Supervaluation on Trees for Kripke's Theory of Truth

Casper Storm Hansen

Abstract: A method of supervaluation for Kripke's theory of truth is presented. It differs from Kripke's own method in that it employs trees; results in a compositional semantics; assigns the intuitively correct truth values to the sentences of a particularly tricky example of Gupta's; and – it is argued – is acceptable as an explication of the correspondence theory of truth.

In his (1982) Gupta presents the following scenario in order to criticize Kripke's (1975) theory of truth: Assume that the sentences

A1: Two plus two is three
A2: Snow is always black
A3: Everything B says is true
A4: Ten is a prime number
A5: Something B says is not true

are all that is said by a person "A", and the sentences

B1: One plus one is two
B2: My name is B
B3: Snow is sometimes white
B4: At most one thing A says is true

are all that is said by a person "B". The sentences A1, A2, and A4 are clearly false and B1, B2, and B3 are clearly true. So it seems unobjectionable to reason as follows: A3 and A5 contradict each other, so at most one of them can be true. Hence at most one thing A says is true. But that is what B says with his last sentence, so everything B says is true. This is again what A says with A3 and rejects with A5, so A3 is true and A5 false.

But counterintuitively Kripke's theory in its strong Kleene scheme, minimal fixed point version (with which I will assume familiarity and hereafter refer to as the "basic version") tells us that A3, A5, and B4 are all undefined. The reason is that the evaluation of A3 and A5 awaits the determination of B4, which in turn cannot receive a truth value before A3 or A5 do.

One way to obtain the intuitively correct truth values is to swear allegiance to one of the theories that assign truth values in a holistic manner.[1] But that

[1] See for example (Walicki 2009), according to which truth values are not assigned in a

should not be necessary; the truth of B4 and A3 and the falsity of A5 are intuitively *grounded*. The truth of B4 is intuitively grounded merely in the facts that A1, A2, and A4 are false and that A3 and A5 are contradictory. No specific assignment of truth values to A3 and A5 is presupposed. However, when B4 has been assigned the value of truth, the truth of A3 and the falsity of A5 are subsequently grounded in the truth of B1–B4.

I will argue that a proper theory of correspondence – or grounded – truth ought to deliver the intuitive result (section 1). It will further be argued that none of the versions of Kripke's theory (sections 1 and 2) nor revision theory (section 3) is a satisfactory theory in that respect. Then an alternative will be explored (from section 4 onwards). It is a modification of Kripke's theory using a method of supervaluation on trees. It is not the ambition to deliver The Correct Theory of Grounded Truth, just to introduce the tree method which I believe is an important step towards such a theory but doesn't get us there by itself.

1 Correspondence and grounded truth

"Grass is green" is true because the sentence corresponds in the right way to the fact that grass is green. This is one instance of the general claim that is the correspondence theory of truth.[2] The correspondence relation is a relation from the class of sentences to the class of facts.

Another instance is that "'grass is green' is true" is true because this sentence corresponds to the fact that "grass is green" is true. This instance reveals that even though the domain and codomain of the correspondence relation are disjoint, every fact gives rise to an iteration of truths. The reason is that each truth begets a fact, namely the fact that the sentence in question is true. So the correspondence relation has its basis in the class of non-semantical facts but also, so to speak, feeds into itself to form an infinite hierarchy of truths further and further away from that basis.

I take "grounding" to be a generalization of correspondence. First, grounding is a transitive relation while correspondence is not. The truth of "'grass is green' is true" is grounded in the fact that grass is green, but it does not correspond to it. Second, grounding is also a relation between (the truth or falsity of) complex sentences and (the truth or falsity of) their constituents. For instance, the truth of a disjunction can be grounded in the truth of one of its disjuncts, while it does not seem right to say that it corresponds to it.[3]

A simplified explanation of what iterated grounding consists of is the following: We begin with a basis consisting of non-semantical facts. Then we

stage-by-stage process as in Kripke's and Gupta's theories, but "simultaneously". A rough formulation of the theory is that any assignment of truth values that satisfies certain compositional demands and minimizes the number of sentences to be declared undefined is acceptable. See also (Wen 2001).

[2] See for instance (Austin 1950) and (Kirkham 1992).

[3] Attempts have also been made to generalize grounding beyond semantics to also cover, for example, the relation that obtains between a set and its members, and even further, to make grounding serve as an all-encompassing metaphysically explanatory concept (Schaffer 2009). This direction of generalization is beyond the scope of this paper.

"make" all those sentences that correspond to those facts true. Now there are a number of semantical facts in addition to the non-semantical facts, namely, facts about which sentences are true. Then we make all those sentences that correspond to the enlarged set of facts true. And so on.

This idea can already be found in Tarski's (1933, 1944) theory of truth. With his basic theory, Kripke contributed two things. First, he showed how to put the different levels of this iteration into one language (although by his own admission (1975, p. 714) the success here is only partial). Second, he showed how to capture formally the intuitive verdict that the idea of grounding provides in the case of the Watergate example and structurally similar cases.

The Watergate example consists of Jones stating that "Most of Nixon's assertions about Watergate are false", Nixon asserting that "Everything Jones says about Watergate is true", and both of them making a number of other assertions about Watergate. Starting out from the basis of non-semantical facts, all of those other assertions may be made true or false in the first iteration. Assuming that Nixon has made a total of seventeen assertions about Watergate, it is now a fact that, say, six of Nixon's assertions about Watergate are true and that ten are false. That is a fact that the quoted utterance of Jones can correspond to in the next iteration so as to be made true. If all Jones' other claims on the subject are also true, then there is thereafter a fact that the quoted utterance of Nixon corresponds to.

The reason why the explanation three paragraphs back is a simplification is that the basis is not completely devoid of semantic facts. Facts concerning what sentences are *about* are semantic facts, and they have to be included in the basis, if the truth values of the Watergate sentences are to come out right. The only semantic facts that cannot be in the basis are those that are defined by the iteration. All facts available at a given level can be used for grounding.

With this in mind, consider Gupta's example again. In the first iteration, A1, A2, A4, B1, B2, and B3 can be given proper truth values (true or false; the value of "undefined" will be called a truth value but not a *proper* truth value) by corresponding (positively or negatively) to non-semantic facts. At this point, the fact that A3 and A5 contradict each other is also available, so in combination with the falsity of A1, A2, and A4, we have the facts that B4 need to correspond to in order to be true. Subsequently, A3 can be made true and A5 false by correspondence to the truth of B1-B4.

All nine sentences can be given grounded truth values. Therefore, these sentences constitute a challenge to any theory which purports to be about grounded truth, such as Kripke's. Let us call it "Gupta's Challenge". Kripke (1975, p. 706) defines "grounded", as a predicate applicable to sentences, as meaning that the sentence has a proper truth value in the minimal fixed point. One thing that Gupta's Challenge teaches us is that the minimal fixed point of the strong Kleene valuation schema does not fully capture the informal notion of groundedness. We will aim at doing better in that respect. In order to focus the paper and avoid having to justify every design choice and engage with every competing theory on the market, it will be assumed, rather than argued for, that capturing the informal notion of groundedness is what a theory of truth should aim for.

Note also that the intuitive reasoning for the truth values of the nine sentences does not employ any "dangerous" principles, i.e. principles that could lead to inconsistency if used in other situations. The only principle used here that goes beyond what is validated by the basic version of Kripke's theory is the inference rule, *from the inconsistency of two sentences, conclude that at most one of them is true*, which is perfectly benign. Hence, there is no good excuse for a theory that delivers another result.

That there are two sentences that cannot both be true is a semantic fact that is not defined by the iteration, but is available from the outset.[4] Hence, when A1, A2 and A4 are made false, the fact of those falsities joins an already existing fact about A3 and A5 being inconsistent. The "reality side" of the correspondence relation is then in place to make B4 true. That is what the basic version of Kripke's theory misses: Gupta's Challenge and structurally similar cases.

The formal iteration of Kripke's basic theory only makes use of 1) non-semantic facts, 2) the semantic facts that can be encoded in the interpretation function which provides the semantics of the constants and the ordinary predicates of the language, and 3) semantic facts consisting of the actual assignment of truth values that have been made at previous levels. One thing is left out with respect to the intuitive iteration he is trying to formalize: facts about which combinations of truth values are possible (given the other information).

This calls for a form of supervaluation: if a particular sentence is true relative to all possible assignments of truth values (given all other facts available at a given level), then that is a fact to which the truth of that sentence can correspond. Hence, the attempt that Kripke makes to formulate a supervaluation version of his theory is well motivated. Unfortunately, that attempt goes too far to be correspondence-theoretically acceptable; in some cases it declares that a sentence is true given all possible assignments of truth values, when in fact it is not. That is one of the issues to be discussed in the next section.

A hierarchical truth theory is correspondence-theoretically acceptable if and only if, every time a sentence is made true (false) at a level, there is an already established (negative) fact for it to correspond to. There may be facts about which combinations of truth values are (im)possible that are established prior to the sentences in question getting specific truth values, and making sentences true by correspondence to such a fact is correspondence-theoretically acceptable – but of course only if it is a genuine fact! If the class of possibilities considered is too narrow, we get "false positive" verdicts about impossibilities of combinations, and using such verdicts will make a theory correspondence-theoretically unacceptable.

[4] Or rather: may be. It is a fact from the outset that ϕ and $\neg\phi$ cannot both be true. If ψ becomes false at level 7, then at level 7 it also becomes a fact that $\psi \vee \phi$ and $\neg\phi$ cannot both be true.

To sum up, I will take two of the success criteria for a theory of truth to be that it can reasonably be seen as an explication of the correspondence theory and that it can handle Gupta's Challenge adequately. A third success criterion for a theory of truth, to be introduced in section 2, is that it is compositional (i.e. that the connectives and quantifiers are truth-functional).

2 Kripke's supervaluation

In this section I discuss Kripke's supervaluation versions of his theory and in the next, Gupta's own revision theory of truth, in order to make the case that these theories are not successful in the stipulated senses of being correspondence-theoretically acceptable, handling Gupta's Challenge adequately and resulting in a compositional semantics.

The simplest version of supervaluation considered by Kripke is the one in which sentences are supervaluated as true (false) if they are evaluated as classically true (false) relative to all total extensions of the evaluations at a given level.[5] This version has no effect upon the sentences of Gupta's example compared to the basic version. The reason is that at every level there is, among the total evaluations quantified over when making the supervaluation, an evaluation in which both A3 and A5 are true, so B4 does not become true. This problem can be remedied by tweaking the theory a bit. Instead of quantifying over all total extensions of the given evaluation, we can restrict attention to the maximally consistent ones.[6] There is no maximally consistent total evaluation in which both A3 and A5 are true, so B4 becomes true; and then at the next level, A3 is made true and A5 false.

This does not really solve the problem, however. After the theory has been tweaked, Gupta can (and does) tweak his example to make the problem reappear by replacing A3 and A5 with A3* and A5*:

A3*: "Everything B says is true" is true
A5*: "Something B says is not true" is true

These sentences are not contradictory in the proper technical sense, only intuitively, so again B4 does not become true.

Taking a step back from the issue of which sentences are given which truth values, there is a more fundamental reason for rejecting the supervaluation versions of Kripke's theory. As argued above, the basic version of the theory can be defended from a correspondence-theoretical point of view; a sentence

[5] More precisely: Let $\mathcal{S}$ be the set of sentences for a suitable formal language. For each classical interpretation of the truth predicate (i.e. a subset of the domain) t, let T_t be the set of classically true sentences relative to t and some fixed model that interprets all constants and all other predicates, and let F_t be $\mathcal{S} \setminus T_t$, the set of classically false sentences relative to t and that model. This simple supervaluation jump takes a partial interpretation (T, F) of the truth predicate to another partial interpretation (ST, SF) where ST (SF) is the set of sentences ϕ such that for all t, if $T \subseteq t \subseteq \mathcal{S} \setminus F$ then $\phi \in T_t$ $(\phi \in F_t)$. See (Kripke 1975, p. 711).

[6] Add the proposition that t is maximally consistent as a conjunct to the antecedent of the condition for ϕ in footnote 5.

is declared true when the state of affairs expressed by the sentence is the case according to the model or a lower level, both of which can reasonably be regarded as (formalistic) counterparts to reality.[7] This is not the case in the supervaluation versions, which declare a sentence true when the state of affairs expressed by the sentence is the case according to a class of fictions. Let me explain.

In the previous section I concluded that if a sentence is true relative to all possible assignments of truth values (given all other facts available at a given level), then that is a fact to which the truth of that sentence can correspond. However, Kripke's form of supervaluation does not check that the sentence is true relative to *all* possible assignments of truth values before the sentence is made true. For the final evaluation (the fixed point) must surely be considered a possible evaluation – what is actual is possible. And the final evaluation can, in the supervaluation versions just as in the original, be non-total, if there is vicious self-reference in the language. In that case the total evaluations quantified over in the supervaluation are not all the possible assignments, for they do not include the evaluation that ends up being the actual evaluation. The total evaluations are, rather, expressions of the fiction that we are in a bivalent setting when in fact we know that we are in a trivalent one. Therefore it is not sufficient to consider what is true in these; doing so gives "false positive" verdicts on matters of which combinations of truth values are possible. Ergo, Kripke's supervaluation theories are not adequate as explications of the correspondence theory of truth.

The idea I am relying on here is the following: Reality consists of some non-semantic facts plus some semantic facts that are grounded in the non-semantic facts. The fixed point consists of the latter and is therefore a part of reality. In the basic version of Kripke's theory, the levels prior to the fixed point contain correct, but partial, information about the totality of semantic facts. What is grounded in these is therefore grounded in reality. On the other hand, what is "grounded" only in evaluations that disagree with the fixed point is based on fictions.

A concrete symptom of this philosophical problem is that Kripke's method for supervaluation declares all classically valid sentences true. This is very misleading in a semantics that is strictly weaker than classical semantics. A prominent example is that the disjunction of the Liar and its negation is true even though neither of the disjuncts is made true. When taking the step from bivalent to trivalent semantics, it seems clear that the property *from the information that a disjunction is true, it is possible to conclude that at least one of the disjuncts is as well*, is more important to preserve than the property *every disjunction of a sentence with its negation is true*. The former is needed for the inference rule of disjunction elimination to be valid. The latter can hardly be considered a desideratum at all (given that there are sentences ϕ such that neither ϕ nor $\neg\phi$ is true); it rather seems like

[7] A lower level contains *no* false information about what turns out to be the final evaluation. For when, in the basic version of Kripke's theory, one evaluation is used to define the next one, undefinedness is interpreted as lack of information rather than as information about lack of truth value.

dishonesty.

The restriction to consistent evaluations does not prevent the disjunction of the Liar and its negation from being made true. Neither does any other restriction, for quantifying over fewer evaluations can only make more sentences true or false. In short, every version of Kripke-style supervaluation results in a non-compositional semantics.[8]

Kripke's basic theory does not go far enough to capture all that is intuitively grounded, while his supervaluation versions go too far. There is a middle ground to be seized, one in which supervaluation is employed, but over trivalent evaluations.

3 Revision theory

We obtain the intuitively correct truth values for A1–A5 and B1–B4 if we accept a revision theory of truth. One version of such a theory[9] is that there are a number of best candidates for the extension of truth, namely the extensions that appear again and again when the revision process has settled into a loop.[10]

These best candidates give the correct truth value to a lot of sentences that, according to intuition, are unproblematically true or false. So for example A1–A5 and B1–B4 (and the *-variant) have the right truth values in all of the

[8]Meadows (2013) – building on work of Leitgeb (2005) – explores a method for doing supervaluation that is quite close to Kripke's but results in a classical evaluation. *A fortiori*, it avoids the problem of true disjunctions with no true disjuncts. However, another problem takes it place: there are true sentences ϕ such that the sentence $\neg T(c_\phi)$ saying that ϕ is not true is also true. This problem is also related to grounding. Meadows is concerned with constructing a grounded extension of the truth predicate ("e-true" in his terminology), and in that endeavor he succeeds (under the assumption that two is the right number of truth values). The set of true ("m-true") sentences, on the other hand, is not grounded according to the standards of this paper. For example, $\neg T(c_\phi)$ lacks a ground. This is possible because what is true depends in part on the anti-extension of the truth predicate, and that anti-extension fails to be grounded: ϕ can be in it when it should not, resulting in $\neg T(c_\phi)$ becoming true.

[9]The aim of this section is to see whether the *machinery of* revision theory can be used to formulate an acceptable theory of grounded truth in the sense of section 1. I am not engaging with revision theory as it is interpreted philosophically by Gupta, Belnap and Herzberger, for their aim is quite different.

[10]Let t, T_t and $\mathcal{S}$ be as in footnote 5. Define $T^\alpha(t)$ by recursion on the ordinal α as follows:

$$T^\alpha(t) = \begin{cases} t & \text{if } \alpha = 0 \\ T_{T^{\alpha-1}(t)} & \text{if } \alpha \text{ is a successor ordinal} \\ X^\alpha \cup ((\mathcal{S} \setminus Y^\alpha) \cap t) & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

where

$$X^\alpha = \left\{ s \;\middle|\; \exists \beta < \alpha \left(s \in \bigcap_{\beta \leq \gamma < \alpha} T^\gamma(t) \right) \right\}$$

and

$$Y^\alpha = \left\{ s \;\middle|\; \exists \beta < \alpha \left(s \notin \bigcup_{\beta \leq \gamma < \alpha} T^\gamma(t) \right) \right\}.$$

The set $\mathfrak{B}$ of best candidates for the set of truths is the set of evaluations B, such that for some interpretation t, for all ordinals α, there is an ordinal $\beta > \alpha$ such that $B = T^\beta(t)$ (Gupta 1982, p. 44–45).

best candidates, and every sentence that is true (false) in the basic version of Kripke's theory is also true (false) in the best candidates. Yet, arguably, none of the *best* candidates are *good*. We are right back at the problems that drove Tarski to "ban" self-reference: the Liar is true in some of the best candidates and false in others, both of which are bad, while "The Liar is true" has the opposite truth value.[11] Ergo, the T-schema is not validated, so according to Tarski's (1944, p. 344) adequacy condition, these candidate truth predicates are not truth predicates at all.

These evaluations cannot reasonably be seen as correspondence-theoretically acceptable solutions to the paradoxes. They can, at best, be credited with being adequate theories of the semantics of non-pathological sentences.[12] But I wouldn't even go that far, for the advantages of revision theory over Kripke's theory are achieved using techniques that are not justifiable from the standpoint of the correspondence theory of truth, for the construction of each of the sequences is not based on reality alone, but also on an initial evaluation that is not reality but a fiction.[13]

A different version of revision theory can be seen as an attempt at dealing with this problem of reliance on fictions in the first version, namely the version were we consider the "real" truths to be the stable truths, and the "real" falsehoods to be the stable falsehoods.[14] We supervaluate over all initial evaluations and all the best candidates they result in, obtaining *one* definite set of truths and *one* definite set of falsehoods. This method also gives us sentences that are neither. Again, we have a trivalent semantics.

The best candidates for the extension of truth are unacceptable because they are each based on just one initial evaluation which is a fiction and not reality. Supervaluations can be seen as an attempt to resolve this problem. The rationale would be that while *one* evaluation is a fiction, what is true relative to *all* possible evaluations is based upon facts; so the truths and falsehoods that are common to all possible evaluations are based on reality, no matter the contingent state of reality, and thus *they* are correspondence-theoretically acceptable. However, this rationale does not hold, for the supervaluation gives rise to a trivalent semantics, and when one accepts such a semantics, one can no longer hold the set of all bivalent truth-value ascriptions to represent all possible realities. Here there is an interesting catch-22 phenomenon: in the attempt to consider all possible realities, one has to admit that they are not all the possible realities after all. For a supervaluation theory to be acceptable, the kind of evaluations (bivalent, trivalent) quantified over must be of the same kind as the possible outcomes of this supervaluation.[15] I will

[11]At least that is the case with the "successor stage candidates". With some of the limit stage candidates, it is not. But in these, the truth values of some sentences are just the arbitrary value of the initial evaluation, so any advantage these candidates may have over the successor stage candidates is purely a result of arbitrariness.

[12]And perhaps as an interesting *analysis* of the paradoxes in the form of the different revision patterns to which they give rise. This seems to be Herzberger's (1982) point with his version of the revision theory.

[13]Gupta most often uses the word "hypothesis", but also in one instance (1982, p. 38) calls such an evaluation a "fiction".

[14]A sentence s is stably true if $s \in \bigcap \mathfrak{B}$ and stably false if $s \notin \bigcup \mathfrak{B}$, where $\mathfrak{B}$ is as defined in footnote 10 (Gupta 1982, p. 46).

[15]I only have the present context, i.e. theories of truth, in mind with this claim. There

propose such a theory.

In addition to this problem of philosophical justification, the supervaluation version of revision theory shares the problem of lack of compositionality with the supervaluation version of Kripke's theory. There are disjunctions that are declared true even though none of the disjuncts are, for example the one with the Liar and its negation.

On top of all this, the revision theory does actually not really meet Gupta's Challenge. For the solutions to the two versions considered up to now depend on the number of iterations of the truth predicate in the analogue of A3 being identical with the number of iterations of the truth predicate in the analogue of A5.

The reason that the theory leads to the intuitively correct result in the original version of the Challenge is as follows: No matter what the initial evaluation is, the six sentences without the truth predicate receive the right value at level 1 and maintain them thereafter. Then, no matter what the value of B4 is, one of the sentences A3 and A5 becomes true and the other false. Therefore, B4 becomes true, which then causes A3 to become true and A5 to become false.

Adding a truth predicate to both A3 and A5 does not change much. It just means that for A3* and A5* there is an extra delay of one level before the truth values settle into the right ones.

But if, for example, A5 is replaced with A5* while A3 is retained, the truth values for the sentences do not stabilize for all initial evaluations; for then the two sentences draw their values from different levels, so to speak, and may become true simultaneously, causing B4 to become false. Using the symbols $\top$ and $\bot$ for truth and falsity respectively, this table shows what happens when the initial evaluation is one that assigns falsity to all the sentences involved (the pattern in levels 1–3 is repeated in levels 4–6 and *ad infinitum*):

Level	0	1	2	3	4	5	6	⋯
A3	$\bot$	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$	⋯
A5*	$\bot$	$\bot$	$\top$	$\bot$	$\bot$	$\top$	$\bot$	⋯
B4	$\bot$	$\top$	$\top$	$\bot$	$\top$	$\top$	$\bot$	⋯

The intuitive argument presented in the introduction of this paper is equally strong no matter how many additional truth predicates are applied to some of the sentences. Therefore this failure shows that revision theory does not get to the heart of the problem.[16]

may be other fields, for example the semantics of vagueness, where the use of supervaluations to go from a set of one kind of evaluations to an evaluation of another kind makes good sense.

[16] Variants of the revision theory with different limit rules are proposed in (Herzberger 1982) and (Belnap 1982). According to Gupta, a sentence that has not reached stability at a given limit ordinal should revert to its truth value in the initial evaluation. Herzberger suggests that it should revert to falsity, which will not help. Belnap thinks that we should consider all possible limit rules as long as they retain the truth values of sentences that

It has been argued above that we can safely infer that at most one of A3 and A5 can be true. To conclude from this that at most one of A3 and A5* (this example is easily generalizable) can be true, all that is needed is that A5 is true iff A5* is true. But that is merely an instance of the rule version of the T-schema which holds in the supervaluation version of revision theory and in all versions of Kripke's.[17] The A3/A5* version of the Challenge is discussed by Gupta and Belnap (1993, p. 228) who claim that "The intuitive argument [...] no longer goes through, since now A's statements [A3] and [A5*] do not contradict each other. To show that [A3] and [B4] are true one needs to appeal to an instance of the T-step, e.g., [A5* iff A5] which is not validated [...]". I beg to differ: the *intuitive* argument does go through. Gupta and Belnap provide no philosophical argument in the text, but simply reiterate the consequences of their theory for the case at hand.

From the considerations of this and the previous section we can draw the conclusion that Kripke's supervaluation theories share three problems with the supervaluation version of revision theory. They both result in non-compositional semantics; they both only quantify over two-valued evaluations, but deliver three-valued evaluations; and both are inadequate to handle all versions of Gupta's Challenge. A version of Kripke's theory with a different form of supervaluation can solve the first two problems and make progress on the third.

4 The alternative

To perform supervaluation in a correspondence-theoretically acceptable way, we need to consider all possible evaluations, i.e. all possible extensions of the evaluation at some given level. Here is a way that this can be done in the case of Gupta's sentences. Consider the tree depicted in figure 1. Here B4 is at the top (the "root") and at the nodes below it are the sentences on which the truth value of B4 depends. (I am here relying on an intuitive notion of dependency. It will be replaced by a precise definition in the next section.) Below A3 and A5 are, in turn, the sentences they depend upon. The different possible evaluations correspond to the different ways truth values can be assigned to the nodes, in such a way that nodes with other nodes below them are assigned truth values in accordance with the

have stabilized. He further believes that we should quantify over all such rules, and only consider a sentence to be true or false if it stabilizes as such under all rules. On this proposal, the A3/A5* version will also not arrive at the intuitively correct values, as one of the limit rules quantified over is the one starting with universal falsity and always reverting to falsity. Technically it is possible to deal with the Challenge within the framework of revision theory, namely by using "fully varied revision sequences"; see pages 168 and 228 in (Gupta and Belnap 1993). Then for each revision sequence there would be a limit ordinal where the "right" values would be assigned and thereafter kept. However, this is to search for consistent evaluations of sentences (see footnote 1), not to make a sentence true when there is a fact it can be grounded in.

[17] A distinction can be made between a weak version of the T-schema in which the biconditional is in the meta-language, namely "$T(c_\phi)$ is true iff ϕ is true" where c_ϕ denotes ϕ, and a stronger version were the biconditional is in the object language, namely "$T(c_\phi) \leftrightarrow \phi$ is true". In this paper attention is restricted to the weak or "rule" version.

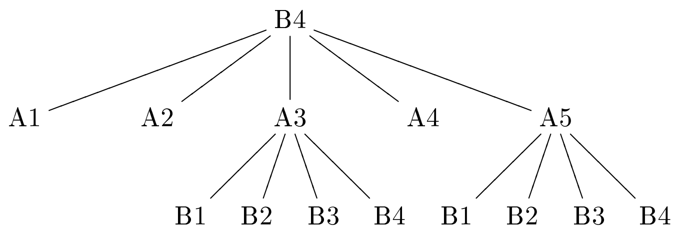

Figure 1

strong Kleene scheme based on the values of those nodes. The set of all such possible evaluations will reveal that certain combinations of truth values are impossible, *in casu* the combination of A3 being true and A5 being true. This corresponds to the crucial step in the intuitive argument for the truth values of the nine sentences: inferring that A3 and A5 contradict each other independently of what their actual truth values are.

The nodes with the sentences B1–B3 should of course be assigned the value $\top$ and the nodes with A1, A2 and A4, the value $\bot$. If we assign $\top$ to the two end nodes with B4, then under the compositional rules we should assign $\top$ to A3 and $\bot$ to A5 and therefore the root should get the value $\top$. If we instead assign $\bot$ to the end nodes with B4, the assignment to A3 and A5 should be the other way around but again the root gets a $\top$.

Lastly, we can assign the value of undefined (for which the symbol $+$ will be used) to B4 at the bottom. Recall that A5 is "Something B says is not true" and that everything else that B says is true. So when B4 is undefined and *a fortiori* not true, A5 is intuitively true. Similarly, A3 ("Everything B says is true") is intuitively false. If we assign these values to the two nodes, the root again becomes true. So in all three of the "possible realities", B4 becomes true and we can supervaluate it as such.[18]

Note that the intuitive argument for the truth values of A3, A5 and B4 does *not* rely on a hidden premise that these sentences have proper truth values. Therefore this is a reasonable thing to do.

So using trees is the basic idea. We will formulate a theory that is a modification of Kripke's. We adopt the technique of reaching a fixed point through a transfinite series of levels of increasingly more extensive partial interpretations of the truth predicate, using a jump rule to get from one level to the next, and taking unions at limit levels. Only the jump rule is new. The assignment of a truth value to a given sentence at a given level is now decided by considering trees for the sentence.

A tree for a given sentence is constructed by placing that sentence at the root, and below that placing nodes with the *constituents* of that sentence,

[18] To understand what is going on here, it is important to recall the distinction made in section 1 between truth values and facts about truth values. The former appear on the language side of the correspondence relation. The latter appear on the reality side, but are induced by truth values that are on the language side of another (more basic) instance of the relation. When B4 can be made true here it is because the supervaluation reveals that there does not have to be a fact on the reality side of the correspondence relation about the truth value of B4 for there to be the facts needed to make B4 true on the language side of the relation.

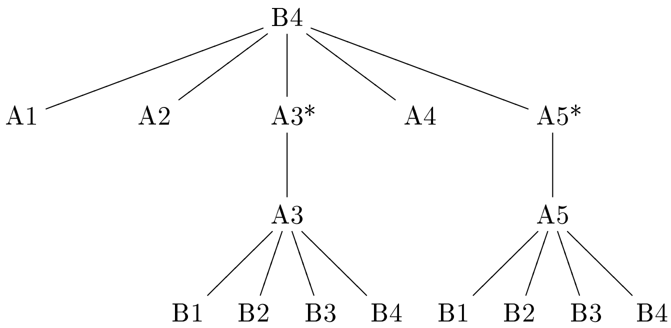

Figure 2

and then iterating. By "constituents" I mean immediate syntactical parts, in the case of connectives; all the instances, in the case of quantifiers; and the sentence referred to, in the case of a sentence claiming the truth of another sentence. The "treebuilding" is stopped when an atomic sentence with an ordinary predicate is reached, when the root sentence reappears, before any other sentence appears so as to have itself as predecessor (in order not to make it a revision theory) or at any earlier point. Also, we only consider trees for which it holds that each "route" from the root down through the tree is finite, so that truth values can be assigned to the nodes in a well-founded way and "travel" all the way from the end nodes to the root.

A tree is evaluated in the direction from the end nodes to the root. The values of some nodes are fixed: if the sentence at a given node has a truth value from an earlier level, the node has that value; if it is an atomic sentence with an ordinary predicate, its value is decided in the normal way; if it is an atomic sentence with the truth predicate and a constant not denoting a sentence, it is false. The remaining end nodes can have any of the values true, false and undefined, albeit with the restriction that if two nodes are labeled with the same sentence, they must have the same truth value. Indeed, we impose this demand not only on the end nodes, but on all nodes except for the root.[19] The remaining non-end nodes are given values based on the values of the nodes immediately below them.

Even though there are still some particulars to be filled out, it should be clear how this can help with Gupta's Challenge. Consider the tree for the A3*/A5* version in figure 2. The result is the same; the truth values merely have to "travel up" one more node. This theory is not vulnerable to any kind of iterated uses of the truth predicate in versions of Gupta's Challenge: also the A3/A5* version would work in exactly the same way.

However, there is a problem: it has not been explained exactly what way that is. In the third paragraph of this section, we accounted *intuitively* for how the assignment of + to the two bottom nodes labeled B4 should result

[19]The root is to be thought of as the language side of the correspondence relation and the other nodes as the reality side. That is why we should not extend the restriction to also involve the root. Doing so would have the consequence that a sentence would be made true simply because truth is the only proper truth value it can consistently have. And it would not be correspondence-theoretically acceptable to have that as a sufficient condition.

in the root node being assigned ⊤, but not *technically*. We will not get the desired result if we adapt Kripke's theory directly to the tree setting. If we use strong Kleene for the nodes labeled with a sentence with a connective or a quantifier with widest scope, while a sentence of the form $T(c_\phi)$, claiming truth of the sentence ϕ, gets the same value as its successor node labeled ϕ, then the assignment of $+$ to the two bottom nodes labeled B4 results in the root also getting the value $+$. In general, such a tree theory would be equivalent to the basic version of Kripke's theory. Some other change is needed as well.

In this paper, I will suggest such a change, but I will not endorse one. I happen to believe that several other, very extensive changes to Kripke's theory should be made, but this paper is not about that; it is about the idea of supervaluating on trees. Therefore, we will be working with the most minimal and uncontroversial change I can think of that is significant enough to make it possible to demonstrate the virtues of the tree approach.

This change is with the semantics of the truth predicate: for each of the three possible truth values of a sentence ϕ, what should the truth value of the sentence $T(c_\phi)$ be? The naive answer, when the problem of paradox is disregarded, is that $T(c_\phi)$ should be true when ϕ is true, false when ϕ is false, and also false when ϕ is undefined.[20] However, general use of that rule leads to paradox. If the Liar is undefined, the sentence claiming that the Liar is true would become false, and then the negation of that sentence, which is the Liar itself, would become true.

Kripke responds to this problem by not using the naive rule under any circumstances. Instead he always has $T(c_\phi)$ undefined when ϕ is undefined. Use of this rule is seemingly required to get the monotonicity which is crucial to his construction. Otherwise we run the risk that at some stage $T(c_\phi)$ is made false because ϕ is (still) undefined, and then at a later stage ϕ becomes true and hence $T(c_\phi)$ becomes true.

In some cases, the naive rule can be safely employed. Going from Kripke's truth rule to this stronger rule is a step back towards the naive theory of truth which the paradoxes have forced us to abandon; hence it requires no special justification to take that step when it can be done without any unpleasant consequences. It *does* require special justification not to use the strong rule in all circumstances when it is used in some, but I will not give one. I will accept the *ad hoc*-ness in the interest of being able to present the idea about trees in near-isolation from other possible modifications of Kripke's theory.

The intuitive result in paragraph three of this section can be achieved by using the strong rule in that evaluation of the tree. This can be done in some

[20] Do you have a *prima facie* intuition that "ϕ is not true" should be true when ϕ is undefined? Then you either share the mentioned intuition about the truth predicate, or you have an intuition that "not" should have the semantics of exclusion negation (taking truth to falsity, falsity to truth, and undefinedness to truth). The latter intuition could be substituted for (or added to) the former in the following; the results for Gupta's Challenge would be the same. To my mind Yablo (1985, p. 301) argues convincingly for the *prima facie* strength of the intuition concerning the truth predicate. Kripke (1975, p. 715) himself mentions it is an "alternative intuition", although he does not endorse it for the object language.

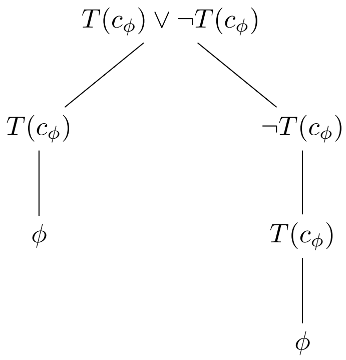

Figure 3

situations without danger to consistency or compositionality. *The situations are those where a node with $T(c_\phi)$ has the root sentence below it with the value* $+$. (This criterion needs to be generalized, but for expository reasons that is postponed to section 7.) The reason is simple: We are only using the strong rule in an evaluation which is considered together with several other evaluations in a supervaluation. In other of those evaluations the end node with ϕ is assigned $\top$, and only when those evaluations assign the same value to the root do we go ahead and assign the root sentence an actual truth value. In this way we are, so to speak, safeguarding ourselves against the possibility that a later stage in the iteration of the jump rule will contradict what we based our truth-value assignments on at the present stage. That is why this limited use of the strong truth rule does not lead to inconsistency.

The use of the rule is restricted in two ways that should be distinguished. First, it is only used in the evaluations that are quantified over. That is, if some sentence ϕ is left undefined at a given level, because the supervaluation for it did not deliver a definite result, that is not taken as a sufficient basis for making $T(c_\phi)$ false. Second, the strong rule is limited in its application to nodes in a tree that has the root sentence appearing again below it, where it is assigned $+$. The first restriction is sufficient for avoiding inconsistency.[21] The second restriction is in place to secure compositionality.[22]

To see why the second restriction is needed for that job, let us contrast the case of B4 with the example $T(c_\phi) \vee \neg T(c_\phi)$ where c_ϕ denotes some sentence ϕ which is not $T(c_\phi) \vee \neg T(c_\phi)$ itself or either of its disjuncts. A tree for this disjunction is the one shown in figure 3. Assigning values to the nodes of this tree in the same way as before, we again arrive at three possibilities. We can assign $\top$ to the end nodes, which results in the root also getting the value $\top$. Assigning $\bot$ to the end nodes has the same result. When the end nodes are given the value $+$, so must the nodes above them including the root, for otherwise we would make any such disjunction true even though there is no guarantee that either of the disjuncts will become true.

As will be demonstrated, this method of supervaluation is "reluctant" in

[21] Theorem 6 below does not depend on the second restriction.

[22] So if we were happy with true disjunctions without true disjuncts (as in Kripke's supervaluation), we could formulate a version of the theory with just the first restriction and handle Gupta's Challenge (unlike in Kripke's supervaluation) with that alone.

assigning truth and falsity, to the point that if it does, the compositional demands for that truth value are guaranteed to be satisfied. For example, if a disjunction is made true, then one of its disjuncts will be too.

In this way, the three problems mentioned at the end of the last section are solved. First, the fixed point is compositional. Second, the three-valued evaluations are reached by quantifying over three-valued evaluations, such that we are genuinely taking all possibilities into account. And third, iterations of the truth predicate in Gupta's Challenge is not a problem. With this theory we check the consequences, for the truth value of a sentence, of the assignment of different truth values to other sentences through several iterations of the truth predicate, while Kripke's theory arbitrarily restricts such consequence-checking to just one iteration.

5 The formal theory

We will formulate the theory using a first-order language with constants and ordinary predicates that are interpreted in the usual way by a domain D and an interpretation function I. In addition, it has a truth predicate T. Negation, disjunction and existential quantification are taken as primitive; and conjunction, the conditional, and universal quantification are defined in the usual way.

Given a wff ϕ, a variable v and a constant c, we understand by $\phi(v/c)$ the wff that is identical with ϕ, with the possible exception that all free occurrences of v are replaced with c. For simplicity's sake it is assumed that there is at least one constant for every object in the domain, so that quantification can be treated substitutionally.

As usual we call a wff a **sentence** if it does not contain any free variables. Let $\mathcal{S}$ be the set of sentences. Also in the interest of simplicity (avoiding the need for discussing Gödel coding and diagonal lemmas) the possibility of self-reference is facilitated, following (Gupta 1982), by assuming $\mathcal{S}$ to be a subset of D and by making certain assumptions, when needed, about constants referring to sentences.

We call an ordered pair $\mathcal{E} = (T, F)$ such that T and F are subsets of $\mathcal{S}$ an **evaluation**.[23] We say that $\mathcal{E}$ is **consistent** if T and F are disjoint. We also say that an evaluation $\mathcal{E}' = (T', F')$ **extends** $\mathcal{E}$ if $T \subseteq T'$ and $F \subseteq F'$.

A **tree**[24] is a triple $Tr = (N, <, l)$ such that N is any set; $<$ is a partial order on N such that for every element of N, the set of predecessors of this element is linearly ordered and finite and there is an element of N called the **root** that is a predecessor of every other element of N; and l is a function from N to $\mathcal{S}$. The elements of N are called **nodes**; a node without successors is called an **end node**; and for each node n, $l(n)$ is called the **label of** n. If a node n has a unique immediate successor labeled ϕ, then "n_ϕ" denotes this successor.

[23] So the letter T is used both for a set of sentences and as a predicate symbol in the language. This should not cause confusion.

[24] This definition is more restrictive than what is standard.

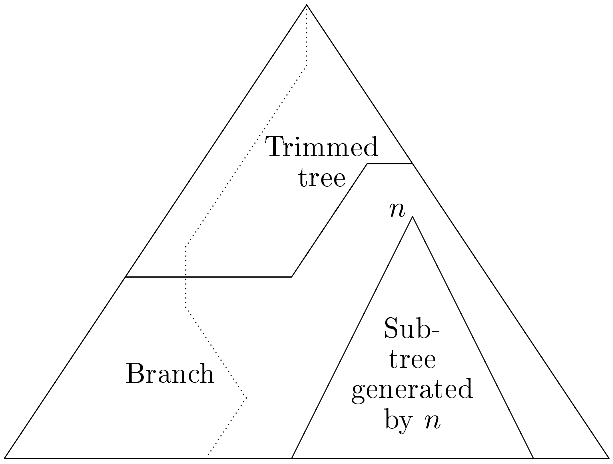

Figure 4

We will consider isomorphic trees to be identical. So, what the elements of N are is of no importance – only the cardinality of N is. (Just think of them as the dots in a graphical representation of the tree.)

Given a tree $Tr = (N, <, l)$, a **trimmed tree of** Tr is a triple $Tr' = (N', <', l')$ in which N' is a subset of N, such that

- for each node n of N, if $n \in N'$, then all the predecessors of n (by $<$) are as well, and
- for each node n of N, of the immediate successors of n, either all of them or none of them are in N',

and $<'$ and l' are the restrictions of $<$ and l, respectively, to N'. This and the next two definitions are illustrated in figure 4.

Given a tree $Tr = (N, <, l)$ and a node $n \in N$, **the sub-tree of** Tr **generated by** n is the triple $Tr' = (N', <', l')$ such that N' is the subset of N consisting of n and all its successors, and $<'$ and l' are again the restrictions of $<$ and l, respectively, to N'. Note that both trimmed trees and sub-trees generated by a node are trees.

A **branch of** Tr is a maximal linearly ordered subset of the nodes of Tr.

Given a sentence s, the **constituents of** s are

- ϕ if s is $\neg\phi$,
- ϕ and ψ if s is $\phi \vee \psi$,
- every sentence of the form $\phi(v/c)$ where c is a constant if s is $\exists v\phi$,
- the sentence $I(c)$ if s is $T(c)$ and $I(c)$ is a sentence,
- nothing if s is $T(c)$ and $I(c)$ is not a sentence, and
- nothing if s is $P(c_1, \ldots, c_n)$ where P is an ordinary predicate.

Given a sentence s, **the full tree for** s is the tree such that the root is labeled with s and for every node n, the following holds: 1) If $l(n)$ is s

and n is not the root, or one of the constituents of $l(n)$ is the label of a non-root predecessor of n, then n has no successors. 2) Otherwise n has one immediate successor for each of the constituents of $l(n)$, and these successors are labeled with these constituents. A trimmed tree of the full tree for s is called **a tree for** s if each branch thereof is finite.

Note that this means that the full tree for s is not *a* tree for s if it has infinite branches, as is the case for, e.g., the sentences of Yablo's (1993) Paradox. To be able to evaluate a tree it has to bottom out in end nodes and therefore we cannot always use the full tree. We therefore have to cut it of at some point. This introduces a complication: there is no canonical way to decide how far down to cut it of, so we need to consider a class of trees for each sentence instead of just one. The complication is small though, for if one tree is adequate to rule out enough combinations of truth values to make s true, say, then those combinations really are impossible and s should be made true. On the other hand, if a given combination is not ruled out that may just be because the tree is too small. (Consider for example the tree in figure 1 with the bottom row of B nodes removed; using that trimmed tree it would not be possible to rule out that both A3 and A5 could be true.) Therefore, it is reasonable to say that a sentence is given a truth value when just one tree provides that judgment, i.e. when all the evaluations for that tree agree.

We define an **evaluation of a tree for some sentence** s **relative to some evaluation** $\mathcal{E} = (T, F)$ (a "tree-evaluation" or, when there is no risk of misunderstanding, just "evaluation") as a function e from the nodes of the tree to $\{\top, \bot, +\}$ such that for every node n,

SEa) if n is not the root, then for every other non-root node n' such that $l(n) = l(n')$, $e(n)$ is identical to $e(n')$;

SEb) if $l(n)$ is in T (F) and n is not the root, then $e(n)$ equals $\top$ ($\bot$);

SEc) if $l(n)$ is of the form $T(c)$ where c is a constant and $I(c)$ is not a sentence, then $e(n) = \bot$; and

SE1) if $l(n)$ is of the form $P(c_1, \ldots, c_n)$ where P is an ordinary n-ary predicate and $c_1, \ldots, c_n$ are constants, then

- $e(n) = \top$ if $(I(c_1), \ldots, I(c_n)) \in I(P)$, and
- $e(n) = \bot$ otherwise,

and for every non-end node n, the following is true:

SE2) If $l(n)$ is of the form $\neg\phi$ where ϕ is a sentence, then

- $e(n) = \top$ if $e(n_\phi) = \bot$,
- $e(n) = \bot$ if $e(n_\phi) = \top$, and
- $e(n) = +$ otherwise.

SE3) If $l(n)$ is of the form $\phi \vee \psi$ where ϕ and ψ are sentences, then

- $e(n) = \top$ if $e(n_\phi) = \top$ or $e(n_\psi) = \top$,
- $e(n) = \bot$ if $e(n_\phi) = \bot$ and $e(n_\psi) = \bot$, and

- $e(n) = +$ otherwise.

SE4) If $l(n)$ is of the form $\exists v\phi$ where v is a variable and ϕ is a wff with at most v free, then

- $e(n) = \top$ if there exists a constant c such that $e(n_{\phi(v/c)}) = \top$,
- $e(n) = \bot$ if for every constant c it holds that $e(n_{\phi(v/c)}) = \bot$, and
- $e(n) = +$ otherwise.

SE5) If $l(n)$ is of the form $T(c)$ where c is a constant, then

- $e(n) = \top$ if $e(n_{I(c)}) = \top$,
- $e(n) = \bot$ if $e(n_{I(c)}) = \bot$,
- $e(n) = \bot$ if $e(n_{I(c)}) = +$ and there is a node labeled $l(n)$ with a successor m with $l(m) = s$ and $e(m) = +$, and
- $e(n) = +$ otherwise.

The third bullet of SE5 is the implementation of the strong truth rule. The second conjunct of the condition imposes the restriction on the use of that strong rule explained in the italicized sentence on page 14, above.

We define the **supervaluation with respect to the evaluation** $\mathcal{E} = (T, F)$ as $\mathrm{SE}_{\mathcal{E}} = (ST_{\mathcal{E}}, SF_{\mathcal{E}})$ where $ST_{\mathcal{E}}$ ($SF_{\mathcal{E}}$) is the set of those sentences s, such that for some tree for s, all evaluations of this tree relative to $\mathcal{E}$ have $\top$ ($\bot$) as the value of the root.[25] Such a tree **decides** s **with respect to** $\mathcal{E}$, and we say that $\mathcal{E}$ **makes** s **true (false)**. In contrast, saying that s **is true (false)** ***in*** $\mathcal{E}$ means that $s \in T$ ($s \in F$).

For all ordinals α, the **supervaluation at level** α, SE^{α}, is defined by recursion:

$$\mathrm{SE}^{\alpha} = \begin{cases} (\emptyset, \emptyset) & \text{if } \alpha = 0 \\ \mathrm{SE}_{\mathrm{SE}^{\alpha-1}} & \text{if } \alpha \text{ is a successor ordinal} \\ \left(\bigcup_{\eta<\alpha} ST_{\mathrm{SE}^{\eta}}, \bigcup_{\eta<\alpha} SF_{\mathrm{SE}^{\eta}}\right) & \text{if } \alpha \text{ is a limit ordinal} \neq 0 \end{cases}$$

A tree **decides a sentence at a successor level** α, if the tree decides the sentence with respect to $\mathrm{SE}^{\alpha-1}$, and the sentence is then **made true/false at level** α.

Figure 5 shows two examples of trees. The example on the left is the full tree, and also *a* tree, for the Liar. That is, c_l is a constant denoting $\neg T(c_l)$. At the first level, this tree has three evaluations. The first assigns $\top$ to the end node and to the "middle node" and $\bot$ to the root. The second is the other way around: $\bot$ to the two bottom nodes and $\top$ to the root. The third assigns $+$ to the end node. Then the third bullet of SE5 kicks in and assigns $\bot$ to the middle node. So again the root is assigned $\top$. Ergo, the evaluations do not agree on a value for the root and hence $\neg T(c_l)$ is not given a truth value in the supervaluation. As the reader can easily verify, the smaller trees for $\neg T(c_l)$ along with the trees for $T(c_l)$ also have disagreeing evaluations. Therefore, the evaluations for these trees are exactly the same at every level.

[25] Note that if there is no evaluation of some tree for a sentence, the sentence becomes both true and false. It needs to be proved that this situation cannot arise. This is done in section 8.

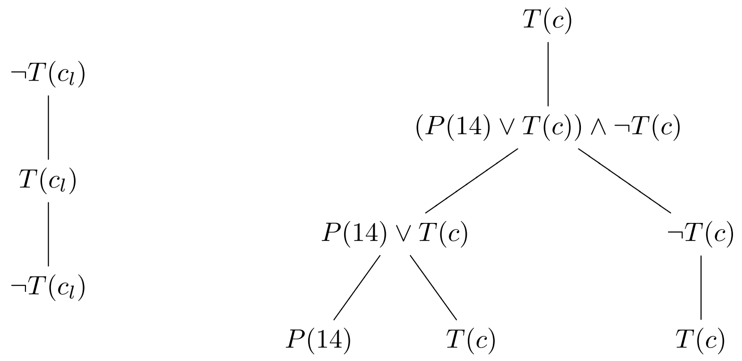

Figure 5

The example on the right is about the sentence $T(c)$ where $I(c) = (P(14) \vee T(c)) \wedge \neg T(c)$. The predicate P means "is prime" and 14 is a constant with the obvious denotation. Intuitively $P(14) \vee T(c)$ and $\neg T(c)$ contradict each other, since 14 is composite. Ergo $T(c)$ should be false. As in the Liar example, the tree depicted here is both the full tree and *a* tree for the root sentence. And at the first level this tree also has three evaluations; SEa and SE1 prevent there from being more. They all assign $\bot$ to the node labeled $P(14)$. The first further assigns $\top$ to the end nodes labeled $T(c)$ and to the node labeled $P(14) \vee T(c)$ and $\bot$ to all the remaining nodes. The second makes those end nodes $\bot$, the node labeled $\neg T(c)$ $\top$, and the rest of the nodes $\bot$. The third leaves the two end nodes with a $+$, resulting in the three nodes in the middle getting $+$ too, whereupon the third bullet of SE5 again takes effect so that the root is assigned $\bot$. In this case, all the evaluations agree on the root, so $T(c)$ is made false at level 1. As a consequence, at all higher levels this tree only has one evaluation, namely the second mentioned of the three. At level 2, $\neg T(c_\phi)$ is made true by the two-node sub-tree generated by the node labeled with that sentence. Similarly, $P(14) \vee T(c)$ and $(P(14) \vee T(c)) \wedge \neg T(c)$ are made false at level 2, securing compositionality.

In order to conclude the statement of the theory we need to prove monotonicity, the existence of a fixed point, and consistency, but this is postponed to section 8. We refer to the fixed point as $\mathcal{SE}$. For all sentences s we set $[\![s]\!] = \top$ if s is in the truth set of $\mathcal{SE}$, $[\![s]\!] = \bot$ if s is in the falsity set of $\mathcal{SE}$ and $[\![s]\!] = +$ otherwise. The value of $[\![s]\!]$ is of course to be thought of as *the* truth value of s.

6 Meeting Gupta's Challenge

We can now apply the theory to Gupta's Challenge. Let A and B be unary predicates interpreted as "is a sentence spoken by A" and "is a sentence spoken by B" respectively; and let $=$ be a binary predicate meaning "are

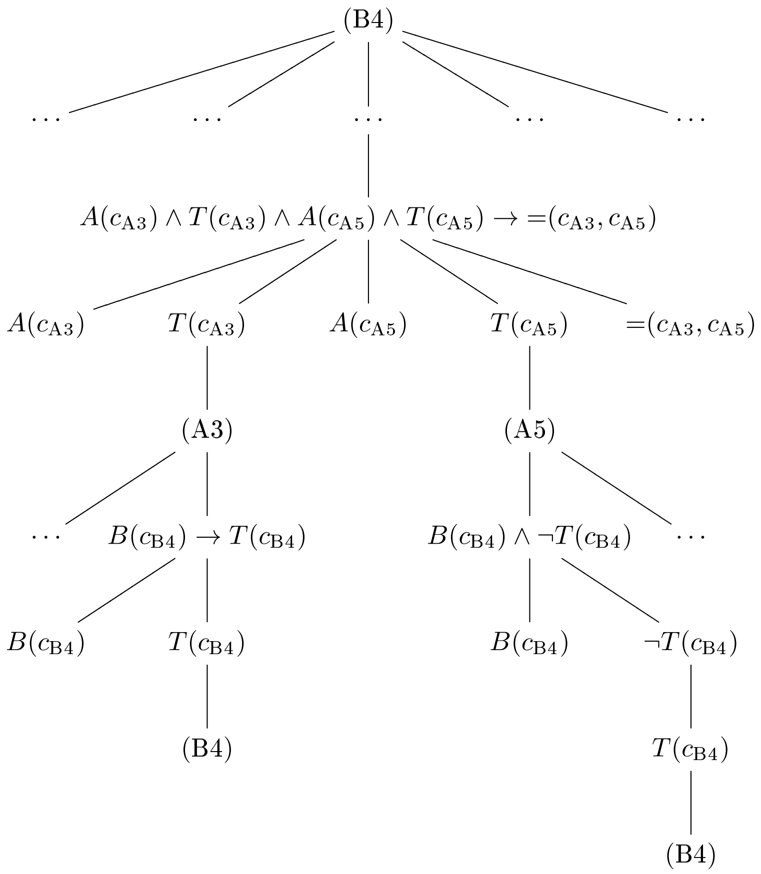

Figure 6

identical". Then the sentences A3, A5, and B4 can be formalized as follows:

$$\forall x(B(x) \rightarrow T(x)) \tag{A3}$$

$$\exists x(B(x) \wedge \neg T(x)) \tag{A5}$$

$$\forall x \forall y(A(x) \wedge T(x) \wedge A(y) \wedge T(y) \rightarrow {=}(x, y)) \tag{B4}$$

A tree for (B4) is outlined in figure 6. The constants c_{A3}, c_{A5}, and c_{B4} refer to (A3), (A5), and (B4) respectively. It is here assumed for the sake of simplicity that no other constants refer to these sentences. Another simplification is that it is pretended that conjunction and the conditional are primitive connectives and that multiple sentences can be concatenated with connectives going just one node up. I leave it to the reader to verify that these simplifications do not affect the conclusion.

It is easily verified that all instances of the doubly universally quantified sentence (B4) which are not displayed in the figure are assigned the value $\top$. Therefore there are only three essentially different evaluations of this tree.

First, there is a tree-evaluation that gives the value $\top$ to the two end nodes labeled (B4). In this evaluation, the nodes labeled $T(c_{B4})$ have the value $\top$, and the node labeled $\neg T(c_{B4})$ the value $\bot$. The nodes with $B(c_{B4})$ have the value $\top$. Ergo, the nodes labeled $B(c_{B4}) \rightarrow T(c_{B4})$ and $B(c_{B4}) \wedge \neg T(c_{B4})$ have the values $\top$ and $\bot$ respectively. Under the node labeled (A3) there is an infinity of nodes, of which all that are not shown also have the value $\top$, so this node also has the value $\top$ in this evaluation. Similarly, all the other nodes under the node labeled (A5) have the value $\bot$, so this node does as well. It follows that the two nodes labeled $T(c_{A3})$ and $T(c_{A5})$ have the values $\top$ and $\bot$ respectively. Hence, the node labeled $A(c_{A3}) \wedge T(c_{A3}) \wedge A(c_{A5}) \wedge T(c_{A5}) \rightarrow =(c_{A3}, c_{A5})$ has the value $\top$. All other nodes with sentences of the same form also have the value $\top$, and therefore the root must too.

Second, there is a tree-evaluation that assigns the value $\bot$ to the two end nodes labeled (B4). The fact that this evaluation also makes the root true follows by analogous reasoning.

The third and final tree-evaluation gives the value $+$ to the two nodes labeled (B4). Then, by the third bullet of SE5, the nodes labeled $T(c_{B4})$ have the value $\bot$, and therefore the rest of the evaluation is exactly as when $\bot$ was assigned to the two end nodes labeled (B4).

Because of SEa, there are no other tree-evaluations. From this, it follows from the definition of supervaluation that (B4) is made true at level 1. From monotonicity it then follows that $[\![(\mathrm{B4})]\!] = \top$.

Then at level 2, we only need to look at the tree for (A3) which ends at the node labeled (B4) (or to put it more precisely: the tree for (A3) which is the sub-tree generated by (A3) in the tree in the figure). Now there is only one evaluation of this tree, namely the one in which the node labeled (B4), and hence the root, is given the value $\top$. That is, (A3) is determined to be true at level 2. Similarly and simultaneously, (A5) is made false. So we have $[\![(\mathrm{A3})]\!] = \top$ and $[\![(\mathrm{A5})]\!] = \bot$.

So this theory handles Gupta's Challenge as desired. And Gupta's "revenge" in the form of replacing A3 and A5 with A3* and A5* does not bite. The proof goes through with minor modifications: Insert two more nodes with labels $T(c_{TA3})$ and $T(c_{TA5})$ in the middle of the tree as immediate predecessors of the nodes labeled $T(c_{A3})$ and $T(c_{A5})$, respectively. The constant c_{TA3} obviously refers to $T(c_{A3})$, and c_{TA5} to $T(c_{A5})$. The three evaluations at level 1 are essentially as before; the truth values just have to "travel up one more node". The combination of A3 with A5* works just as well.

7 A generalization

However, the theory, as stated, does not handle all Gupta-style challenges adequately. If we modify the story, so that person B says one more sentence, namely

B5: At least two things A says are true

intuition still insists that all ten sentences have proper truth values. The reasoning is only a slight modification of that in the introduction: A3 and A5 contradict each other, so at most one of them can be true. Hence at most one thing A says is true. So B4 is true and B5 is false. Ergo, A3 is false and A5 is true. These truth values are also, to use the language of section 1, grounded in facts about which combinations of truth values are possible.

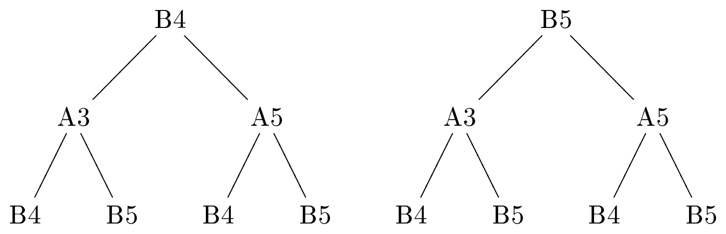

Figure 7

Figure 7 shows, in simplified form, trees for B4 and B5. The following tables detail the nine evaluations of each of the trees.

B4 (root)	⊤	+	⊤	⊤	⊤	⊤	⊤	⊤	⊤
A3	⊤	+	⊥	⊥	⊥	⊥	⊥	⊥	⊥
A5	⊥	+	⊤	⊤	⊤	⊤	⊤	⊤	⊤
B4	⊤	⊤	⊤	+	+	+	⊥	⊥	⊥
B5	⊤	+	⊥	⊤	+	⊥	⊤	+	⊥

B5 (root)	⊥	⊥	⊥	+	⊥	⊥	⊥	⊥	⊥
A3	⊤	⊥	⊥	+	⊥	⊥	⊥	⊥	⊥
A5	⊥	⊤	⊤	+	⊤	⊤	⊤	⊤	⊤
B4	⊤	⊤	⊤	+	+	+	⊥	⊥	⊥
B5	⊤	+	⊥	⊤	+	⊥	⊤	+	⊥

The top row of the first table only contains the values true and undefined and so "almost" makes B4 true. Not quite, though, for in one tree-evaluation the root is undefined. Similarly, B5 is not supervaluated as false.

Making B4 true and B5 false can be done without taking the risk of applying the strong rule for the truth predicate to an undefined sentence that at a later stage becomes true. We just need to consider the two sentences "simultaneously", a complication that the present rules do not take into account.

We can remedy the situation by amending the theory as follows: For each set of sentences S, we consider all sets of trees containing exactly one tree Tr_s for each $s \in S$, such that elements of S are only labels of root nodes and end nodes. Then we replace "$l(m) = s$" in the third bullet of SE5 with "$l(m) \in S$". If it holds for each Tr_s that all evaluations of it assign the same value to the root, then all the sentences in S are given these root values. With this change, which is a straightforward generalization of the original theory in which S was only allowed to be a singleton, B4 is made true and B5 false, in accordance with the intuitive verdict. A3 and A5 are similarly given the right truth values (now already at level 1).

To be more precise, we change four of the definitions. First, the definition of full tree is altered to read as follows: given a sentence s *and a set of sentences S containing s*, **the full tree for** ***s relative to S*** is the tree such that the root is labeled with s and for every node n, the following holds: 1)

If $l(n)$ is *an element of* S and n is not the root, or one of the constituents of $l(n)$ is the label of a non-root predecessor of n, then n has no successors. 2) Otherwise, n has one immediate successor for each of the constituents of $l(n)$, and these successors are labeled with these constituents. For some examples, consider figure 5 again. The tree on the left is the full tree for $\neg T(c_l)$ relative to S for any S that does not include $T(c_l)$. If S does include $T(c_l)$, the full tree stops one node higher up. Similarly, the tree on the right is the full tree for $T(c)$ relative to any S that does not include any of the three sentences $(P(14) \vee T(c)) \wedge \neg T(c)$, $P(14) \vee T(c)$ and $\neg T(c)$. If it does include, say, $P(14) \vee T(c)$, remove the two left-most end nodes to get the full tree for $T(c)$ relative to S.

Second, the definition of a tree is relativized to S in the obvious way. Third, the definition of tree-evaluation is relativized to S by changing "$l(m) = s$" to "$l(m) \in S$". And fourth, **supervaluation with respect to the evaluation** $\mathcal{E} = (T, F)$ is redefined as $\mathrm{SE}_{\mathcal{E}} = (ST_{\mathcal{E}}, SF_{\mathcal{E}})$ where $ST_{\mathcal{E}}$ $(SF_{\mathcal{E}})$ is the set of those sentences s, such that for *some set of sentences* S *containing* s *and* some tree for s *relative to* S, all evaluations of this tree relative to $\mathcal{E}$ *and* S has $\top$ $(\bot)$ as the value of the root*; and further, that for all the other elements of* S *there are also trees for them relative to* S*, all the evaluations of which agree on assigning* $\top$ *or agree on assigning* $\bot$ *to the root.* Such a tree **decides** s **with respect to** $\mathcal{E}$, and we say that $\mathcal{E}$ **makes** s **true (false)** ***(among*** S***)***.

8 Theorems and proofs

For this generalized theory we can now prove the promised theorems.

Lemma 1. *For any evaluation* $\mathcal{E}$*, extension* $\mathcal{E}'$ *of* $\mathcal{E}$*, sentence* s*, set of sentences* S *containing* s*, tree* Tr *for* s *relative to* S*, and evaluation* e *of* Tr *relative to* S *and* $\mathcal{E}'$*, the tree-evaluation* e *is also an evaluation of* Tr *relative to* S *and* $\mathcal{E}$*.*

Proof. If one of the antecedents of SEb is satisfied for $\mathcal{E}$, then the same antecedent is satisfied for $\mathcal{E}'$. So a restriction on what counts as an evaluation of the tree imposed by this clause in the case of $\mathcal{E}$ also applies in the case of $\mathcal{E}'$. The same holds trivially for the other clauses. □

Lemma 2. *For any evaluation* $\mathcal{E}$ *and extension* $\mathcal{E}'$ *of* $\mathcal{E}$*,* $\mathrm{SE}_{\mathcal{E}'}$ *is an extension of* $\mathrm{SE}_{\mathcal{E}}$*.*

Proof. For any sentence s, set of sentences S containing s, and tree for that sentence relative to S, the set of evaluations of the tree relative to S and $\mathcal{E}'$ is a subset of the set of evaluations of the tree relative to $\mathcal{E}$. This follows from lemma 1. So s satisfies the criterion for being in $ST_{\mathcal{E}'}$ $(SF_{\mathcal{E}'})$ if it satisfies the criterion for being in $ST_{\mathcal{E}}$ $(SF_{\mathcal{E}})$. □

Theorem 3 (Monotonicity)**.** *For all ordinals* α *and* β*, if* $\alpha < \beta$ *then* SE^{β} *is an extension of* SE^{α}*.*

Proof. As SE^0 is empty, this follows from lemma 2 by induction. □

Theorem 4 (Fixed point)**.** *There is an ordinal α such that for all ordinals $\beta > \alpha$, $\mathrm{SE}^\beta = \mathrm{SE}^\alpha$.*

Proof. This follows from Theorem 3 by the usual cardinality argument. $\square$

As prematurely mentioned, we refer to the fixed point as $\mathcal{SE}$.

The basic idea for the consistency proof is to show that the largest intrinsic fixed point of the strong Kleene jump is an extension of $\mathcal{SE}$. For this, we need some definitions.[26] "Fixed point of the strong Kleene jump" will be shortened to **FPSK** (and following Kripke, we will take these to include only consistent evaluations). An evaluation $\mathcal{E}$ is an **intrinsic FPSK** if $\mathcal{E}$ is an FPSK and it is the case that for any FPSK $\mathcal{E}'$ there exists an FPSK $\mathcal{E}''$ that is an extension of both $\mathcal{E}$ and $\mathcal{E}'$. In that case, the elements of the truth (falsity) set of $\mathcal{E}$ are called **intrinsically true (false)**. Of the intrinsic FPSKs, one is the largest (Kripke 1975, p. 709). We denote it $\mathcal{I}$.

For any set of sentences S, function $\pi : S \to \{\top, \bot\}$ and FPSK $\mathcal{E} = (T, F)$, let $\mathcal{E}^\pi$ be $(T \cup \pi^{-1}(\top), F \cup \pi^{-1}(\bot))$. The **closure of** $(\mathcal{E}, \pi)$, denoted $cl(\mathcal{E}, \pi)$, is defined as the smallest extension (CT, CF) of $\mathcal{E}^\pi$ such that if $\phi \in CT$ or $\psi \in CT$, then $\phi \vee \psi \in CT$; and if $\phi \in CF$ and $\psi \in CF$ then $\phi \vee \psi \in CF$ and similarly for negation, the existential quantifier and the truth predicate (we refer to these as **closure rules**). The **semi-closure of** $(\mathcal{E}, \pi)$, denoted $cl^-(\mathcal{E}, \pi)$, is defined as the smallest extension (CT, CF) of $\mathcal{E}^\pi$ such that if $\phi \in CT$ or $\psi \in CT$ *and* $\phi \vee \psi \notin S$, then $\phi \vee \psi \in CT$; and if $\phi \in CF$ and $\psi \in CF$ *and* $\phi \vee \psi \notin S$, then $\phi \vee \psi \in CF$ and similarly for negation, etc.

There are two things about these definitions that should be noted. First, closure could just as well have been defined as a function on $\mathcal{E}^\pi$, for it does not matter which truths and falsities "comes from" $\mathcal{E}$ and which from π. The only reason for not defining it like that is the desire to have the wording of the definition as close as possible to that of semi-closure. For semi-closure it *does* matter what comes from $\mathcal{E}$ and what from π. Second, $cl(\mathcal{E}, \pi)$ may not be an FPSK. As no restrictions have been placed on π, it may, for example, take a disjunction to $\top$, so that that disjunction is also true in $cl(\mathcal{E}, \pi)$, without either of the disjuncts being true in $cl(\mathcal{E}, \pi)$. Another reason that $cl(\mathcal{E}, \pi)$ may not be an FPSK is that it can be inconsistent. Taking the closure of $(\mathcal{E}, \pi)$ is like doing the Kripke iteration starting from $\mathcal{E}^\pi$ except that monotonicity is forced. The following lemma gives a condition under which $cl(\mathcal{E}, \pi)$ *is* an FPSK:

Lemma 5. *Let S be a set of sentences, π a function $S \to \{\top, \bot\}$ and $\mathcal{E}$ an FPSK such that $cl(\mathcal{E}, \pi)$ is consistent. For each $s \in S$, let $Tr_s = (N_s, <_s, l_s)$ be a tree for s relative to S with more than one node. Let e_s be the function from N_s to $\{\top, \bot, +\}$ defined by having, for each $n \in N_s$, $e(n) = \top$ ($\bot$; $+$) if $l(n)$ is true (false; undefined) in $cl(\mathcal{E}, \pi)$. If for each s, e_s is an evaluation of Tr_s relative to S and $(\emptyset, \emptyset)$, then $cl(\mathcal{E}, \pi)$ is an FPSK.*

Proof. Let $\mathcal{E}^*$ be the result of applying the strong Kleene jump to $cl(\mathcal{E}, \pi)$. We need to show that $cl(\mathcal{E}, \pi)$ is an extension of $\mathcal{E}^*$ and that $\mathcal{E}^*$ is an extension of $cl(\mathcal{E}, \pi)$. The former follows directly from the definition of cl. To

[26] The strong Kleene jump is defined on pages 700–703 of (Kripke 1975) where it is denoted "ϕ". The notion of a fixed point is defined on page 703.

demonstrate the latter it is enough to show that for all $s \in S$, s is false in $\mathcal{E}^*$ if $\pi(s) = \bot$ and true in $\mathcal{E}^*$ if $\pi(s) = \top$. So let an $s \in S$ be given.

As e_s is an evaluation of Tr_s and assigns $\pi(s)$ to the root labeled s, e_s and thereby $cl(\mathcal{E}, \pi)$ assign values to the immediate successors of the root/the constituents of s in such a way that (if s is an atomic sentence with the truth predicate) the T-scheme is satisfied (because the assignment of $\pi(s')$ to end nodes labeled with a $s' \in S$ implies that the third bullet of SE5 is not applied) or (if s is any other type of sentence) the strong Kleene scheme is satisfied.

Each of those constituents ϕ satisfies the T-schema/the strong Kleene scheme in relation to *its* constituents: if ϕ is an element of S this follows from similar considerations about the root and its immediate successors in Tr_ϕ, and if not it follows from the definition of $cl(\mathcal{E}, \pi)$.

For each constituent, continue like this until either 1) reaching sentences that do not have constituents or 2) passing from atomic sentences with the truth predicate to their constituents. Having the latter (with their associated truth values) in $cl(\mathcal{E}, \pi)$ is sufficient for having s false in $\mathcal{E}^*$ if $\pi(s) = \bot$ and having s true in $\mathcal{E}^*$ if $\pi(s) = \top$. □

Theorem 6. *$\mathcal{SE}$ is consistent.*

Proof. We prove by induction that SE^α is consistent and that $\mathcal{I}$ is an extension of SE^α. The base and limit cases are trivial. So for the successor case, let an α be given and assume that $\mathcal{I}$ is an extension of SE^α. Let s be a sentence.

Claim 6.1. *s is not made both true and false by* SE^α *among the same S.*

Proof. To prove the claim it must be demonstrated that a) every tree for s relative to S has an evaluation relative to SE^α and b) there is not one tree for s relative to S all the evaluations for which assign $\top$ to the root, and another tree for s relative to S all the evaluations for which assign $\bot$ to the root.

Let Tr_s be a tree for s relative to S. We construct an evaluation for Tr_s as follows:[27]

1. For each node n of Tr_s, if $l(n)$ is true (false) in $\mathcal{I}$, then assign the value $\top$ ($\bot$) to n.
2. If there are end nodes labeled with an element of S that were not assigned a value in step 1, assign them values: $\top$ if the label is made true by SE^α, $\bot$ otherwise.

[27] The most straightforward approach to constructing such an evaluation will not work. If we simply assign values to the end nodes according to the rule "$\top$ if the label is in the truth set of SE^α/$\bot$ if the label is in the falsity set/$+$ otherwise", and then assign values to the non-end nodes according to SE2-SE5, the result is not necessarily a tree-evaluation relative to the given evaluation. For example, let ϕ and ψ be labels of two end nodes and assume they are in neither the truth set nor the falsity set. Just above them may be a node labeled $\phi \vee \psi$ which *is* in the truth set. According to SEb, the node should be assigned $\top$, while SE3 dictates that it should instead be given the value $+$. (As is proved below, either ϕ or ψ would eventually be made true, but compositionality is something that holds in the fixed point, not at every level leading up to it.)

3. Starting from those end nodes labeled with an element of S and going through their predecessors from the bottom up, ensure that the first two bullets of each of SE2–SE5 are satisfied.
4. For each non-end node that is assigned a value in step 3, if there are other non-end nodes with the same label, copy the assigned value to them. Then repeat steps 3 and 4 starting from those nodes.
5. Assign $+$ to all remaining nodes.

(To prevent confusion: In fact no node is assigned $\top$ in step 2. But that will only have been established at the end of this proof. It does not follow *directly* from the induction hypothesis, for being made true *by* SE^α is not the same as being true *in* SE^α.)

To see that this is an evaluation we first need to verify that every node is assigned exactly one of the three values. As step 5 assigns $+$ to all nodes that have not already received a truth value, and $+$ is assigned in no other step, this reduces to showing that no node is assigned both $\top$ and $\bot$ in the first four steps. The consistency of $\mathcal{I}$ implies that this does not happen in step 1, and step 2 expressly avoids nodes that have been given a value in step 1 and at most assigns one value to other nodes. For steps 3 and 4 I will just give the intuitive idea for what should formally be an induction argument concerning a transfinite, monotonic sequence of partial tree-evaluations where values are added to nodes one at a time (in steps 3 and 4 from the bottom up), and I will stick to the example of a node n that is labeled with a disjunction $\phi \vee \psi$ and assigned $\bot$ at some point in the "process", leaving the other cases to the reader. We need to show that this node was not already assigned $\top$ previously in the process, and we can assume that at this point no other node is assigned both $\top$ and $\bot$.

If n is assigned $\bot$ in step 3, it has successors labeled ϕ and ψ that are both assigned $\bot$ and therefore not $\top$. Hence, those two sentences are not true in $\mathcal{I}$, so neither is $\phi \vee \psi$, ergo n was not assigned $\top$ in step 1. It also follows directly from those two nodes not having been assigned $\top$ that n has not been assigned $\top$ in step 3. Since n is a non-end node, it was not assigned any value in step 2. Finally, it can not have been assigned $\top$ in step 4 (even though *a* step 4 can precede *a* step 3), for that would imply that the node from which it was copied had successors labeled ϕ and ψ at least one of which would have been assigned $\top$. That is impossible, for had that assignment happened in step 1 or 2, it would also have happened to the successor of n with the same label (in the latter case because they would then both be end nodes); and had it happened in step 3, it would have been copied in step 4.

If instead n is assigned $\bot$ in step 4, then some other node m labeled $\phi \vee \psi$ was assigned $\bot$ in step 3 and then the same reasoning can be used to show that n was not assigned $\top$ in steps 1, 3 or 4. That it was also not in step 2 is due to the fact that it follows from m being assigned a value in step 3 that m is a non-end node and that therefore $\phi \vee \psi$ is not in S.

We can now proceed to show that what is constructed is an evaluation of Tr_s relative to S and SE^α by verifying that all the clauses are satisfied. The induction hypothesis implies that SEb is satisfied after step 1. SEc and SE1 are as well. SEa is satisfied for the following reason: it is obvious that it is after step 1; in step 2 values are only assigned to nodes with a label that only appear on other end nodes, save perhaps the root, so the same values are assigned to all other non-root nodes with the same labels in that step; and step 4 is in place to ensure that violations of SEa in step 3 are taken care of. Bullets one and two of SE2–SE5 are satisfied after step 1, may not be after step 2, but then are again after step 3. Let m be a node that has a value "copied" to it in step 4, and assume that m has immediate successors (there is nothing to show if not). The node from which the value was copied has immediate successors with the same labels, and there the first two bullets of SE2–SE5 were satisfied, so they are also satisfied for m and its immediate successors. Step 2 ensures that the third bullet of SE5 is vacuously satisfied. The last bullet of each of SE2–SE5 is satisfied by step 5.

With that, part a) is dealt with. For part b), let Tr'_s and Tr''_s be two trees for s relative to S. Let Tr_s be their "union" (in the obvious but not literal meaning of this word). Tr_s is also a tree for s relative to S, as the properties of having only finite branches and only having the root and end nodes labeled with elements of S are preserved. It is easily seen that, excluding evaluations which assign $+$ to end-nodes labeled with elements of S, any evaluation of Tr_s can be restricted to an evaluation of Tr'_s and an evaluation of Tr''_s. As it has been shown that Tr_s has such an evaluation, Tr'_s and Tr''_s have evaluations that assign the same value to their roots. ⊣

Now assume that s is made true among S (the case of falsity is similar).

Claim 6.2. *s is intrinsically true.*

Proof. Let π be the function that describes the truth values that the elements of S are given. Obviously, s is not intrinsically false, for then the constructed tree-evaluation makes the root false, and therefore s would not have been made true. By generalization it follows that $\mathcal{I}^\pi$ is consistent. As a closure rule can only produce a truth value for a sentence that contradicts an intrinsic truth or falsity from other truth values that contradict intrinsic truths or falsities, it further follows that $cl^-(\mathcal{I}, \pi)$ is also consistent.

Under the assumption that s is made true *among* S, what happened in step 2 of the construction was that each end node with a label s' from S was given exactly the value $\pi(s')$. Ergo, for all non-root nodes, the value assigned is the same as the one the label has according to $cl^-(\mathcal{I}, \pi)$. From the assumption that s is made true, it follows that the root is assigned $\top$; and hence it follows from the fact that the two first bullets of SE2–SE5 are satisfied that s cannot be made false by one more application of a closure rule to $cl^-(\mathcal{I}, \pi)$. Generalizing this observation to all the elements of S, it follows that $cl(\mathcal{I}, \pi) = cl^-(\mathcal{I}, \pi)$, i.e. that $cl(\mathcal{I}, \pi)$ is consistent. So according to lemma 5, $cl(\mathcal{I}, \pi)$ is an FPSK, since the constructed evaluation of Tr_s is exactly the function e_s mentioned in that lemma.

Assume for *reductio* that s is false in some FPSK. By the definition of $\mathcal{I}$ it follows that there is an FPSK, $\mathcal{I}^*$, which is an extension of $\mathcal{I}$ and in which s is false. Define a new evaluation of Tr_s the same way as above but using $\mathcal{I}^*$ instead of $\mathcal{I}$ in step 1. This is an evaluation of Tr_s relative to $\mathcal{I}^*$ and hence according to lemma 1 and the induction hypothesis also relative to SE^{α}, and it assigns $\perp$ to the root. This is a contradiction.

It can be concluded that s is true in an FPSK and not false in any. Hence, no sentence that is true or false in $\mathcal{I}^{\pi}$ has the opposite truth value in an FPSK. It follows that no sentence that is true or false in $cl(\mathcal{I}, \pi)$ has the opposite truth value in an FPSK. Ergo $cl(\mathcal{I}, \pi) = \mathcal{I}$, i.e. s is intrinsically true. $\dashv$

As s was arbitrary, $\mathcal{I}$ is an extension of $\mathrm{SE}^{\alpha+1}$. From this it finally follows that s is not made both true and false in $\mathrm{SE}^{\alpha+1}$, period. □

Theorem 7 ($\mathcal{SE}$ is compositional and satisfies the T-schema)**.** *For all sentences ϕ and ψ, wff's ξ with at most the variable v free, and constants c such that $I(c) \in \mathcal{S}$, the following holds:*

- $[\![\neg\phi]\!] = \top$ *iff* $[\![\phi]\!] = \perp$.
- $[\![\neg\phi]\!] = \perp$ *iff* $[\![\phi]\!] = \top$.
- $[\![\phi \vee \psi]\!] = \top$ *iff* $[\![\phi]\!] = \top$ *or* $[\![\psi]\!] = \top$.
- $[\![\phi \vee \psi]\!] = \perp$ *iff* $[\![\phi]\!] = \perp$ *and* $[\![\psi]\!] = \perp$.
- $[\![\exists v\xi]\!] = \top$ *iff for some constant k,* $[\![\xi(v/k)]\!] = \top$.
- $[\![\exists v\xi]\!] = \perp$ *iff for all constants k,* $[\![\xi(v/k)]\!] = \perp$.
- $[\![T(c)]\!] = \top$ *iff* $[\![I(c)]\!] = \top$.
- $[\![T(c)]\!] = \perp$ *iff* $[\![I(c)]\!] = \perp$.

Proof. The right-to-left direction is, in each case, simple: Consider a level where the right-hand side is satisfied. Then at the next level, the tree for the sentence on the left-hand side consisting of just the root and immediate successors thereof will do the job.

The left-to-right direction can be proved by induction on the smallest level that decides the sentence on the left-hand side. Such a level is always indexed by a successor ordinal, so the base and limit cases are trivial. Let α be an ordinal, and s a sentence that is decided at level $\alpha + 1$ and not at any lower level. Let S be a set of sentences that s was decided among, and let Tr_s be a tree that decided s relative to S at level $\alpha + 1$. Again we construct a tree-evaluation, this time by taking these steps:

1. For each node n of Tr_s, if $l(n)$ is true (false) in SE^{α}, then assign the value $\top$ ($\perp$) to n.
2. Starting from each node assigned a value in step 1 and working downwards, ensure that the first two bullets of each of SE2–SE5 are satisfied, by assigning values to immediate successors that their labels have in $\mathcal{SE}$. This is possible according to the induction hypothesis.
3. For each node that is assigned a value in step 2, if there are other nodes with the same label, copy the assigned value to them. Then repeat steps 2 and 3 starting from those nodes.
4. Assign the value of each element of S according to $\mathcal{SE}$ to end nodes labeled with those sentences.

5. Starting from each node assigned a value in steps 1, 3 and 4 and working upwards, ensure that the first two bullets of SE2–SE5 are satisfied.
6. For each node that is assigned a value in step 5, if there are other nodes with the same label, copy the assigned value to them. Then repeat steps 5 and 6 starting from those nodes.
7. Assign + to all remaining nodes.

This is an evaluation of Tr_s relative to SE^α: SEb, SEc and SE1 are satisfied after step 1. That SEa and SE2–SE5 are satisfied is demonstrated in the same way as in the previous proof.

If follows that this evaluation assigns the same value to the root as s has in $\mathcal{SE}$. From this it again follows that, if this value is $\top$ ($\bot$) and s is of the form $\neg\phi$, then the immediate successor of the root (labeled ϕ) is assigned $\bot$ ($\top$), and similarly for the other bullets of the theorem; in the last case because step 4 ensures that bullet 3 of SE5 does not come into play. Whenever $\top$ ($\bot$) is assigned to a node, the label of that node is true (false) in $\mathcal{SE}$. This follows from monotonicity in the case of step 1, from the induction hypothesis in the case of step 2, from the right-to-left direction of this proof in the case of step 5, and is trivial for the other steps. From this the desired conclusion can be deduced. □

9 Comparison and discussion

Gupta has one more trick up his sleeve; a trick that cannot be dealt with using the approach of this paper alone. Replace A3 and A5 with these sentences:

A3†: "A3† is true" is true
A5†: "A3† is not true" is true

Under the present theory these sentences become undefined, and as a consequence B4 does as well. But the same old reason for B4 being intuitively true seems to have undiminished power: A3† and A5† are intuitively contradictory, so the fact that they are not both true is independent of their specific truth values.

What is required to make the formal theory match the intuition is that the sentence

$$A(c_{TTA3^\dagger}) \wedge T(c_{TTA3^\dagger}) \wedge A(c_{T\neg TA3^\dagger}) \wedge T(c_{T\neg TA3^\dagger}) \rightarrow =(c_{TTA3^\dagger}, c_{T\neg TA3^\dagger})$$

(with obvious notation) is made true even though two of the conjuncts of the antecedent are undefined, the other two true, and the consequent false. We would need a theory for which theorem 7 does not hold, and a point of this paper has been to investigate how far we can go with supervaluation while retaining exactly that.

I will return to this version of Gupta's Challenge below and first undertake a comparison between the theory presented in this paper and the various versions of Kripke's theory.

The basic version can be reformulated using trees as follows: For each sentence, consider only the tree for that sentence that stops one node below any node labeled with a sentence of the form $T(c)$ for some constant c. Further, remove the third bullet from SE5, and let SEa and SEb apply only to end nodes. Then check only one evaluation of this tree, namely the one that assigns + to all end nodes to which none of SEb, SEc or SE1 applies. If the root is assigned a proper truth value by this tree-evaluation, then (and only then) is the sentence in question given that value.[28]

The weak Kleene scheme version can of course be obtained similarly if proper modifications are made to clauses SE3 and SE4.

The simple Kripkean supervaluation version can be reconstructed using the same trees and by restricting attention to evaluations that assign proper truth values to all nodes. The more sophisticated supervaluation versions then correspond to considering even fewer different tree-evaluations.[29]

So all these versions of Kripke's theory can be reformulated using trees, and hence, in a sense, the framework of this paper is a generalization of Kripke's. In fact, they can be reformulated using just a certain class of rather small trees. With the perspective afforded by the more general framework, this can be seen to be an arbitrary and unreasonable restriction. Kripkean supervaluation only looks forward through one iteration of the truth predicate, so to speak. The starred versions of Gupta's Challenge show that a supervaluation method should look forward through an arbitrary number of iterations.

However, there is another sense in which the present theory is not a generalization compared to Kripke's work. He does consider the set of all fixed points for the jump operation in abstraction from methods of "arriving" at these fixed points. And the theory considered in this paper is just one of these fixed points for the strong Kleene jump. (This follows from theorem 7, together with the fact that atomic sentences with ordinary predicates are assigned truth values the "right" way.) So in that sense, the merit of the tree framework is merely that it provides a natural method for reaching this fixed point "from below". But that is philosophically important. Knowing that an evaluation is a fixed point tells you that there is an answer of sorts to any question about why a given sentence is true. However, it is consistent with the evaluation being a fixed point that such answers go around in a circle, which is unacceptable; the reality side of the correspondence relation must be logically prior to the language side. Thus, the existence of a method for reaching a fixed point from below is a necessary condition for it to be correspondence-theoretically acceptable.

The proof of theorem 6 reveals that the fixed point is intrinsic, i.e. no sentence has a proper truth value in it if it has the opposite truth value in some other

[28] An alternative is to remove clause 1) in the definition of "full tree" (and hence also the word "otherwise" in clause 2); equip the trees (still only having finite branches) for a given sentence with only the one evaluation just described; and count a sentence as true (false) iff some tree for it assigns $\top$ ($\bot$) to the root. This would result in the fixed point being already reached at level 1. See (Davis 1979; Hazen 1981).

[29] Gupta's theory can be reformulated using the same trees with just one evaluation, namely the one in which the end nodes are assigned values based on the evaluation of the previous level.

fixed point. However, the fixed point is not the largest intrinsic fixed point, for the following sentence is "intrinsically true" but undefined in the fixed point: "This sentence is true or the Liar is false".[30] My contention is that this example tells against the largest intrinsic fixed point, not the theory of this paper: this sentence can consistently be true and cannot consistently be false (because then both disjuncts should be false and the second cannot be), but that should not be sufficient cause for actually counting it as true, when we require groundedness and not just non-arbitrariness. That the sentence can consistently be true and cannot consistently be false does not imply that there is a grounded fact to which the sentence can correspond. At any level where not at least one of the two sentences referred to in the disjuncts already has a truth value there is no such grounded fact. Hence, it should remain undefined.

"It is not at all clear how the largest intrinsic fixed point fits in with the intuitive picture of truth that we get from Kripke", Gupta (1982, p. 37) writes, and asks rhetorically "By what sort of stage-by-stage process do we reach this fixed point?" The largest fixed point is indeed not intuitively satisfactory. We should aim somewhat lower,[31] and when we do so, we *can* hit the target through a stage-by-stage process.

The point of this paper has been to carve out a larger class of grounded truths and falsities than Kripke managed to. I believe that all sentences that are true (false) according to the formal theory are genuinely true (false), assuming of course that the model accurately represents the non-semantic facts. That is, the formal theory is sound with respect to The Correct Theory of Grounded Truth (which is just another way of saying that it is correspondence-theoretically acceptable).

Is it possible to carve out an even larger class? I leave that question open. In section 1 I argued that all facts that are available at a given level should be utilized by the formal theory when assigning truth values. I have formulated a theory that exploits more of the available facts (while stopping short of exploiting non-facts as Kripke's and revision theory supervaluation do), but I am in no position to claim that it exploits all of the available facts. That is, I do not claim that the formal theory is *complete* with respect to The Correct Theory of Grounded Truth.

The version of Gupta's Challenge with $A3^\dagger$ and $A5^\dagger$ may be considered a counterexample to completeness. We are forced to so consider it if we accept the claims made about the first version of the Challenge in the beginning of this paper *and* find ourselves unable to point to a relevant difference between the two versions to justify the unequal results. The only plausible candidate for a relevant difference is compositionality: in one case all the sentences can be assigned proper truth values in a compositional way and in the other we cannot, given that $A3^\dagger$ and $A5^\dagger$ should be undefined by virtue of being ungrounded. Even though they are ungrounded, it is still a fact that is

[30] The formalization of the sentence is $T(c_t) \vee \neg T(c_l)$ where $I(c_t) = T(c_t) \vee \neg T(c_l)$ and $I(c_l) = \neg T(c_l)$.

[31] How much lower are we aiming? I leave it as an open problem to characterize the relationship between my fixed point and the largest intrinsic fixed point with more precision.

available at the first level that they contradict each other, so the necessary facts in which to ground the truth of B4 seem to be there. Does the demand for compositionality trump the force of the intuitive argument that is the same for all the versions of Gupta's Challenge?

If not, we have ended up in a dilemma: the desideratum to use all the information available at each level to produce new truth values conflicts with the compositionality desideratum. Let me very briefly and very superficially consider the possibility of resolving that dilemma in favor of the former desideratum. I argued in sections 2 and 3 that the lack of compositionality that consists in having true disjunctions with undefined disjuncts (together with undefined disjunctions with undefined disjuncts) is unacceptable. However, the kind of non-compositionality needed to get the same result for the $A3^\dagger/A5^\dagger$ version as for the other versions is different and perhaps less bad: false conjunctions with one conjunct undefined and the other true or undefined (together with undefined conjunctions with one conjunct undefined and the other true or undefined).[32] Pursuing that idea, however, involves an even more drastic modification of Kripke's theory than the one investigated in this paper and must be postponed.

However, even if the present theory is incomplete, I believe there are two ideas in this paper that would have to be used also in formulating the "ultimate" theory: 1) The idea of using all available information at each stage while not using "information" that turns out to be false when the final evaluation is reached. 2) The idea of doing supervaluation that is not limited to one iteration of the truth predicate, using the tree method. The problem of the †-version seems to be connected with the limits on the use of the strong truth rule, not directly with the central ideas of this paper.

There are also other issues I have ignored. Kripke's theory is confronted by the problem of semantic openness and the revenge problem.[33] So is the present theory. I have employed the strong rule for truth where it is safe. The challenge of providing a principled explanation of why it should not be used where it is unsafe, i.e. an explanation that does not merely appeal to the fact that it is unsafe, is left untouched.

Acknowledgments

This paper has been changed significantly because of feedback from Aaron Cotnoir, Øystein Linnebo, Crispin Wright, audiences in Munich and Rio de Janeiro, and anonymous reviewers.

[32]I think that the essential property of disjunctions is that a disjunction is true iff one of the disjuncts is, but that the essential property of conjunctions is *not* that a conjunction is false iff one of the conjuncts is, but rather that it is true iff both its conjuncts are. Truth criteria are primary, while falsity criteria (and undefinedness criteria) are derivative.

[33]See (Kripke 1975, p. 714) and (Beall 2007).

References

Austin, J. L. (1950). Truth. *Proceedings of the Aristotelian Society, Supplementary Volumes 24*, 111–128.

Beall, J. C. (Ed.) (2007). *Revenge of the Liar: New Essays on the Paradox.* Oxford University Press.

Belnap, N. (1982). Gupta's rule of revision theory of truth. *Journal of Philosophical Logic 11*, 103–116.

Davis, L. (1979). An alternative formulation of Kripke's theory of truth. *Journal of Philosophical Logic 8*, 289–296.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic 11*, 1–60.

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

Hazen, A. (1981). Davis's formulation of Kripke's theory of truth: A correction. *Journal of Philosophical Logic 10*, 309–311.

Herzberger, H. (1982). Notes on naive semantics. *Journal of Philosophical Logic 11*, 61–102.

Kirkham, R. L. (1992). *Theories of Truth: A Critical Introduction.* MIT Press.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy 72*, 690–716.

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic 34*, 155–192.

Meadows, T. (2013). Truth, dependence and supervaluation: Living with the ghost. *Journal of Philosophical Logic 42*, 221–240.

Schaffer, J. (2009). On what grounds what. In D. J. Chalmers, D. Manley, and R. Wasserman (Eds.), *Metametaphysics: New Essays on the Foundations of Ontology.* Oxford University Press.

Tarski, A. (1933). *Pojecie Prawda w Jezykach Nauk Dedukcyjnych.* Translated as "The concept of truth in formalized languages" in A. Tarski (1956), Logic, semantics, metamethematics. Oxford University Press.

Tarski, A. (1944). The semantic conception of truth. *Philosophy and Phenomenological Research 4*, 341–376.

van Fraassen, B. (1966). Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy 63*, 481–495.

Walicki, M. (2009). Reference, paradoxes and truth. *Synthese 171*, 195–226.

Wen, L. (2001). Semantic paradoxes as equations. *Mathematical Intelligencer 23*, 43–48.

Yablo, S. (1985). Truth and reflection. *Journal of Philosophical Logic 14*, 297–349.

Yablo, S. (1993). Paradox without self-reference. *Analysis 53*, 251–252.