(G) = 1$. Following the spirit of the Strong Kleene logic, if the condition *fails*, the conditional can still be taken to be true in case its consequent is (i.e., $I(G) = 1$). But if $I(F) = \alpha$ and $I(G) \neq 1$, a nonclassical value is obtained—which yields something very close to the truth conditions in (34) (the nonclassical value obtained is $\text{Max}(I(F), I(G))$. We leave a deeper discussion of conditionals for future research.

[22] The proof is by induction on the construction of formulas, assuming $\beta > \alpha$. The atomic case is trivial. If G and H have value β, so does $(G \wedge H)$. Finally, if G has value β and if $\neg^{\gamma} G$ occurs in F, $\gamma = \alpha < \beta$ (since $\text{rk}(F) = \alpha$), and thus $\neg^{\gamma} G$ has value β as well.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

English display anything like the strong semantics that Feferman and Field desire. They seem to be asking a question about how people *reason* about paradoxes, rather than about natural language conditionals used to express such reasonings. The two enterprises are conceptually distinct, and there is no guarantee that they will converge.[23]

2. Still, if we are willing to restrict attention to a *finite* number of indeterminate truth values, the problem of finding an adequate notion of implication can be solved in a satisfactory fashion, thanks to a more general result. Under broad conditions, our logic enjoys a limited form of expressive completeness, which guarantees that *any truth function defined over a finite set of values can be expressed by some formula*. In particular, any notion of implication that one wishes to define over a finite set of truth values will turn out to be expressible. Needless to say, this result does not invalidate Galinon's observation that no definition of $\rightarrow$ can make $p \rightarrow p$ valid. As we will see in the next paragraph, it is only in case we stipulate that p takes its value within a finite set V that we can find a definition of implication on which $p \rightarrow p$ is always true (if p is assigned an arbitrarily large indeterminate value, $p \rightarrow p$ will end up indeterminate as well).

4.3. Limited expressive completeness. We saw earlier that Kripke's theory of truth was expressively incomplete: there is a connective, namely weak negation, which cannot be defined within the Strong Kleene logic favored by Kripke. What is the situation in the logic we just defined? Despite Galinon's limitative results, a weakened form of expressive completeness can be achieved, as shown by the following theorem.

THEOREM 4.1. *Let* V *be a finite set of truth values, and suppose that for every* α *in* V *the ground interpretation i allows for the definition of the formulas* $\underline{\chi}^{\alpha}$ *defined below.*[24] *Then any n-ary function on V (for finite* n*) can be expressed by a formula of the logic. Specifically, if* f *is an n-ary function on* V*, there exists a formula* F *such that:*

(a) if $a_1, \ldots, a_n \in$ V *then for each fixed point* I *in which (i) all atomic propositions of* F *take their values in* V *and (ii)* $\mathrm{I}(p_1) = a_1, \ldots, \mathrm{I}(p_n) = a_n$*, it is also the case that* $f(a_1, \ldots, a_n) = \mathrm{I}(\mathrm{F})$

(b) the rank of F is at most $\mathrm{Max}(\mathrm{V}) + 1$.

This result will prove important when we compare our account to some alternatives in Section 6. For a logic that has a finite number of truth values, limited expressive completeness is the same thing as standard expressive completeness—and for a rich enough language the latter cannot be obtained on pain of inconsistency. In this respect, our semantics is crucially different: although it does not achieve full expressive completeness, it enjoys a limited form of it.

The proof of the theorem is in two steps. First, we show that under certain conditions any ordinal α can be characterized by a formula $\mathrm{C}^{\alpha}[p]$, in the sense that $\mathrm{C}^{\alpha}[p]$ has value 1 just in case p has value α, and has value 0 otherwise. Second, we use this result to provide a description in the object language of any truth table with values in V.

---

[23] Still, one could try to add to our logic a primitive notion of implication—not definable from other connectives—which is 'as strong as possible' (possibly along the lines of Field 2003a,b, 2007, 2008). Alternatively, one may define a hierarchy of implications of increasing strengths, as is done by Cook (2008a) in a system that shares important features with the present one. See Sections 6.3 and 6.4 for discussion.

[24] If necessary additional sentence-denoting constants can be added to the language to ensure that this is indeed the case.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

LEMMA 4.2. *Let* α *and* β *be two ordinals with* $2 \leq \alpha < \beta$. *Then some formula* $C^\alpha[p]$ *(of rank* β *with a single propositional variable p) characterizes* α *over the set* β, *in the sense that if* $I(p) < \beta$, *then (a)* $I(C^\alpha[p]) = 1$ *iff* $I(p) = \alpha$, *and (b)* $I(C^\alpha[p]) = 0$ *iff* $I(p) \neq \alpha$.

Proof of the lemma. We assume throughout that $\alpha < \beta$.

a. If $\alpha = 1$, set $C^\alpha[p] := (p \wedge \neg^2\neg^\beta p)$.
We note that if $I(p) = 1$, $I(C^\alpha[p]) = 1$ as well. If $I(p) \neq 1$, $I(\neg^\beta p) = 1$ and hence $I(\neg^2\neg^\beta p) = 0$, and $I(C^\alpha[p]) = 0$ as well.

b. If $\alpha = 0$, set $C^\alpha[p] := (\neg^2 p \wedge \neg^2\neg^\beta\neg^2 p)$.
If $I(p) = 0$, $I(C^\alpha[p]) = 1$. If $I(p) = 1$, $I(\neg^2\neg^\beta\neg^2 p) = 0$ and thus $I(C^\alpha[p]) = 0$ as well. Finally, if $I(p) = \beta'$ with $1 < \beta' < \beta$, $I(\neg^2 p) = \beta'$, and $I(\neg^\beta\neg^2 p) = 1$, hence $I(\neg^2\neg^\beta\neg^2 p) = 0$. Thus $I(C^\alpha[p]) = 0$ as well.

c. If $1 < \alpha < \beta$, set $C^\alpha[p] := (\neg^{\alpha+1}(\neg^\alpha p \vee p) \wedge \neg^2\neg^\beta\neg^{\alpha+1} p)$.

— If $I(p) = \alpha$, $I(\neg^\alpha p) = \alpha$ and thus $I(\neg^\alpha p \vee p) = \alpha$, and $I(\neg^{\alpha+1}(\neg^\alpha p \vee p)) = 1$. Since we also have $I(\neg^2\neg^\beta\neg^{\alpha+1} p) = 1$ (because $I(\neg^{\alpha+1} p) = 1$), it follows that $I(C^\alpha[p]) = 1$, as is desired.

— If $I(p) = 0$ or 1, $I(\neg^{\alpha+1}(\neg^\alpha p \vee p)) = 0$, and hence $I(C^\alpha[p]) = 0$.

— If $1 < I(p) < \alpha$, $I(\neg^\alpha p) = 1$, and hence $I(\neg^\alpha p \vee p) = 1$. It follows that $I(\neg^{\alpha+1}(\neg^\alpha p \vee p)) = 0$, and again that $I(C^\alpha[p]) = 0$.

— If $\alpha < I(p) < \beta$, $I(\neg^{\alpha+1} p) = I(p)$, hence (since $\alpha+1 \leq I(p) < \beta$) $I(\neg^\beta\neg^{\alpha+1} p) = 1$ and $I(\neg^2\neg^\beta\neg^{\alpha+1} p) = 0$. It follows that $I(C^\alpha[p]) = 0$.

Proof of the theorem. Let f be a function which takes a (fixed) number n of ordinal arguments and returns an ordinal argument within a given finite set. We can think of f as a table in which the possible values of the arguments $a_1, \ldots, a_n$ are arranged in successive columns, followed by the value of $f(a_1, \ldots, a_n)$.

	a_1	a_2	...	a_n	$f(a_1, \ldots, a_n)$
L_1					1
L_2					5
L_3					1
...					...
					0
L_m					1

In order to find a formula that defines f, we start by describing *in the object language* the situations that give rise to a *1* in the last column. In general this will be a disjunction; thus in the table as represented, a *1* is obtained if one is in the configuration of L_1 or in the configuration of L_3 or in the configuration of L_m, and so on. We do the same thing for lines that have an indeterminate truth value α in the last column: we find a formula of the language which is true *precisely* in case 1 is in a situation that corresponds to any of these lines (and which is false otherwise). But we will add to this formula a conjunct that ensures that in precisely these cases the entire formula has value α.

We define $\beta = \text{Max}(V) + 1$ and we use the notations of the lemma.

Case 1. Describing lines that have a 1 in their last cell.

We define:

$F_1 := \vee_{f(a_1,\ldots,a_n)=1} (\wedge_{1 \leq i \leq n} C^{a_i}[p_i])$.

Case 2. Describing lines that have an indeterminate truth value in their last cell.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

We define:
for $\alpha \in V$ with $\alpha \notin \{0, 1\}$, $F_\alpha := (\vee_{f(a_1,\ldots,a_n)=\alpha}\ (\wedge_{1\leq i\leq n} C^{a_i}\ [p_i])) \wedge \underline{\chi}^\alpha$
where the ground interpretation allows for the definition of $\underline{\chi}^\alpha$:
$< \chi^\alpha, \neg^\alpha \mathrm{Tr}(\chi^\alpha) \wedge \wedge_{\gamma \in V_{\text{and } \gamma\neq\alpha \text{and } \gamma\neq 0}} \neg^2 C^\gamma\ [\mathrm{Tr}(\chi^\alpha)] >$.
Adequacy of the formula

1) Let us reason by cases to show that if I is a fixed point and if $Tr(\chi^\alpha)$ has a value in V, then $\underline{\chi}^\alpha$ has value α.

   — If $I(\underline{\chi}^\alpha) = \alpha$, no contradiction is derived: the first conjunct has value α, and the second conjunct has value 1.
   — If $I(\underline{\chi}^\alpha) = 0$, $I(\neg^\alpha \mathrm{Tr}(\chi^\alpha)) = 1$ and $I(\wedge_{\gamma \in V_{\text{and } \gamma\neq\alpha \text{ and } \gamma\neq 0}} \neg^2 C^\gamma\ [\mathrm{Tr}(\chi^\alpha)]) = 1$. Contradiction.
   — If $I(\underline{\chi}^\alpha) = \gamma$ with $\gamma \in V$, $\gamma \neq 0$ and $\gamma \neq \alpha$, $I(C^\gamma[\mathrm{Tr}(\chi^\alpha)]) = 1$ and $I(\wedge_{\gamma\in V_{\text{and } \gamma\neq\alpha \text{ and } \gamma\neq 0}} \neg^2 C^\gamma [\mathrm{Tr}(\chi^\alpha)]) = 0$. But this entails that $I(\underline{\chi}^\alpha) = 0$. Contradiction.
   If I is a fixed point, then, $\underline{\chi}^\alpha$ can only have value α. (It must be pointed out, however, that this result holds *only* on the assumption that the atomic formulas of $\underline{\chi}^\alpha$ take their values within V. Without this constraint we cannot prevent $\underline{\chi}^\alpha$ from taking any value γ greater than β when we would want it to take value α instead).

2) We define $F := \vee_{\alpha\in V \text{ and } \alpha\neq 0} F_\alpha$ with the F_α's as described above, that is,
   $F_1 := \vee_{f(a_1,\ldots,a_n)=1}\ (\wedge_{1\leq i\leq n} C^{a_i}\ [p_i])$
   and for $\alpha \in V$ with $\alpha \notin \{0, 1\}$, $F_\alpha := (\vee_{f(a_1,\ldots,a_n)=\alpha}\ (\wedge_{1\leq i\leq n} C^{ai}\ [p_i])) \wedge \underline{\chi}^\alpha$.
   We reason by cases to show that the translation is adequate.

   — If $f(a_1,\ldots, a_n) = 0$, we are in none of the configurations described by any F_α. So for each $\alpha \in V$ and $\alpha \neq 0$, $I(F_\alpha) = 0$, and thus $I(F) = 0 = f(a_1,\ldots, a_n)$.
   — If $f(a_1,\ldots, a_n) = 1$, $I(F_1) = 1$ and thus $I(F) = 1 = f(a_1,\ldots, a_n)$.
   — If $f(a_1,\ldots, a_n) \in V$ and $f(a_1,\ldots, a_n) \notin \{0, 1\}$, let us call $\alpha^* := f(a_1,\ldots, a_n)$. For every α such that $\alpha \in V$, $\alpha \neq 0$ and $\alpha \neq \alpha^*$, $I(F_\alpha) = 0$. Thus each disjunct F_α of F for $\alpha \neq \alpha^*$ has value 0. Furthermore, the first conjunct of F_{α^*} has value 1, and the second conjunct has value α^*, so F_{α^*} has value α^*. It follows that $I(F_{\alpha^*}) = \alpha^*$ and thus that $I(F) = \alpha^* = f(a_1,\ldots, a_n)$.

4.4. (Im)possible variants.

4.4.1. Alternative definition of negation One could be tempted to give a different semantics for negation. In the previous paragraphs we decided that $\neg^\alpha F$ must have the same value as F when F has a value of at least α. What would happen if we decided instead that in the latter case $\neg^\alpha F$ must uniformly have value α? The resulting logic might have no fixed points.

Let us call $\underline{\neg}^\alpha$ this alternative negation which, as its 'kosher' counterpart, comes in as many varieties as there are ordinals. Unlike $\neg^\alpha$, $\underline{\neg}^\alpha$ does not (cautiously) return the value of its argument F when F has a value greater than α; rather, in such a case it returns the value α:

(35) $I(\underline{\neg}^\alpha F) = 1$ iff for some ordinal $\beta < \alpha$ with $\beta \neq 1$, $I(F) = \beta$; $I(\underline{\neg}^\alpha F) = 0$ iff $I(F) = 1$; otherwise, $I(\underline{\neg}^\alpha F) = \alpha$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

To see that in general a language with $\underline{\neg}^{\alpha}$ does not have any fixed point, consider the pair $< \mu, \underline{\neg}^3\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu) >$. We assume that I is a fixed point and we reason by cases to obtain a contradiction:

i. If $I(\underline{\mu}) = 1$, then $I(\underline{\neg}^2\mathrm{Tr}(\mu)) = 0$ and thus $I(\underline{\neg}^3\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu)) = 0$. Contradiction.

ii. If $I(\underline{\mu}) = 0$, then $I(\underline{\neg}^2\mathrm{Tr}(\mu)) = 1$ and thus $I(\underline{\neg}^3\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu)) = 1$. Contradiction.

iii. If $I(\underline{\mu}) \geq 2$, then $I(\underline{\neg}^2\mathrm{Tr}(\mu)) = 2$, $I(\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu)) = 2$, and $I(\underline{\neg}^3\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu)) = 1$. Contradiction again.
In fact, the point is more general: if $\beta > \alpha$, we can replace $\underline{\neg}^3\underline{\neg}^2\underline{\neg}^2\mathrm{Tr}(\mu)$ with $\underline{\neg}^{\beta}\underline{\neg}^{\alpha}\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)$ and obtain exactly the same conclusion. Lines i and ii are the same as before, and we replace line iii with iii':

iii'. If $2 \leq I(\underline{\mu}) < \alpha$, $I(\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)) = 1$, hence $I(\underline{\neg}^{\alpha}\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)) = 0$ and $I(\underline{\neg}^{\beta}\underline{\neg}^{\alpha}\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)) = 1$. Contradiction. If $I(\underline{\mu}) \geq \alpha$, $I(\underline{\neg}^{\alpha}\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)) = \alpha$ and $I(\underline{\neg}^{\beta}\underline{\neg}^{\alpha}\underline{\neg}^{\alpha}\mathrm{Tr}(\mu)) = 1$. Contradiction again.
By contrast, we will see shortly that some fixed points *can* be found for the logic as we have defined it.

4.4.2. Alternative definition of conjunction (and disjunction) Suppose now that we attempted to replace *Max(F, G)* with *Min(F, G)* in the semantics for conjunction given in (17). We call $\underline{\wedge}$ the 'new' conjunction, defined in (36).

(36) $I(F \underline{\wedge} G) = 1$ iff $I(F) = I(G) = 1$; $I(F \underline{\wedge} G) = 0$ iff $I(F) = 0$ or $I(G) = 0$; otherwise, $I(F \underline{\wedge} G) = \mathrm{Min}(I(F), I(G))$.

Under certain conditions, this semantics will make it impossible to find fixed points.

We start by noting that we can define from $\underline{\wedge}$ and $\neg^2$ a Min-based version of disjunction, as shown in (37) (and we could just as well define $\underline{\wedge}$ from $\underline{\vee}$ and $\neg^2$).

(37) a. *Definition:* $(F \underline{\vee} G) := \neg^2(\neg^2 F \underline{\wedge} \neg^2 G)$.
b. *Result:* $I(F \underline{\vee} G) = 1$ iff $I(F) = 1$ or $I(G) = 1$; $I(F \underline{\vee} G) = 0$ iff $I(F) = I(G) = 0$; otherwise, $I(F \underline{\vee} G) = \mathrm{Min}(I(F), I(G))$.

(The result is immediate if there is a most one indeterminate value in $\{I(F), I(G)\}$. If both *F* and *G* are indeterminate, $I(\neg^2 F) = I(F)$ and $I(\neg^2 G) = I(G)$ and thus $I(F \underline{\vee} G) = I(\neg^2(F \underline{\wedge} G)) = I(F \underline{\wedge} G) = \mathrm{Min}(I(F), I(G)))$.

Now the key observation is that under some conditions, we can define $\underline{\neg}^{\alpha}$ (from (35)) in terms of $\neg^{\alpha}$, $\underline{\wedge}$, and $\underline{\vee}$, hence also in terms of $\neg^{\alpha}$ and $\underline{\wedge}$ only.

LEMMA 4.3. *Suppose that for some formula* L^{α}, $I(L^{\alpha}) = \alpha > 1$. *Then for any formula* F, $I(\underline{\neg}^{\alpha}F) = I(\neg^{\alpha}F \underline{\vee} ((F \underline{\wedge} \neg^{\alpha}F) \underline{\wedge} L^{\alpha}))$.

Proof. We provide below a condensed truth table, with the minimal amount of information needed to prove the result. The value of a formula is represented under its main connective. We assume that $\beta \geq \alpha$ and that $1 < \gamma < \alpha$ (if such a γ exists), and we write the value of each sentence in bold.

$\underline{\neg}^{\alpha}F$		$(\neg^{\alpha}F$		$\underline{\vee}$	$((F \underline{\wedge} \neg^{\alpha} F)$			$\underline{\wedge} L^{\alpha}))$		
1	0	1	0	**1**						
0	1	0	1	**0**	0	0	1	0		
α	β	β	β	**α**	β	β	β	β	α	α
1	γ	1	γ	**1**						

It can then be shown that some ground interpretations have no fixed points.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Trivalization Lemma: Suppose that the logic is based on $\neg$, $\underline{\wedge}$, and $\underline{\vee}$ (one of the latter two connectives can be eliminated if the language has $\neg^2$, as was shown in Section 3.1.1) and that for some α, β with $1 < \alpha < \beta$, the language has negations of rank α and β. Then some ground interpretations have no fixed points which assign both the value α and the value β.

Proof. Suppose that for each pair of formulas L, L', the ground interpretation i allows for the definition of $< \nu_{L,L'}, \underline{\neg}^{\beta}{}_{L'}\underline{\neg}^{\alpha}{}_{L}\underline{\neg}^{\alpha}{}_{L}\mathrm{Tr}(\nu_L) >$, where

$\underline{\neg}^{\alpha}{}_{L}F$ abbreviates $\neg^{\alpha}F\underline{\vee}((F\,\underline{\wedge}\neg^{\alpha}F)\,\underline{\wedge}\,L)$, and

$\underline{\neg}^{\beta'}{}_{L}F$ abbreviates $\neg^{\beta}F\underline{\vee}((F\,\underline{\wedge}\neg^{\beta}F)\,\underline{\wedge}L')$.

(Note that a ground interpretation with the desired properties can be found if the language has at least as many names as there are formulas).

Now suppose that I is a fixed point that extends i and assigns the value α to some formula L^{α} and the value β to some formula L^{β} with $1 < \alpha < \beta$. Then for $L = L^{\alpha}$ and $L' = L^{\beta}$ the formula $\underline{\nu_{L,L'}}$ cannot be assigned any truth value, as was shown in Section 3.4.1.

§5. Existence of fixed points As was shown in Section 3.2.2, Monotonicity does not hold in the present system, at least on a standard ordering of the truth values. But a weakened form of the property ('Weak Monotonicity') turns out to be available, which suffices to prove the existence of a variety of fixed points.

5.1. Weak Monotonicity We start with two definitions.

DEFINITION 5.1. *Semantically closed classes*[25]
A class S of formulas is semantically closed in an interpretation i for L^{-} just in case

(i) S contains the subformulas of each of its members, and

(ii) for any interpretations I and I' which extend i to L, if I and I' agree on the classification of the members of S among the various extensions of Tr, then for each formula F in S, I(F) = I'(F).

(Note: For applications to the analysis of quantificational languages in Sections 5.2 and 5.3, subformulas *in (i) will be replaced with* constituents*, a notion which is defined in Section 5.2.2).*

DEFINITION 5.2. *Jump operator*
If I is an interpretation that extends i to L, f(I) is that interpretation obtained from i by setting, for each formula F: F belongs to $f(I)^{\alpha}(Tr)$ iff $I(F) = \alpha$.
With these definitions in mind, we can define a notion of Weak Monotonicity which holds in the present system.

Weak Monotonicity: Let S be a semantically closed set of formulas, and suppose that for each member F of S, I specifies that F belongs to an α-extension of *Tr* for $\alpha \leq \beta$. Define a partial order $\leq_S^{\beta}$ over interpretations of the language as follows:

$I \leq_S^{\beta} I'$ iff (i) all the sentences of S are in an α-extension of *Tr* for $\alpha \leq \beta$, and (ii) for every $\alpha < \beta$, $I^{\alpha}(\mathrm{Tr}) \subseteq I'^{\alpha}(\mathrm{Tr})$. Then: if $I \leq_S^{\beta} I'$, $f(I) \leq_S^{\beta} f(I')$.

We note that $I \leq_S^{\beta} I'$ just in case I and I' fully agree on the classification of members of S, except that some sentences classified by I in the β-extension of *Tr* are classified by I' in some 'lower' α-extensions.

[25] Related notions are defined in Leitgeb (2005) and Schlenker (2007).

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Proof. The proof is by induction on the construction of formulas. In each case we assume that $I \leq_S^\beta I'$, and we show that if $F \in S$ and $I(F) = \alpha < \beta$, then $I'(F) = \alpha$.

$F = p$ Since the value of p only depends on i (and not on the interpretation of *Tr*), it is immediate that $I(F) = I'(F) = i(F)$.

$F = Tr(s)$ Since S is semantically closed, $\underline{s} \in S$ (otherwise we could find an interpretation I* which agrees with I on the classification of the members of S but not of $\underline{s}$; but then we would have $I(F) \neq I^*(F)$, contrary to our hypothesis that $F \in S$ and that S is semantically closed). If $I(F) = \alpha < \beta$, $\underline{s} \in I^\alpha(Tr)$. But by assumption $I \leq_S^\beta I'$, hence $\underline{s} \in I'^\alpha(Tr)$. It follows that $I'(F) = \alpha$.

$F = (G \wedge H)$ Suppose that $I(F) = \alpha < \beta$. (i) If $\alpha = 1$, $I(G) = I(H) = 1$, and thus by the Induction Hypothesis $I'(G) = I'(H) = 1$, whence $I'(F) = 1$. (ii) If $\alpha = 0$, $I(G) = 0$ or $I(H) = 0$, and by the Induction Hypothesis $I'(G) = 0$ or $I'(H) = 0$, whence $I'(F) = 0$. (iii) If $1 < \alpha < \beta$, there are just two cases to consider. (a) If $I(G) = 1$ and $I(H) = \alpha$ (or the symmetric case obtained by permuting *G* and *H*), by the induction Hypothesis $I'(G) = 1$ and $I'(H) = \alpha$, whence $I'(F) = \alpha$. (b) Otherwise, it must be that $I(G), I(H) > 1$ and $Max(I(G), I(H)) = \alpha$. By the Induction Hypothesis, $I'(G) = I(G)$ and $I'(H) = I(H)$, whence $I'(F) = \alpha$.

$F = \neg^\alpha G$ Suppose that $I(F) = \alpha' < \beta$. Since by assumption $\alpha < \beta$, it could not be that $I(G) \geq \beta$ (because if so it would also be the case that $I(\neg^\alpha G) \geq \beta$). So $I(G) < \beta$, and by the Induction Hypothesis $I'(G) = I(G)$. Therefore $I'(F) = I(F)$.

5.2. The Extension Theorem We are now in a position to define the notion of a *local fixed point*, and to show—thanks to Weak Monotonicity—that under certain conditions a local fixed point over a set of sentences can be extended to a local fixed point over certain larger sets (the 'larger sets' in question may be the entire language if it forms a set).

DEFINITION 5.3. *Local fixed points*

An interpretation I is a local fixed point[26] *over a class S of formulas just in case for each member F of S and for each ordinal α, $I(F) = \alpha$ iff $F \in I^\alpha(Tr)$. (We sometimes say that a fixed point is 'global' if it is a local fixed point over the entire language).*

Extension Theorem: Suppose that the logic satisfies Weak Monotonicity. Let S and S' be two semantically closed sets of sentences, and suppose that I is a fixed point over S. Suppose further that (i) β is greater than all the values assigned by I to members of S, and (ii) that β is also greater than the rank of each member of S'. Then an interpretation I' can be found which (i) is a fixed point over S and over S', (ii) agrees with I over S, and (iii) assigns no value above β to any sentence of S∪S'.

Proof. We start by defining by transfinite induction an increasing series of interpretations. This series must have a fixed point, which is the desired interpretation. We note for future reference that only Weak Monotonicity is needed to obtain the result.

Construction

(i) I_0 is determined by: (i) classifying all members of S as I requires, and (ii) classifying all other formulas in the β-extension of *Tr*, where β is greater than the rank of each negation in S', and also greater than all the values assigned by I to the members of S.

(ii) $I_{k+1} = f(I_k)$.

[26] This notion is also defined in Schlenker (2007).

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(iii) For limit ordinal k, I_k is determined by:

(a) for each ordinal $\alpha < \beta$, $I_k^\alpha(\mathrm{Tr}) := \cup_{k'<k} I_{k'}^\alpha(\mathrm{Tr})$.

(b) $I_k^\beta(\mathrm{Tr}) := L - \cup_{\alpha<\beta} I_k^\alpha(\mathrm{Tr})$.

(c) for all other ordinals γ, $I_k^\gamma(\mathrm{Tr}) = \emptyset$.

REMARK 5.4.

(i) It is immediate (by 'Semantic Restraint') that for any ordinal k, for every formula F, $I_k(F) \leq \beta$.

(ii) A trivial induction shows that for every ordinal k, I_k agrees with I over S (because I is a local fixed point over S, S is semantically closed, and by construction I_0 agrees with I over S).

LEMMA 5.5. $I_0 \leq^\beta_{S\cup S'} f(I_0)$ $(= I_1)$

By construction, the members of (S'-S) start out in the β-extension of Tr, and thus they automatically satisfy the claim. As for the members of S, by Remark (ii) above their semantic status remains constant throughout the series of interpretations.

LEMMA 5.6. *For all ordinals a, b, if $a < b$, then $I_a \leq^\beta_{S\cup S'} I_b$.*

The proof is by transfinite induction on b.

(i) If $b = 0$, the claim is vacuously true.

(ii) If $b = b'+1$, we only have to show that $I_{b'} \leq^\beta_{S\cup S'} I_b$.

(a) If $b' = 0$, the claim has already been proven in Lemma 1.

(b) If $b' = b"+1$, we have $I_{b"} \leq I_{b'}$. By Weak Monotonicity, $f(I_{b"}) \leq f(I_{b'})$, that is, $I_{b'} \leq I_b$.

(c) If b' is a limit ordinal, for every ordinal $\alpha < \beta$: $I_{b'}^\alpha(Tr) = \cup_{b"<b'} I_{b"}^\alpha(Tr) \subseteq \cup_{b"<b'} I_{b"+1}^\alpha(Tr)$. But for every $b" < b'$, $I_{b"} \leq^\beta_{S\cup S'} I_{b'}$, hence (by Weak Monotonicity) $I_{b"+1} \leq^\beta_{S\cup S'} I_{b'+1}$. Thus for every $\alpha < \beta$, $I_{b"+1}^\alpha(Tr) \subseteq I_{b'+1}^\alpha(Tr)$, and in the end $I_{b'}^\alpha(Tr) \subseteq I_{b'+1}^\alpha(Tr) = I_b^\alpha(Tr)$.

(iii) If b is a limit ordinal, for every $\alpha < \beta$ $I_b^\alpha(Tr) = \cup_{b"<b} I_{b"}^\alpha(Tr)$ and the result follows immediately.

Proof of the Extension Theorem. In short, since $(I_k)_{k\in Ord}$ is increasing, it has a fixed point, which we call I*.[27]

5.3. Some fixed points The Extension Theorem is powerful enough to allow for the construction of a variety of fixed points. In this section we restrict attention to the case in which the language forms a set (in Section 5.1 we show that fixed points can be constructed for any propositional language, even those that do not form a set).

[27] To be more specific, suppose that $(I_k)_{k\in Ord}$ has no fixed point. Then for any ordinal k there is a sentence of S∪S' (in fact, of S') which first 'leaves' the β-extension of *Tr* at level k+1. We can thus define a function g which for each formula F of S∪S' returns k = g(F) if F first leaves the the β-extension of *Tr* at level k+1; and which returns 0 if F never leaves the β-extension of *Tr*. But this is impossible: g would be a function whose domain is a set and whose range is the class of all ordinals—which contradicts the axiom of replacement. (It can be noted that the argument crucially relies on the fact that S∪S' forms a set.)

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

5.3.1. The boring fixed point When we take S to be empty and S' to be the entire language, we obtain a fixed point which is not particularly interesting, the 'boring fixed point'. All sentences start out in the β-extension of *Tr*. Thus (by Semantic Restraint) {0, 1, β} are the *only* values that are ever assigned in the construction of the fixed point. In fact, we construct in this way a notational variant of the 'least fixed point' obtained by Kripke with the Strong Kleene logic, as shown by the following lemma.

Emulation Lemma: Let S be empty in the construction of the fixed point I*, and let S' be the entire language. For any formula F of the language, let F^{-} be the result of 'stripping' all negations from their superscripts, and let I** be the least fixed point obtained for the resulting language in the Strong Kleene logic. Then for any formula *F* of the language, if we write the third truth value of the Strong Kleene logic as β,

$I^*(F) = I^{**}(F^{-})$.

Proof. The result follows from two observations.

(i) First, when the only possible values of our logic are {0, 1, β}, the semantic rules we have are notational variants of those of the Strong Kleene logic. This follows because (a) β is the only indeterminate value that is ever assigned by the interpretations $(I_k)_{k \in Ord}$, and (b) β is greater than the rank of any formula of the language, with the effect that $I(\neg^{\alpha}F) = \beta$ whenever $I(F) = \beta$. It is routine to check that the rules for *Tr* and $\wedge$ are notational variants of the Strong Kleene rules in the present situation (this also holds of $\vee$, though as discussed earlier this connective is redundant if we have $\wedge$ and $\neg^2$).

(ii) Second, the construction of the boring fixed point mirrors that of Kripke's least fixed point. We start by putting all the sentences of the language in the β-extension of *Tr*, and apply a rule of revision which is a notational variant of Kripke's (by (i)). Thus each step in the construction of the boring fixed point mirrors the corresponding step in the construction of Kripke's least fixed point. The end result should thus be the same.

5.3.2. Interesting fixed points For any ordinal α, we can in principle define an α-Liar as: $< \lambda_\alpha, \neg^{\alpha}\mathrm{Tr}(\lambda_\alpha) >$ (I write 'in principle' because we have assumed that the language is a set, and thus that it does not contain all possible negations, since there are ordinal-many of those). It is immediate that if $\underline{\lambda_\alpha}$ can be expressed:

(i) for each $\alpha' \geq \alpha$, a local fixed point over $\{\underline{\lambda_\alpha}\}$ can be found in which $\underline{\lambda_\alpha}$ has value α', and

(ii) for no $\alpha' < \alpha$ can a local fixed point over $\{\underline{\lambda_\alpha}\}$ be found in which $\underline{\lambda_\alpha}$ has value α'.

We can thus construct a local fixed point by classifying each $\underline{\lambda_\alpha}$ in the α-extension of *Tr*. By the Extension Theorem, this local fixed point can be extended to a global fixed point for the entire language. Unlike the boring fixed point, the interpretation we obtain does not mirror any of Kripke's fixed points, since it makes crucial use of different indeterminate truth values.

§6. Extensions. In this section we consider some refinements of our main results. First, we discuss the case in which the language does not form a set, and show that under certain conditions it is still possible to extend any local fixed point to a global one. Second, we examine how our results can be generalized to quantificational languages (the analysis

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

remains roughly the same when the language forms a set, but significant changes are required when this is not so).

6.1. Languages that do not form a set. Let us consider a language that makes full use of all the negations afforded by the logic. Since there are ordinal-many such negations, the language cannot form a set. For this reason, the proof that some fixed points exist must be refined.

Under certain conditions, we can construct a fixed point 'by stages'.[28] This will be the case if for each sentence s there exists a semantically closed class C(s) which contains s and forms a set. We give ourselves an ordinal enumeration $(s_i)_{i\in Ord}$ of all the sentences of the language, and we define a local fixed point for $C(s_0)$, which we extend to $C(s_1)$; we extend the result to $C(s_2)$,..., and keep going in this way until the entire language has been 'covered' (this requires a lot of patience, since ordinal-many stages are needed in the construction). This procedure will be shown to be applicable to propositional languages and to certain quantificational languages.

Before we provide the construction, we introduce a piece of terminology:

DEFINITION 6.1. *Order of a set of sentences*
If S is a set of sentences, we define Ord(S) (the order of S) as:
Ord(S) = 1 if S contains no negation, and
Ord(S) = α if α is the least upper bound of $\{rk(s): s \in S\}$.

THEOREM 6.2. *Let L be a language whose sentences can be enumerated by the ordinals, and let i be a ground interpretation in which, for each sentence s of L, there is a semantically closed class C(s) which contains s and forms a set (L itself need not form a set). Then any local fixed point I over a semantically closed set* S_0 *can be extended 'by stages' to a global fixed point for the entire language.*

Proof. Let $(F_i)_{i>0(i\in Ord)}$ be an ordinal enumeration of the language (for notational reasons the series starts with F_1 rather than F_0). We simultaneously define the construction and explain why it yields local fixed points for increasingly larger sets of sentences.

Stage 0: Let β be the least upper bound of the values assigned by I to members of S. We set $\beta_0 := Max\{Ord(S_0)+1, \beta+1\}$. We classify the members of S_0 in the extensions of *Tr* specified by I, and we put all remaining sentences in the β_0-extension of *Tr*. We obtain in this way an interpretation I_0 which only assigns values as great as β_0.

Stage k+1: —If $F_{k+1} \in S_i$, we set $I_{k+1} := I_k$, $\beta_{k+1} := \beta_k$, and $S_{k+1} := S_k$.
—Otherwise, we set $\beta_{k+1} := Max\{Ord(C(F_{k+1})+1, \beta_k+1\}$ and $S_{k+1} := S_k \cup C(F_{k+1})$. For each formula F: (i) if $F \in S_k$, we leave F in the extension of *Tr* which is specified by I_k; (ii) otherwise, we initially put F in the β_{k+1}-extension of *Tr*. By the Extension Theorem, we obtain an interpretation I_{k+1} which is a local fixed point for S_{k+1} and agrees with I_k on S_k.

Stage k for limit ordinal k:
We call β'_k the least upper bound of $\{\beta_{k'}: k' < k\}$.
Case 1. If $F_k \in \cup_{k'<k} S_{k'}$, we set $\beta_k := \beta'_k+1$, $S_k := \cup_{k'<k} S_{k'}$, and we define I_k by:

a. if $F \notin S_k$, we put F in the β_k-extension of *Tr*.

[28] A construction by stages is also used in Cook (2008a). See Section 6.4 for a discussion of Cook's system.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

b. if $F \in S_k$, we define k* to be the least stage k' for which F belongs to $S_{k'}$, and we put F in the extension of *Tr* which is specified by I_{k*}. I_{k*} is a local fixed point over S_{k*}, and all 'later' interpretations $I_{k'}$ (for $k^* < k' < k$) agree with I_{k*} on the classification of the members of S_{k*}, so this is also the case of I_k. Thus I_k is a local fixed point over S_{k*}, and for any ordinal α, $I_k(F) = \alpha$ iff $F \in I^{\alpha}(\mathrm{Tr})$.

Case 2. If $F_k \notin \cup_{k'<k} S_{k'}$, we define $\beta_k := \mathrm{Max}\{\beta'_k+1, \mathrm{Ord}(C(F_k))+1\}$ and $S_k := C(F_k) \cup (\cup_{k'<k} S_{k'})$ and we define our initial interpretation $I_{k,0}$ by:

a. if $F \notin \cup_{k'<k} S_{k'}$, we put F in the β_k-extension of *Tr*.
b. if $F \in \cup_{k'<k} S_{k'}$, we define k* to be the least stage k' for which F belongs to $S_{k'}$, and we put F in the extension of *Tr* specified by I_{k*}.

By the same reasoning as in Case 1, it can be shown that $I_{k,0}$ is a local fixed point over $\cup_{k'<k} S_{k'}$. By the Extension Theorem, $I_{k,0}$ can be extended to an interpretation I_k which is a local fixed point over S_k and which agrees with $S_{k,0}$ over $\cup_{k'<k} S_{k'}$.

It is immediate that for any formula F, we can find a least ordinal k for which $F \in S_k$. And since I_k is a local fixed point over S_k, it is clear that I_k and all the interpretations of later stages assign the value α to F if and only if they classify F in the α-extension of *Tr*. In this sense, we have extended the local fixed point I to a global fixed point for the entire language.

COROLLARY 6.3. *If L is a propositional language, any local fixed point I over a semantically closed set S_0 can be extended 'by stages' to a global fixed point for the entire language.*

Proof. It suffices to observe that for each sentence F of a propositional language, we can define C(F) as follows:

$C_0(F) = \{F\}$
$C_{i+1}(F) = C_i(F) \cup \{\underline{s}: \mathit{Tr}(s) \in C_i(F)\} \cup \{F'': \exists F' \ (F' \in C_i(F) \text{ and } F'' \text{ is a subformula of } F'\}$
$C(F) = \cup_{i \in |N} C_i(F)$
It is immediate that C(F) is semantically closed.

6.2. A quantificational language that forms a set. Up to this point we have restricted attention to propositional languages. Adding quantifiers over sentences leads to inessential changes when the language forms a set, as is shown in this section. When the language does not form a set, quantifiers must be equipped with set restrictions if the language is to have fixed points, as is shown in Section 5.3. (When the language forms a set, it cannot contain all of the ordinal-many negations we have defined; so the case in which the language does not form a set is quite important.)

6.2.1. Enriched syntax and semantics for L. Given our treatment of disjunction and conjunction, it is natural to enrich our syntax and semantics with the following rules, which are defined on the assumption that the language forms a set (quantification must be given a different treatment when this is not the case).

The propositional letters, relational symbols, sentence names, and variables are all indexed by a set J, and the negations are indexed by a set K (which of course cannot contain all ordinals).

(38) Quantificational Syntax for L
The enriched language L^- is now formed by rules a, b, c, e; the enriched language L is formed by rules a-e.
a. Propositional constants: $p := p_i$ (for $i \in J$)

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

b. Sentence-denoting terms: $s := s_i, x_i$ (for $i \in J$)
(the s_i's are constants; the x_i's are variables).
c. Sentence-denoting relations: $R^n := R^n_i$ (for n, $i \in J$)
d. Truth predicate: Tr
e. Formulas: $F := p \mid Tr(s) \mid \neg^{\alpha} F \mid (F \wedge F) \mid (F \vee F) \mid$
$R^n(x,\ldots, x)$ (n times) $\mid \exists x\, F \mid \forall x\, F$
(for α any ordinal such that $\alpha \in K$ and $\alpha \geq 2$).

(39) Quantificational semantics for L
Let i assign a classical value (0 or 1) to every propositional letter, and for every ordinal α, let $I^{\alpha}(Tr)$ be a class of sentences of L, with the condition that for every sentence S of L, there is exactly one ordinal α such that S belongs to $I^{\alpha}(Tr)$. We assume that i guarantees that each sentence $\underline{s}$ has a canonical name s, which allows us to give a substitutional semantics to quantifiers. We use $\wedge$ in the meta-language to refer to the least upper bound of a set, and we write $G[s/x_i]$ for the formula obtained from G by replacing each free occurrence of x_i with s.
$I(p) = i(p)$
$I(R^n(s_{i_1},\ldots, s_{i_n})) = 1$ iff $<i(s_{i_1}),\ldots, i(s_{i_n})> \in i(R^n)$; otherwise, $I(R^n(s_{i_1},\ldots, s_{i_n})) = 0$.
$I(Tr(s)) = \alpha$, where α is the (unique) ordinal for which I(s) belongs to $I^{\alpha}(Tr)$.
$I(F \wedge G) = 1$ iff $I(F) = I(G) = 1$; $I(F \wedge G) = 0$ iff $I(F) = 0$ or $I(G) = 0$; otherwise, $I(F \wedge G) = Max(I(F), I(G))$.
$I(F \vee G) = 1$ iff $I(F) = 1$ or $I(G) = 1$; $I(F \vee G) = 0$ iff $I(F) = I(F) = 0$; otherwise, $I(F \vee G) = Max(I(F), I(G))$.
$I(\forall x_i\, F) = 1$ iff for each sentence $\underline{s}$, $I(F[s/x_i]) = 1$; $I(\forall x_i\, F) = 0$ iff for some sentence $\underline{s}$, $I(F[s/x_i]) = 0$; otherwise, $= \wedge\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$.
$I(\exists x_i\, F) = 1$ iff for some sentence $\underline{s}$, $I(F[s/x_i]) = 1$; $I(\exists x_i\, F) = 0$ iff for each sentence $\underline{s}$, $I(F[s/x_i]) = 0$; otherwise, $I(\exists x_i\, F) = \wedge\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$.
$I(\neg^{\alpha} F) = 1$ iff for some ordinal $\beta < \alpha$ with $\beta \neq 1$, $I(F) = \beta$; $I(\neg^{\alpha} F) = 0$ iff $I(F) = 1$; otherwise, $I(\neg^{\alpha} F) = I(F)$.

We note immediately that when the language does not form a set, $\wedge\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$ may not be defined. In fact, if the language and the ground interpretation i allow for the definition of an α-Liar for each ordinal $\alpha \geq 2$, we can prove that $\wedge\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$ is defined in no fixed point whatsoever for $F = (Tr(x_i) \vee \neg^2 Tr(x_i))$. This is because in this case there is no upper bound on $I(F[s/x_i])$, since for each ordinal $\alpha \geq 2$ some sentence, the α-Liar, has a value $\geq \alpha$.

6.2.2. Main properties.

□ *Interdefinability*

As was the case with conjunction and disjunction, existential and universal quantification are interdefinable given $\neg^2$:

(40) In any interpretation I,
a. $I(\exists x_i\, F) = I(\neg^2 \forall x_i \neg^2 F)$.
b. $I(\forall x_i\, F) = I(\neg^2 \exists x_i \neg^2 F)$.

Proof. Consider (40)a. The identity is trivial if $I(\exists x_i\, F) \in \{0, 1\}$. If not, $\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$ either (i) includes 0 and some indeterminate values and nothing else, or (ii) only includes indeterminate values. In case (i), $\{I(\neg^2 F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$ includes 1 and some indeterminate values and nothing else, and thus $I(\forall x_i\, \neg^2 F) > 1$ and

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

$I(F) = I(\forall x_i \neg^2 F) = \wedge\{I(F[s/x_i])$: s is a sentence of L$\} = I(\exists x_i\ F)$. The argument is similar in case (ii). The proof of (40)b follows from (40)a and the observation that for any formula F, $I(\neg^2\neg^2 F) = I(F)$ (see (31)).

As was the case for connectives, the interdefinability of quantifiers makes it possible to take one of them (say, the universal quantifier) as primitive and the other one as defined (this is possible as long as the language has the negation $\neg^2$).

□ *Weakened Semantic Restraint*

We saw earlier that the propositional language exercised Semantic Restraint: if all the atomic formulas of *F* have values in V, the value of F is in $V \cup \{0, 1\}$. However Semantic Restraint as we have defined it does not make sense in the quantificational case; this is because subformulas with free variables are not assigned a value by our interpretation procedure. We will thus define a generalization of the notion of subformula, which we call 'constituent'. This will allow us to define a weakened form of Semantic Restraint, which will hold in the quantificational case.

Let us say that a formula G is an *immediate constituent* of a formula F just in case:

(a) G is a conjunct or a disjunct of F, or
(b) $F = \neg^\alpha G$ for some ordinal α, or
(c) $F = \forall x_i$ G' or $F = \exists x_i$ G', and for some sentence s, $G = G'[s/x_i]$.

The relation *x is a constituent of y* is obtained by taking the transitive closure of the relation *x is an immediate constituent of y*. An atomic constituent is simply a constituent which is an atomic formula.

Semantic Restraint, properly generalized by replacing 'subformula' with 'constitutent', does *not* generally hold for a quantificational language. This is because $I(\forall x_i\ F) = \wedge\{I(F[s/x_i])$: s is a sentence of L$\}$, and there is no guarantee that the least upper bound of $\{I(F[s/x_i])$: s a sentence of L$\}$ is itself in that set. But it is immediately apparent that a weakened form of Semantic Restraint still holds:

Weakened Semantic Restraint: If all the atomic constituents of F have values in V, *and if the least upper bound of each series of values in V is itself in V,* then the value of F is in $V \cup \{0, 1\}$.

□ *Weak Monotonicity*

Weak Monotonicity also holds for a quantificational language. However we must modify the definition of a 'semantically closed class'; where we mentioned the 'subformulas' of a sentence, we will now talk about its 'constituents'. The difference only matters in the case of quantificational formulas: with the modified definition, we ensure that if a semantically closed set S contains a formula $\forall x_i F$, then it also contains each occurrence of $F[s/x_i]$ for each sentence s of the language.

A modified proof of Weak Monotonicity can be given by induction on the complexity c of formulas, where (a) $c(F) = 1$ if F is atomic, (b) $c(F \wedge G) = c(F) + c(G)$, and (c) $c(\neg^\alpha F) = c(\forall x_i\ F) = c(F) + 1$. The proof of Section 4.1 can be reproduced: we assume that $I \leq_S^\beta I'$, and we show that if $I(F) = \alpha < \beta$, then $I'(F) = \alpha$. For complex *F*, we assume that the result holds for formulas of lower complexity, and hence in particular for the immediate constituents of F (which by the modified definition of a semantically closed set must all belong to S). Since in the propositional case the immediate constituents of F are just its immediate subformulas, there is no need to revise the beginning of the proof, and we only have to consider the quantificational case:

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

$F = \forall x_i\ G$ Suppose that $I(F) = \alpha < \beta$. (i) If $\alpha = 1$, for each sentence $\underline{s}$, $I(G[s/x_i]) = 1$, and thus by the Induction Hypothesis $I'(G[s/x_i]) = 1$, whence $I'(F) = 1$. (ii) If $\alpha = 0$, for some sentence $\underline{s}$, $I(G[s/x_i]) = 0$, and thus by the Induction Hypothesis $I'(G[s/x_i]) = 0$, whence $I'(F) = 0$. (iii) If $1 < \alpha < \beta$, for each sentence $\underline{s}$, $1 \leq I(G[s/x_i]) < \beta$ and $\alpha = \wedge\{I(G[s/x_i])$: $\underline{s}$ is a sentence of L$\}$. By the Induction Hypothesis, for each sentence $\underline{s}$, $I'(G[s/x_i]) = I(G[s/x_i])$ and thus $I'(F) = I(F)$.

□ *Existence of fixed points*

Our results on the existence of various fixed points only depended on Weak Monotonicity. Since the latter also holds for a quantificational language, the various results carry over as well.

6.3. A quantificational language that does not form a set. As was already pointed out, if the quantifiers range over all sentences, there is in general no fixed point for a quantificational language that does not form a set. Specifically, if the language and the ground interpretation i allow for the definition of an α-Liar for each ordinal $\alpha \geq 2$, we can prove that for $F = (Tr(x_i) \vee \neg^2 Tr(x_i))$, $\wedge\{I(F[s/x_i])$: $\underline{s}$ is a sentence of L$\}$ is defined in no fixed point whatsoever.

The natural solution is to require that every quantifier carry a restrictor that denotes a set. This condition can be implemented by replacing the unrestricted quantifier $\forall x_i$ with a restricted quantifier $[\forall x_i : F]$, and by requiring that (i) F be a classical formula with one free variable, x_i,and that (ii) the class $\{\underline{s}$: $\underline{s}$ is a sentence and $i(F[s/x_i]) = 1\}$ form a set (we could try to liberalize the constraint a bit by allowing F to contain additional free variables, though this would require a more complex definition of the constraint that F denotes a set).

With this modification, we can preserve our results about quantificational languages, replacing the notion of 'constituent' with 'constituent*', where a constituent* of F is the same thing as a constituent of F, except if $F = [\forall x_i$: G]H. In that case, the constituents* of F are the members of $\{\underline{s}$: $\underline{s}$ is a sentence and $i(G[s/x_i]) = 1\}$ (and by our definition of restricted quantification, we are guaranteed that this class is indeed a set). This modification matters for the definition of a 'semantically closed class' of sentences, which must now include the constituents* of each of its members.

With these modifications, we can extend a local fixed point to a global fixed point by proceeding 'by stages', as we did in Section 5.1. All we need to do is to provide a definition of C(F). We retain the definition of the Corollary 5.3 of Section 5.1, except that we replace 'subformula' with 'constituent*':

$C_0(F) = \{F\}$

$C_{i+1}(F) = C_i(F) \cup \{\underline{s}$: *Tr(s)* $\in C_i(F)\} \cup \{F''$: $\exists F'$ ($F' \in C_i(F)$ and F" is a constituent* of F'$\}$

$C(F) = \cup_{i \in |N} C_i(F)$

Let us check that C(F) is semantically closed (in the new sense). Clearly, if $F' \in C(F)$, then for some $i \in |N$, $F' \in C_i(F)$. Thus:

(i) the constituents* of F' belong to $C_{i+1}(F)$, and hence also to C(F).

(ii) if F' contains any occurrence of *Tr,* there are just two possibilities:

— the argument of *Tr* is a constant t, and $\underline{t}$ belongs to $C_{i+1}(F)$.

— the argument of *Tr* is a variable x_i. If so, *Tr* must be part of H in a subformula $[\forall x_i : G]H$, and for each sentence $\underline{t}$ for which $i(G[t/x_i]) = 1$, $\underline{t}$ belongs to $C_{i+1}(F)$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

As a result, if I and I' are two extensions of i that agree on the classification of all the members of C(F) among the various extensions of *Tr*, they agree in particular on the value of all atomic formulas of F', and thus I(F') = I'(F'), as was desired.

§7. Comparisons. While the present approach should be systematically compared to other accounts of the Liar, we restrict attention to three recent philosophical theories and one older technical result. On a methodological level, our account shares with Burge (1979) a desire to take seriously the *linguistic underpinnings* of the Liar.[29] On a technical level, it shares with Gupta and Martin (1984) and Field (2008) the goal to construct fixed points for logics that do not satisfy standard properties of monotonicity. On a substantive level, our endeavor is fully consonant with analysis which was developed independently by Cook (2008a, 2008b), who claims that one should *embrace the Liar* and do so within a logic that countenances ordinal-many truth values.

7.1. Burge (1979). Starting from the puzzle of the Strengthened Liar as it arises in natural language, Burge (1979) offers a hierarchical solution to the paradoxes, based on an infinite series of truth predicates. He takes the truth predicate to be an indexical expression, whose extension varies with the context of use; which truth predicate is used it taken to depend on pragmatic considerations. This general direction is close to the spirit of second fragment we explored (L*). There the truth predicate is taken to have *selectional restrictions*, which are revised as one learns that there are more possible truth values than one initially thought. The technical results of the two theories are rather different, however;[30] we leave a systematic comparison for future research.

7.2. Gupta and Martin (1984). Gupta and Martin (1984) show that a fixed point can be found for a trivalent logic which is not monotonic. Specifically, they prove that the Weak Kleene Logic can contain both its own truth predicate *Tr* and an 'undefinedness' predicate N, which is true of a sentence just in case it is neither true nor false. It is immediate that this predicate is nonmonotonic: $N(s)$ is true when $\underline{s}$ is indeterminate, but it becomes false when $\underline{s}$ becomes classical.

The basis of Gupta and Martin's construction is the observation that two Weak Kleene interpretations that agree on the union of the extension and antiextension of any predicate are also bound to make the same sentences indeterminate. To put it differently, in order to determine whether a Weak Kleene sentence is indeterminate, one does not need to know how the 'pie' is divided between the extension and antiextension of the various predicates of the language; all that matters is the union of the extension and antiextension.

---

[29] For lack of space, we leave a comparison with Glanzberg (2001, 2004) for future research.

[30] Let us consider a fixed point in which the sentence $\neg Tr_2(\lambda_2)$, denoted by λ_2 itself, has the value 2. If we assess $\underline{\lambda_2}$ with respect to this same truth predicate Tr_2, as Burge initially does, we also obtain the value 2 (since $Tr_2(\lambda_2)$ also has the value 2). On the other hand, if we evaluate $\underline{\lambda_2}$ with respect to Tr_3, we conclude that $\underline{\lambda_2}$ is untrue (since $Tr_3(\lambda_2)$ has the value 0). When indeterminate truth values are collapsed, we reach the conclusion that $\underline{\lambda_2}$ evaluated with respect to Tr_2 has the value #, while it has the value 0 when evaluated with respect to Tr_3.

Due to a twist of Burge's analysis, his technical results are different. For him, a truth_i schema is only applied to sentences that are not pathological_i. But he has an axiom that guarantees that a sentence that is pathological_i is untrue_i (axiom (6) p. 188). Thus $\underline{\lambda_2}$ is pathological_2, and for this reason is it also untrue_2. If we evaluate $\underline{\lambda_2}$ with respect to Tr_3, the result is not pathological_3 (because Tr_3 is higher in the hierarchy of truth predicates than Tr_2), and the truth_3 schema applies. Because $\underline{\lambda_2}$ is untrue_2, the content of $\underline{\lambda_2}$ (namely $Tr_2(\lambda_2)$) is true, and thus $\underline{\lambda_2}$ is declared true by Tr_3—whereas in our analysis it is declared untrue.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Gupta and Martin's construction is in two stages. We represent as $<<Tr^+, Tr^- >, <N^+, N^- >>$ the extension and antiextension of the predicates *Tr* and *N* respectively. If we identify an interpretation I to this pair, we consider a monotonic operation defined as follows:

μ(I) = <<true sentences in I, false sentences in I>, <indeterminate sentences in I, determinate sentences in I>>.

(i) First, a fixed point for *Tr* is constructed by stages using Kripke's procedure, starting from $<<\emptyset, \emptyset>, <D, \emptyset>>$. Let us call this fixed point $<<T^+, T^->, <D, \emptyset>>$.

(ii) Second, the same construction is performed again, but starting this time from $<<\emptyset, \emptyset>, <D\text{-}(T^+ \cup T^-), (T^+ \cup T^-) >>$.

Comparing the procedures in (i) and (ii), we see that the starting points are identical except for the reassignment of some sentences from the extension to the antiextension of *N*. Thanks to the basic observation mentioned above, this guarantees that in any given stage of both constructions, exactly the same sentences are determinate/indeterminate. It follows that in the fixed point obtained in (ii), the same sentences are determinate/indeterminate as in the fixed point obtained in (i)—which means that the interpretation of *N* is indeed adequate.

From the predicate *N* it is straightforward to define an operator *n*, where *nF* means that *F* is neither true nor false. Thus by definition:

nF is true iff *F* is in the extension of *N*

nF is false iff *F* is in the antiextension of *N*

It is immediate that this operator is nonmonotonic: if an atomic sentence *P* is indeterminate according to I, and if I' is identical to I except that it makes *P* true, *nP* is true according to I but false according to I'.

There are two major differences between Gupta and Martin's results and ours.

1. First, their analysis is based on Weak Kleene, whereas we start from Strong Kleene. Unlike the latter, the former is exceedingly cautious in its attribution of classical truth values, since it makes a formula indeterminate as soon as any of its atomic components is. This, in turn, is essential to enforce the 'basic property' on which the construction is based.
2. Second, Gupta and Martin's result does not produce even a limited form of expressive completeness. True, a predicate can be defined which adequately captures the notion of 'indeterminate sentence'. However a weak negation cannot be added to the language—in other words, no operator *neg* can be defined which guarantees that *neg F* is true just in case *F* is false or indeterminate, and false otherwise. For if such an operator existed, one could define a strengthened liar $\underline{\lambda^*}$, which says: *neg Tr*(λ^*). It is immediate that $\underline{\lambda^*}$ cannot be assigned any coherent truth value.[31]

7.3. Field (2008). In a series of recent works, synthesized in Field (2008), Field has offered an ambitious analysis of the Liar that has a number of appealing properties (insightfully discussed in Yablo, 2003):

Property 1. It validates full substitution between *Tr(c)* and $\underline{c}$.

[31] It may seem puzzling that such an operator cannot be defined if we have *Tr* and *N* at our disposal. But the obvious definition fails due to the very weakness of the evaluation scheme: let *neg F* abbreviate $N(c_F) \vee \neg Tr(c_F)$, where c_F denotes *F*; by the semantics of the Weak Kleene disjunction, this formula is *indeterminate* as soon as $\neg Tr(c_F)$ is—which means that it does not deliver the intended semantics for the operator *o*.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Property 2. The theory has a 'strong' conditional $\rightarrow$, which among others validates $Tr(c) \leftrightarrow \underline{c}$ (i.e., $(\mathrm{Tr(c)} \rightarrow \underline{c}) \wedge (\underline{c} \rightarrow \mathrm{Tr(c)})$).

Property 3. The theory is trivalent, but not compositional (a compositional variant can be devised, in which the values are considerably more fine grained—and more numerous).

Property 4. One can define a hierarchy of Liars of increasing strength. Specifically, thanks to $\rightarrow$, one can define a 'determinacy' operator D which gives rise to the Liars $<Q_0, \neg \mathrm{Tr}(Q_0)>$, $<Q_1, \neg \mathrm{DTr}(Q_1)>$, $<Q_2, \neg \mathrm{DDTr}(Q_2)>$, ... , $<Q_n, \neg \mathrm{D^nTr}(Q_n)>$, and so on.[32]

Property 5. The language has the resources to characterize the defective status of each of these Liars. Specifically, for each n, $\underline{Q}_n$ has a nonclassical value, but the sentence $D^{n+1}\underline{Q}_n$ has value 0.

Field's theory is explicitly designed to make 'sustained reasoning' possible—hence a sophisticated semantics for $\rightarrow$, intended to make it as 'strong' as possible without yielding inconsistency. This is done by way of a transfinite series of Kripkean fixed points, each of which is built on top of a 'seed' interpretation that treats $\rightarrow$ as a nonlogical connective. At each stage, the semantics of $\rightarrow$ is given by a revision-theoretic semantics in the spirit of Herzberger and Gupta's systems. Field then shows (in his 'fundamental theorem', e.g., Field, 2008, p. 257) that there are stages at which every sentence has its 'eventual' value.

Property 1 is shared by our main analysis (the fragment L). On the other hand, no version of Property 2 can be obtained in our theory, for as was already mentioned in Section 3.2.3, if $\beta > \alpha$, any formula of rank α is bound to have value β whenever each of its atomic formulas has value β. In particular, for $\mathrm{S} =: \mathrm{A} \rightarrow \mathrm{A}$ we *fail* to get that S is true when A has value β. Field solves this problem, but the price is that his semantics is not compositional (Property 3).[33] This is a radical departure from assumptions that are usually made about

---

[32] Field's determinacy operator is defined as $\mathrm{DA} := \mathrm{A} \wedge \neg(\mathrm{A} \rightarrow \neg \mathrm{A})$. When n is infinite, the definition of $D^{\alpha} A$ must be indirect (in brief, one can use definitions such as $(\forall \alpha < n)\ Tr(< D^{\alpha} A >)$, where $Tr(< D^{\alpha} A >)$ is a description within the language of the appropriate formula; see Field (2008, chap. 22).

[33] To see an example, consider Field's treatment of Curry's paradox, defined as $<\mathrm{K}, \mathrm{Tr(K)} \rightarrow \mathrm{F}>$, where F is a contradiction. In Field's construction, if Kripkean fixed points have been constructed for stages 0, 1, ..., the 'seed' valuation for stage α is given by (i), with the ordering $0 < \# < 1$ (Field 2008 p. 250; Field uses $1/2$ instead of #):

(i) $|\mathrm{A} \rightarrow \mathrm{B}|_{\alpha} = 1$ iff $(\exists \beta < \alpha)\ (\forall \gamma)\ [\beta \leq \gamma < \alpha \Rightarrow |\mathrm{A}|_{\gamma} \leq |\mathrm{B}|_{\gamma}]$
$= 0$ iff $(\exists \beta < \alpha)\ (\forall \gamma)\ [\beta \leq \gamma < \alpha \Rightarrow |\mathrm{A}|_{\gamma} > |\mathrm{B}|_{\gamma}]$
$= \#$ otherwise.

Clearly, $\underline{K}$, $Tr(K)$ and $\neg \underline{K}$ must all have the indeterminate value. By Property 2, the sentence $(\underline{K} \rightarrow Tr(K))$ is true (and by Property 1, so is $(\underline{K} \rightarrow \underline{K})$); in other words, $\rightarrow$ can yield truth when its antecedent and consequent are both indeterminate. On the other hand, it turns out that $(\underline{K} \rightarrow \neg \underline{K})$ has the indeterminate value as well, despite the fact that its antecedent and consequent are also indeterminate.

The latter observation follows because $\underline{\mathrm{K}}$'s value oscillates without end in the construction (see Yablo 2003 p. 317). For instance, if the first 'seed' valuation S_0 gives $\underline{\mathrm{K}}$ the value #, so will the first fixed point obtained from S_0. By Field's revision rule, the second 'seed' valuation S_1 must give $\underline{\mathrm{K}}$ the value 0 because its antecedent $Tr(K)$ has value # while its consequent F has value 0, and $\# > 0$. The next seed valuation S_2 - and associated fixed point - will give $\underline{\mathrm{K}}$ the value 1 (because in the previous stage $Tr(K)$ got value 0 and F got value 0, and $0 = 0$). And S_3 will give $\underline{\mathrm{K}}$ the value 0 (because in the previous stage $Tr(K)$ got value 1, and F got value 0). So the value of $\underline{\mathrm{K}}$ oscillates between 0 and 1 across the positive integers, and hence gets value # at stage ω; more generally, $\underline{\mathrm{K}}$ gets value 0 at odd ordinals, 1 at even successor ordinals, and # at limit ordinals. With the obvious inversion of values for $\neg \underline{\mathrm{K}}$, we see that $(\underline{K} \rightarrow \neg \underline{K})$ has a value that oscillates as well (e.g. it is 1 when $\underline{K}$ has value 0, and 0 when $\underline{K}$ has value 1), which shows that its 'ultimate' value is #.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

natural language semantics. Furthermore, for sentences that do not involve conditionals, Field's analysis is similar to Kripke's; which means that for the (conditional free) examples we used to motivate our own approach, Field's theory *as considered as a semantics for natural language* does not improve on Kripke's theory.[34] But this is not so much a criticism of Field's analysis as a reminder that his goal was not to give a linguistically motivated treatment of English connectives, but rather to find a logic which can accommodate the Liar while still preserving a notion of implication that is strong enough to make 'sustained reasoning' possible.

Still, our own theory was built in part to obtain an analogue of Property 4 and Property 5. Instead of relying on the iteration of a determinacy operator (which yields complications when the iteration is infinite), our analysis generates a hierarchy of α-Liars of the form $< \lambda_\alpha, \neg^\alpha \mathrm{Tr}(\lambda_\alpha) >$; and it also to make it possible to state of each of them that it is something other than true, thanks to the statement $\neg^{\alpha+1} Tr(\lambda_\alpha)$, which comes out as true in our system.

Importantly, two questions about the expressive power of the theory should be clearly distinguished, as Yablo (2003) emphasizes,: (i) "Is the language able to characterize as defective every sentence that deserves to be so characterized?" (ii) "Are there intelligible semantic notions such that paradox is avoided only because those notions are not expressible in the language?" Field's and the present theory give a positive answer to (i), but also to (ii). For in each case there is a semantic property of 'having a value different from true' within the domain of *all* values; but if this notion were expressible in the language, a strengthened Liar could be expressed that would lead to inconsistency. Still, our theory enjoys a limited form of expressive completeness, which is possible because the domain of truth values is infinite—so that within any finite domain of truth values expressive completeness does not lead to inconsistency. Field's theory has no analogue, since when there is a finite number of truth values limited expressive completeness is the same thing as standard expressive completeness—which leads straight back into inconsistency.[35]

7.4. Cook (2008a, 2008b). Cook (2008a) forcefully argues that the Revenge Liar should be embraced, not avoided:

> The problems associated with the Liar paradox and similar phenomena can be traced to the fact that language is, in principle, indefinitely extensible. (. . .) Recall that a concept is indefinitely extensible if and

[34] The 'determinately' operator D, used to generate Field's hierarchy of Liars, is itself defined from $\rightarrow$ and therefore has a noncompositional semantics. Furthermore, even if *D* were used to give a semantics for the English word *determinately*, this would still be insufficient to yield a linguistically realistic account of our initial examples (e.g., with *is something other than true*), which were constructed without any operator-like expression.

[35] Field (2008) develops a version of his theory which is compositional, but has many more values. In a nutshell, in this 'fine-grained' semantics the value of a sentence *S* is a function from ordinals α to the value *S* takes in the revision procedure at stage $\Delta_0 + \alpha$, where Δ_0 is the first 'acceptable ordinal', that is, the first ordinal at which every sentence has its 'ultimate' value. A comparison of this version of Field's system would require a much longer discussion; but it is clear that both the nature of the values (which are ordinals for us, but functions from ordinals to truth values for Field) and their interpretation (revision theoretic for Field but not for us) distinguish sharply between the two systems. (In Field, 2008, pp. 264–266, Field does offer an interpretation of his conditional in terms of Lewis's semantics for counterfactuals; but Field's 'worlds' are entirely different from the possible worlds posited by Lewis, and there is certainly no claim on Field's part that there independent linguistic evidence to have them play a role semantics of *if*-clauses).

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

> only if there is a rule such that, given any definite collection of objects falling under the concept, application of the rule to that collection provides a new object also falling under the concept. The Revenge Problem, properly understood, provides just such a rule. Given a definite language L, let TV_L be the collection of truth values required to correctly interpret L. Then the appropriate version of the Strengthened Liar (i.e., *This sentence has a truth value in TV_L other than true*) provides a truth value which is not in TV_L (namely, the truth value of that very sentence, the 'next' pathological value).

I couldn't agree more. Still, there are a number of differences between Cook's account and the present one.

On a *methodological level*, Cook appears to draw his inspiration from logical rather than strictly linguistic considerations: his concern is primarily to define material implications that make it possible to reason within a multivalent system; in this respect, Cook's general goals are closer to those of Field (2008) than of the present paper (see also Myhill, 1984, for a proof-theoretic system that includes a hierarchy of implications).

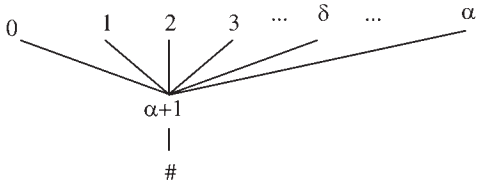
On a *technical level*, there are interesting similarities and differences between Cook's system and our own. Although he doesn't define his language and his fixed points as we do, there are deep similarities between the two methods. On the other hand, the semantics he assumes for his connectives is derived from the Weak Kleene logic, which makes his semantics considerably more cautious than ours in the assignment of classical values.

□ *Construction of the language and of a fixed point*

Cooks constructs a series of languages by stages; by contrast, we define our semantics for an entire language. But the difference might not be significant: each of Cook's languages is a strict superset of the preceding ones, and thus we can consider the union of all his languages to be comparable to the language that we define from the outset. It is also striking that Cook constructs his fixed points 'by stages', just like we do (technically, he constructs an interpretation for each language, whereas we had to rely on the Extension Theorem to guarantee that local fixed points could be extended to global ones). For technical reasons, Cook must stipulate that a sentence of a lower language that tries to refer to a sentence that only exists in a higher language gives rise to a semantic failure—hence the use of an additional truth value *n*, henceforth written as #, to take care of these cases. But when we consider in detail the way in which he constructs his fixed points, we see that it parallels our construction of 'local fixed points' for 'referentially closed sets' of sentences. Specifically:

— In the base of his revision procedure for the language $L_{\alpha+1}$, he puts all sentences that were not in L_α in the α+1-extension of the truth predicate.

— As I understand him (correcting the letter but, I believe, not the spirit of his paper), he posits the following ordering of truth values, where δ is any value smaller than α+1:



https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

This is exactly the strategy we followed when we defined our notion of 'weak monotonicity':

(41) $I \leq_S^\beta I'$ iff (i) all the sentences of S are in a δ-extension of *Tr* for $\delta \leq \beta$, and (ii) for every $\delta < \beta$, $I^\delta(Tr) \subseteq I'^\delta(Tr)$. Then: if $I \leq_S^\beta I'$, $f(I) \leq_S^\beta f(I')$.

Modulo the absence of the value # (i.e., Cook's n) in our system, taking $\beta = \alpha+1$ we get exactly the same ordering as Cook, which is indeed monotonic.

□ *Semantics of the connectives*

As mentioned, Cook's system is a generalization of the Weak Kleene logic, whereas ours extends Strong Kleene. This doesn't exhaust the differences, however: Cook bases his system on a 'smart'—and ramified—notion of implication, whereas we have no primitive implication but many 'smart' negations. The differences are worth investigating in greater detail.

Cook defines connectives which we will call $\neg$, $\wedge^c$, $\vee^c$, and $\rightarrow_\beta$ (for each ordinal β). His logic countenances ordinal-many truth values, just like ours (he uses different names for these values; we will normalize the notations). And it includes predicates that are intended to capture the fact that a sentence has value 0, 1, or β (for any indeterminate value β); we will leave these aside in the following discussion. As mentioned, an additional value, #, is obtained when a predicate has an argument that denotes a sentence that is not in the language. The semantics for conjunction, disjunction, and negation is a straightforward generalization of the Weak Kleene semantics:[36]

(42) a. $I(G \wedge^c H) =$
iff $I(G) = \#$ or $I(H) = \#$
1 iff $I(G) = I(H) = 1$
0 iff $I(G)$, $I(H) \leq 1$ and ($I(G) = 0$ or $I(H) = 0$)
Max {I(G), I(H)} otherwise.
b. $I(G \vee^c H) =$
iff $I(G) = \#$ or $I(H) = \#$
1 iff $I(G)$, $I(H) \leq 1$ and ($I(G) = 0$ or $I(H) = 0$)
0 iff $I(G) = I(H) = 0$
Max {I(G), I(H)} otherwise.
c. $I(\neg G) =$
iff $I(G) = \#$
1 iff $I(G) = 0$
0 iff $I(G) = 1$
I(G) otherwise.

Cook's semantics for $\rightarrow_\beta$ is also in the spirit of Weak Kleene, in the sense that the connective returns an indeterminate value greater than β as soon as one of its argument has such a high indeterminate value. Thus if $I(G) > \beta$, and if $I(H) = 1$, $G \rightarrow_\beta H$ gets the value I(G), although one might be tempted to say that in this case the implication is true because its consequent is. When both I(G) and I(H) are at or below β, $I(G \rightarrow_\beta H)$ has value 1 in case $I(G) = 0$, or $I(H) = 1$, or $I(G) \geq I(H) \geq 2$, and otherwise it has value 0:

[36] We reproduce Cook's semantics with our own notational conventions. In particular, we write # for Cook's n, α for Cook's p_α, and 0 and 1 for Cook's f and t. For conjunction, Cook gives a definition in terms of Min rather than Max, but that's because he posits (for this definition only) a *reversed* ordering of truth values.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

(43) Let $\rightarrow_\beta$ belong to the language L_α obtained at stage α.
$I(G \rightarrow_\beta H) \neq \#$ iff $I(G) \neq \#$ and $I(H) \neq \#$.
If $\neq \#$, $I(G \rightarrow_\beta H) =$
$\text{Max}\{I(G), I(H)\}$ if $I(G) > \beta$ or $I(H) > \beta$,
1 if $I(G) \leq \beta$ and $I(H) \leq \beta$ and ($I(G) = 0$ or $I(H) = 1$ or $I(G) \geq I(H) \geq 2$),
0 otherwise.

There are two immediate questions to ask: can Cook's connectives be defined with our own? Can our connectives be defined with Cook's? Four results can be stated at the outset for the propositional case (we leave a deeper comparison for future research). For convenience we assume that our connectives and Cook's are part of the same language, and we restrict attention to the case in which none of the atomic formulas has value #. Since it is trivial that Cook's negation is essentially our $\neg^2$, we do not discuss this case any further.

(44) Let I be any interpretation in which (i) I(G) and I(H) do not have the value # and (ii) $I(\underline{\beta}) = \beta$.
Defining our connectives with Cook's
a. $I(\neg^{\beta+1}G) = I(\underline{\beta} \rightarrow_\beta \neg G)$
b. $\wedge$, $\vee$ cannot be defined in terms $\{\wedge^c, \vee^c, \neg, \rightarrow_\alpha: \alpha \in \text{Ord}\}$
Defining Cook's connectives with ours
c. In general, $\rightarrow_\beta$ cannot be defined in terms of $\{\wedge, \vee, \neg^\alpha: \alpha \in \text{Ord}\}$
d. $I(G \wedge^c H) = I((G \wedge H) \vee \neg^2(G \wedge \neg G) \vee \neg^2(H \wedge \neg H))$
$I(G \vee^c H) = I((G \vee H) \wedge (G \vee \neg G) \vee (H \vee \neg H))$.

Together, these results show that Cook's system and ours are not intertranslatable, at least not without additional predicates or connectives. Something similar to our negation can be defined with his connectives, but our conjunction and disjunction cannot be so defined. His conjunction and disjunction can be defined with our connectives, but his implications cannot be.

Proof of (44)a. If $I(G) \geq \beta+1$, $I(\neg G) = I(G)$ and $I(\underline{\beta} \rightarrow_\beta \neg G) = I(G)$. If $I(G) < \beta+1$, $I(\underline{\beta} \rightarrow_\beta \neg G) = 1$ iff $0 < I(\neg G) < \beta+1$ and $I(\underline{\beta} \rightarrow_\beta \neg G) = 0$ iff $I(\neg G) = 0$; hence $I(\underline{\beta} \rightarrow_\beta \neg G) = 1$ iff $I(G) = 0$ or $1 < I(G) < \beta+1$ and $I(\underline{\beta} \rightarrow_\beta \neg G) = 0$ iff $I(G) = 1$.

Proof of (44)b. We note that Cook's connectives have the property (P):

(P) If $F = (G \wedge^c H)$, $(G \vee^c H)$, $\neg G$, $(G \rightarrow_\alpha H)$, there is an ordinal β such that if at least one of the immediate arguments of F has value $\geq \beta$, so does F.

It is immediate that F cannot define $(G \wedge H)$ or $(G \vee H)$ unless G and H figure in F. It follows from (P) (by induction on the construction of F) that there is an ordinal β such that if at least one of G, H has value $\geq \beta$, so does F.But such is not the case of $(G \wedge H)$ or $(G \vee H)$ (the former has value 0 as soon as G or H does, and the latter has value 1 as soon as G or H does).

Proof of (44)c. If F is a formula formed from the atoms G and H, and if $I(G) = g$ and $I(H) = h$, let us abbreviate $I(F)$ as $f(g, h)$. The following property is immediate given Cook's semantics:

(C) For infinite α, if $F = (G \rightarrow_\alpha H)$, for every g for which $2 < g < \omega$, $f(g, g) = 1$ while $f(g, g+1) = 0$.

But we will prove that for any formula F formed from G, H and $\{\wedge, \vee, \neg^\alpha: \alpha \in \text{Ord}\}$, (C) must fail. We can restrict attention to formulas that do not contain $\vee$, since $\vee$ is definable from $\wedge$ and $\neg^2$.

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

Let i* be defined as i* = 2 if F contains no negation with a finite index; otherwise, i* = the greatest finite index that appears on a negation of F. We consider g with i* < g < ω, and prove that it cannot be that f(g, g) = 1 and f(g, g+1) = 0.

We note that for every subformula F' of F, one of three cases arises:

(i) {f'(g, g), f'(g, g+1)} $\subseteq$ {g, g+1};

(ii) {f'(g, g), f'(g, g+1)} = {0};

(iii) {f'(g, g), f'(g, g+1)} = {1}.

— The atomic case is trivial: since the only atoms are G and H, {f'(g, g), f'(g, g+1)} $\subseteq$ {g, g+1}.

— If the property holds of A and B, it holds of $F' = (A \wedge B)$ as well.

- If A or B is in case (ii), it is immediate that F' is as well.
- If both A and B are in case (i), so is F'; similarly, if both A and B are in case (iii), so is F'.
- If A is in case (i) and B is in case (iii), F' is in case (i); and similarly if A in case (iii) and B is in case (i).

— If the property holds of A, it holds of $F' = \neg^{\alpha} A$ as well. The result is immediate if A is in case (ii) or (iii). If A is in case (i), F' is in case (i) if $\alpha \leq$ i* (since in this case I($\neg^{\alpha}$A) = I(A)), and F' is in case (iii) if $\alpha \geq \omega$ (since in this case I($\neg^{\alpha}$A) = 1).

The property applies to every subformula of F, including F itself. But it is clear that for F = (G $\rightarrow_{\alpha}$ H) the property fails, since {f(g, g), f(g, g+1)} = {0, 1}.

Proof of (44)d.

(i) G $\wedge^{c}$H

The argument is most easily laid out in a truth table. For brevity, we write I(G) and I(H) (rather than specific values) when I(G) $\geq$ 2 or I(H) $\geq$ 2.

G	**H**	**G $\wedge$H**	**$\neg^2$(G $\wedge\neg$G)**	**$\neg^2$(H $\wedge\neg$H)**	**(G $\wedge$H) $\vee\neg^2$(G $\wedge\neg$G) $\vee\neg^2$(H $\wedge\neg$H)**
0	0	0	0	0	0
1	1	1	0	0	1
0	1	0	0	0	0
1	0	0	0	0	0
I(G)	0	0	I(G)	0	I(G)
I(G)	1	I(G)	I(G)	0	I(G)
0	I(H)	0	0	I(H)	I(H)
1	I(H)	I(H)	0	I(H)	I(H)
I(G)	I(H)	Max(I(G), I(H))	I(G)	I(H)	Max(I(G), I(H))

(ii) G $\vee^{c}$H

Both in Cook's system and in ours, standard rules of duality hold (using as negation the connective $\neg^2$ in our system). The result immediately follows from (i).

§8. Conclusion. It has become standard to assume that a third truth value must be introduced to handle paradoxes in natural language. But once this step is taken, our analysis of the Strengthened Liar suggests that it should be iterated indefinitely. The main difficulties already arise at the very first step: the semantics of natural language must somehow be 'tweaked' to handle the third truth value; once a mechanism is found for this case (typically

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

along the lines of the Strong Kleene scheme), it can be extended to logics that have further truth values—in fact, ordinal-many truth values.

We have attempted to offer a detailed analysis of the linguistic underpinnings and formal properties of a hierarchy of Super Liars in natural language. Focusing on the expression of negation, we saw that the quantificational nature of the expression *is something other than true* can be analyzed literally, which naturally yields a hierarchy of Liars of increasing strength as the implicit domain restriction of *something* gets gradually expanded. A related analysis can be offered by focusing on the predicate *true* alone, which may come with selectional restrictions that may change as one learns that there are more truth values than one initially thought. Of course these attempts are still fragmentary, and they should in the end be integrated into a general study of the linguistic means by which the Liar can be expressed. This is a difficult task, but one which cannot be avoided if the goal is to give a semantics for natural language.

We should also emphasize some limitations of the present analysis.

—As was clear from our discussion of quantificational languages, we sometimes need restrictions on the expressive power of the language if we wish to guarantee that some fixed points can be found. It must be noted, however, that this limitation only holds when the language does not form a set.

—Even in the simplest cases, however, a version of the Revenge Liar cannot be handled by the theory. We might want to introduce an absolute negation $\neg^{Ord}$, whose domain restriction is the class of all ordinals. And we might want to use it to define an absolute Super Liar as $< \lambda^{\mathrm{Ord}}, \neg^{\mathrm{Ord}}\mathrm{Tr}(\lambda^{\mathrm{Ord}}) >$. But it is immediate that no truth value could be assigned to $\underline{\lambda}^{\mathrm{Ord}}$: for the usual reasons, it could not be assigned a classical value; but in addition it could not be assigned an indeterminate truth value, because if so it would have an ordinal value different from 1, which should make it true, contrary to hypothesis. The Super Liar gets his revenge in the end—but I refer the reader to Cook (2008a) for an argument that this is entirely in order if one takes truth to be an indefinitely extendible concept.

Still, our analysis improves on Kripke's theory in two related respects. First, we account for a considerably broader fragment of English than Kripke, since we give a plausible analysis of several forms of weak negation, and explain how speakers can come to have judgments about a variety of Super Liars. Second, our logic has a limited form of expressive completeness that was absent from Kripke's logic. We can express *in the language* (though with some limitations) that any given Super Liar is something other than true. And our logic enjoys a form of limited expressive completeness which is not available in most other accounts.

§9. Acknowledgments. Many thanks to Henri Galinon for very helpful comments, and to two anonymous referees for detailed suggestions and criticisms; the first referee mentioned the existence of Cook (2008a), which proved essential to the final version of this paper. Special thanks to Roy T. Cook for agreeing to send me his manuscripts, and for discussing some issues by e-mail. Helpful remarks were also provided by audiences at University of California, Los Angeles and Institut Jean-Nicod ('3rd Paris-Amsterdam Logic Meeting of Young Researchers'). The author gratefully acknowledges the past financial support of the American Council of Learned Societies ('Ryskamp Fellowship') and of UCLA. This work was supported in part by a 'Euryi' grant of the European Science Foundation ('Presupposition: a formal pragmatic approach').

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

BIBLIOGRAPHY

Burge, T. (1979). Semantical paradox. *Journal of Philosophy*, **76**, 169–198.
Chemla, E. (2009). Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics*, **17**(4), 299–340.
Cook, R. T. (2008a). Embracing revenge: On the indefinite extendibility of language. In Beall, J. C., editor. *Revenge of the Liar*. Oxford University Press.
Cook, R. T. (2008b). *What Is a Truth Value, and How Many Are There?* Manuscript, University of Minnesota.
Feferman, S. (1984). Towards useful type-free theories I. *Journal of Symbolic Logic*, **49**, 75–111.
Field, H. (2003a). A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, **32**, 139–177.
Field, H. (2003b). The semantic paradoxes and the paradoxes of vagueness. In Beall, J. C., editor. *Liars and Heaps: New Essays on Paradox*. Oxford University Press.
Field, H. (2007). Solving the paradoxes, escaping revenge. In Beall, J. C., editor. *Revenge of the Liar*. Oxford University Press, pp. 78–144.
Field, H. (2008). *Saving Truth from Paradox*. Oxford University Press.
Fitch, F. (1964). Universal metalanguages for philosophy. *Metaphysics*, **17**(3), 396–402.
Gajewski, J. (2005). *Neg-Raising: Polarity and Presupposition*. Doctoral Dissertation, MIT: Cambridge, MA.
Galinon, H. (2006). Comments on Schlenker. Hand-out of a talk given at the 3rd Paris-Amsterdam Logic Meeting of Young Researchers (*PALMYR 3*), held at Institut Jean-Nicod, Paris on June 28, 2006.
Glanzberg, M. (2001). The liar in context. *Philosophical Studies*, **103**, 217–251.
Glanzberg, M. (2004). A contextual-hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic*, **33**, 27–88.
Gupta, A., & Martin, R. L. (1984). A fixed point theorem for the weak Kleene valuation scheme. *Journal of Philosophical Logic*, **13**, 131–135.
Horn, L. (1989). *A Natural History of Negation*. CSLI Publications. Reprinted in 2001.
Kadmon, N. (2001). *Formal Pragmatics*. Blackwell.
Kratzer, A. (1991). Conditionals. In von Stechow, A., and Wunderlich, D., editors. *Semantik: ein internationales Handbuch der zeitgenössischen Forschung*, pp. 651–656.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy*, **72**, 690–716.
Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic*, **34**, 155–192.
Myhill, J. (1984). Paradoxes. *Synthese*, **60**, 129–143.
Schlenker, P. (2004). Conditionals as definite descriptions: A referential analysis. *Research on Language and Computation*, **2**(3), 417–162.
Schlenker, P. (2007). The elimination of self-reference: Generalized Yablo-series and the theory of truth. *Journal of Philosophical Logic*, **36**(3), 251–307.
von Fintel, K. (2001). Counterfactuals in a dynamic context. In Kenstowicz, M., editor. *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press, pp. 123–152.
Yablo, D. (2003). New grounds for naive truth theory. In Beall, J. C., editor. *Liars and Heaps: New Essays on Paradox*. Oxford University Press.

INSTITUT JEAN-NICOD
ECOLE NORMALE SUPÉRIEURE
29, RUE d'ULM
75005 PARIS

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press

DEPARTMENT OF LINGUISTICS
NEW YORK UNIVERSITY
10 WASHINGTON PLACE
NEW YORK, NY 10003
E-mail: philippe.schlenker@gmail.com

https://doi.org/10.1017/S1755020310000067 Published online by Cambridge University Press