

# Knights, Knaves, Truth, Truthfulness, Grounding, Tethering, Aboutness, and Paradox

Stephen Yablo

September, 2015

## 1 Knights and knaves

Knights, as we know, always tell the truth; knaves always lie. Knight and knave puzzles ask us to figure out who is who on the basis of their answers to cleverly contrived questions. For instance,

*A*, *B*, and *C* were standing together in a garden. A stranger passed by and asked *A*, “Are you a knight or a knave?” *A* answered, but rather indistinctly, so the stranger could not make out what he said. The stranger then asked *B*, “What did *A* say?” *B* replied, “*A* said that he is a knave.” At this point the third man, *C*, said, “Don’t believe *B*; he is lying!” The question is, what are *B* and *C*? ([Smullyan(1986)], 20)

Smullyan begins by observing that

It is impossible for either a knight or a knave to say, “I’m a knave,” because a knight wouldn’t make the false statement that he is a knave, and a knave wouldn’t make the true statement that he is a knave.

He concludes on this basis that *B*, since he is lying about what *A* said, is a knave; *C* must be a knight since he is right about *B*; *A*’s status cannot be determined.

A variant of the puzzle can be imagined in which *B* replies, not “*A* said he was a knave,” but “*A* said that he was a knight.” *B* in saying this shows himself, it seems, to be a knight. For knights and knaves *both* say, “I am a knight”—knights because “I am a knight” is true in their mouths, and knaves because it is false in theirs. *B* might equivalently have replied that *A* said he always told the truth, for that is the kind of speech behavior that is definitive of a knight.

Straightforward as this reasoning appears, there is, to go by current theories of truth and self-reference, something badly wrong with it. Knights cannot, on current theories, truly describe themselves as always telling the truth. That the problem is not apparent even to veteran paradox-mongers (see below) is a datum in need of explanation. This paper seeks mainly to *explain* the problem. But we will take a shot, toward the end, at addressing it.

## 2 Russell and Moore

The Smullyan puzzle recalls a remark of Kripke’s about semantic humility. Russell supposedly asked Moore, *Do you always tell the truth?* Moore replied that he didn’t. Russell

regarded Moore's negative reply as the sole falsehood Moore had ever produced. Surely no one had a keener nose for paradox than Russell. Yet he apparently failed to realize that if, as he thought, all Moore's other utterances were true, Moore's negative reply was not simply false but paradoxical ([Kripke(1975)], 691-2)

Why paradoxical? Assume first that the statement is false. Then Moore does sometimes lie, in which case the statement is true after all. If on the other hand it is true, then Moore never lies, in which case the answer he gives Russell is just incorrect. A statement that cannot consistently be assigned either truth-value is normally considered paradoxical. "Even the subtlest experts," Kripke says, "may not be able to avoid utterances leading to paradox."

### 3 Moore be(k)nighted

And yet, there seems to be something right about Russell's claim that Moore spoke falsely. How else are we to describe the situation, if we cannot call Moore's mea culpa a lie? All of Moore's other statements are true, we're supposing. His statement *I sometimes lie* has, therefore, no basis in fact. To call it untrue seems like our only option if we want to give voice to this observation. And yet to call it untrue is self-refuting.

Russell may have put his point in an unnecessarily paradoxical way. Perhaps he meant, not that Moore's actual statement, *I sometimes lie*, was untrue, but that the opposite statement, *I always tell the truth*, would have been true, had he made it. That *I (Moore) always speak the truth* would have been true does seem intuitively rather similar to what Russell alleges, viz. that *I (Moore) sometimes lie* is false. One feels that had Moore said instead that he never lied, or that all his statements were true, he would have spoken truly. An honest person ought to be able to assert their own honesty!<sup>1</sup> And that is what Moore would be doing in the imagined scenario.

Where does this leave us? Even if Moore did not *lie*, when he said *I sometimes lie*, Russell can be forgiven, so it seems, for thinking that he did. The judgment is forgivable for it is easily confused with (what seems so far to be) the *correct* judgment that Moore would have done better to say, *I always tell the truth*, since he would then have been speaking truly. This seems like a very satisfactory resolution. It allows us to agree with Kripke that Russell misconstrued a paradox as a lie, while also agreeing with Russell that Moore's reply to *Do you ever lie?* was an unforced error, in this sense: the answer he did give (*YES*) was indefensible, while the answer he didn't give (*NO*) would have been true. Russell had the right idea, on this interpretation; he simply didn't say it right.

### 4 The problem

To explain the false-seemingness of *I sometimes lie* as reflecting the truth of *I never lie* seems like a satisfactory resolution. But the plot now begins to thicken. Granted that *I (Moore) never lie* is not paradoxical, there is still the problem of seeing why it should be regarded as *true*. It is after all self-referential; it attributes truth to itself. Statements like that may not be consigned to the *first* circle of hell, but they *are* often sent to the second.

There's an intuitive aspect to this and a technical aspect. The intuitive aspect is as follows. You all know of the Liar sentence *L*, which describes itself as untrue ( $L = \neg T(L)$ ). The Liar cannot consistently be regarded either as true or as false; that is more or less what it means to be paradoxical. Paradox is not the only form of semantic pathology, however, as remarked by Kripke:

---

<sup>1</sup>Self-identified knights are the group Smullyan admires the most. If they were talking nonsense, he would have noticed it.

It has long been recognized that some of the intuitive trouble with Liar sentences is shared with such sentences as

(K) K is true

which, though not paradoxical, yield no determinate truth conditions ([Kripke(1975)], 693)

Where the Liar can consistently be assigned *neither* truth-value, the Truth-Teller *K* can consistently be assigned *either*. Suppose we call it true; then what it says is the case; and so it deserves the description we gave it. Likewise if we call it false. We can assign it whatever truth-value we like and that assignment will bear itself out. Borrowing a term from Kripke, the Truth-Teller is not paradoxical but *indeterminate*—much as the Liar’s truth-value is overdetermined, the Truth-Teller’s is underdetermined.

Return now to *Everything I say is true*. I will call it the Truthfulness-Teller, because the speaker (Moore, we suppose) is declaring him or herself to be generally truthful, and write it *H*, for honesty. *H* is, it may seem, in the same boat as the Truth-Teller *K*, assuming that the speaker’s other statements are true. It is equivalent after all to *Everything else I say is true, and this statement too is true*. If we postulate that Moore lies when he calls *I always tell the truth* false, the postulate is self-supporting. What the sentence says really is false, on the assumption of its falsity, because it describes itself as true. If we assume for argument’s sake that it is true, that assessment is self-supporting too.

So, the Truthfulness-Teller is true on the assumption of its truth, and false on the assumption of its falsity. A sentence that can consistently be supposed either true or false, compatibly with the non-semantic facts, is, it seems, indeterminate. The Truthfulness-Teller was *introduced*, though, precisely as a *truth* that Moore had available to him to utter, when he said instead that he was not always truthful, thus involving himself in paradox. The statement’s truth was indeed proposed as what lent the appearance of falsity to *I sometimes lie*.

That’s the intuitive aspect. The technical aspect is that if you look at the various formal truth theories that have been proposed — Tarski’s, Kripke’s, the Herzberger/Gupta theory, McGee’s theory, Field’s theory—not a single one of them supports the thought that Moore could truthfully have declared himself to be honest. Kripke’s theory doesn’t, for instance, because a sentence attributing truth to itself is *ungrounded* in the manner of the Truth-Teller and the Liar. Gupta’s theory doesn’t make *I never lie* true, for it is stably true in some revision-sequences but not others. Herzberger’s version of the revision theory makes the Truthfulness-Teller just *false*, for it assigns the truth-predicate, initially, an empty extension, a setback from which *I never lie* cannot recover.<sup>2</sup>

## 5 Kripke and dependence trees

There are really two puzzles here. One, the comparative puzzle, asks why the Truthfulness-Teller should be truer than the Truth-Teller, despite making a stronger claim. The absolute puzzle asks why the Truthfulness-Teller should be true full stop. Insofar as the first puzzle is to do with *H* seeming less grounded than *K*, and the second with *H* being ungrounded full stop, the natural context for either is Kripke’s theory, for it was Kripke who put grounding at the center of the things.<sup>3</sup>

---

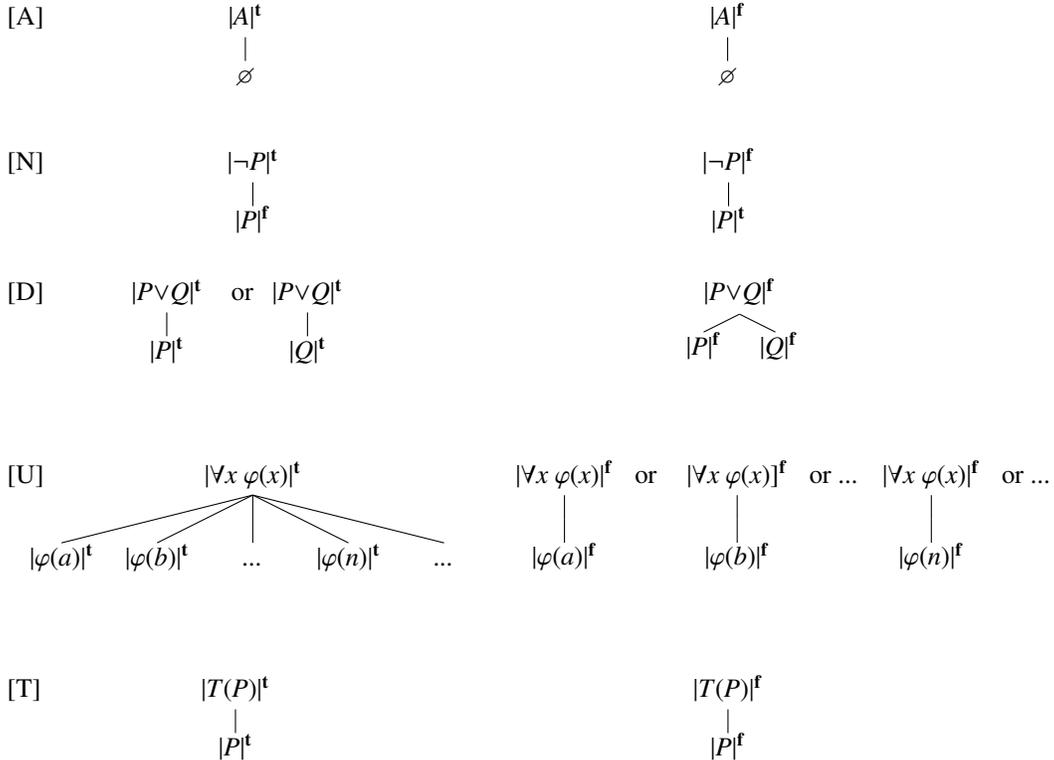
<sup>2</sup>Kripke does allow ungrounded sentences to be *intrinsically* true: true in a fixed point none of whose assignments are reversed in other fixed points. But the Truthfulness-Teller cannot claim that lesser status either, for there are fixed points in which it is uniquely false.

<sup>3</sup>Kripke cites [Herzberger(1970)]. See also [Davis(1979)], [Hazen(1981)], [Yablo(1982)], and [Yablo(1993)]. For the relation to grounding in set theory, see [Mirimanoff(1917)], [Yuting(1953)], [Boolos(1971)], [Barwise and Etchemendy(1989)], [McLarty(1993)], and [Yablo(2004)].

To appreciate how the theory works, let's associate with each sentence  $P$  two “attributions”  $|P|^t$  and  $|P|^f$ , one assigning truth to  $P$ , the other falsity. A relation  $\Delta$  on the set of attributions is a *dependence* relation iff it satisfies these conditions:

- (A) if  $A$  is atomic,  $|A|^t$  and  $|A|^f$  bear  $\Delta$  to nothing (written  $\emptyset$ )
- (N)  $|\neg P|^t$  bears  $\Delta$  to  $|P|^f$ ;  $|\neg P|^f$  bears  $\Delta$  to  $|P|^t$
- (D)  $|P \vee Q|^t$  bears  $\Delta$  either to  $|P|^t$  or  $|Q|^t$ ;  $|P \vee Q|^f$  bears  $\Delta$  both to  $|P|^f$  and  $|Q|^f$
- (U)  $|\forall x \varphi(x)|^t$  bears  $\Delta$  to  $|\varphi(n)|^t$  for each name  $n$ ;  $|\forall x \varphi(x)|^f$  bears  $\Delta$  to  $|\varphi(n)|^f$  for some particular name  $n$
- (T)  $|T(A)|^t$  bears  $\Delta$  to  $|A|^t$ ;  $|T(A)|^f$  bears  $\Delta$  to  $|A|^f$

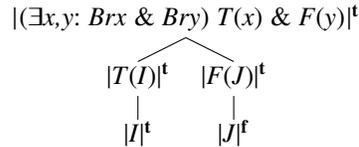
$P$  is *grounded-true* iff there is a dependence relation  $\Delta$  such that every  $\Delta$ -path starting from  $|P|^t$  leads to a fact—an atomic attribution  $|A|^t$  ( $|A|^f$ ) such that  $A$  really is true (false) in reality, as represented by the underlying model. Equivalently,  $|P|^t$  sits atop a factual  $\Delta$  tree—a dependence tree all of whose branches terminate in facts. The rules in tree form:



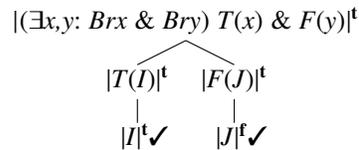
One way to define grounded-truth is in terms of trees whose branches terminate in *facts*: atomic attributions in which the sentence really does have the indicated truth-value. A different, but equivalent, way uses *decorated* trees whose attributions are marked  $\checkmark$  if they're factual and  $\times$  if they conflict with the facts. To get a decorated tree from a plain one, one starts by tagging terminal nodes with  $\checkmark$ s and  $\times$ s according to the rule just stated. One then marks parent nodes as factual when all their children have been so marked, and as

anti-factual when at least one their children is anti-factual.  $P$  is grounded-true, on this way of doing it, iff some decorated dependence tree has  $|P|^t\checkmark$  at the top.

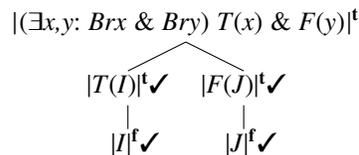
Here for instance is an undecorated tree for *Something Russell believed was true, and something he believed was false*, on the hypothesis that Russell believed (at least) that *Ice is cold (I)*, which is true, and that *Jello is hot (J)*, which is false.



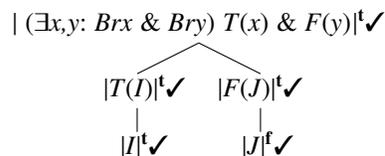
To decorate it, we start by appending  $\checkmark$  to any terminal node that is factual. As it happens they both are, so we have two  $\checkmark$ s to tack on.



That was stage 1 of the operation. Now we move gradually upward, checking off at stage  $n+1$  any nodes each of whose children were checked off at stage  $n$ . This yields, at stage 2,



and at stage 3,



A decorated tree headed by  $|\varphi|^t\checkmark$  means that  $\varphi$  is grounded-true. So, *Not everything Russell said was true, nor was it all false* is true by the lights of Kripke's grounding semantics.

Now let's try the rules out on some trickier examples, starting with the Liar  $L (= \neg T(L))$ , the the Truth-Teller, and so on.



(1) Liar

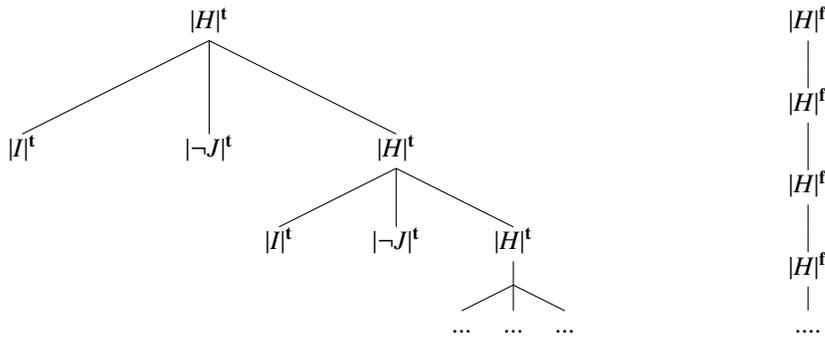
That neither tree terminates means that  $L$  is neither grounded-true nor grounded-false. Attempts to decorate either one never get off the ground since there are no terminal nodes to start from. Note that the Liar trees not only conflict with each other (that's by design) but also each with itself; each contains  $|P|^t$  and  $|P|^f$  for the same sentence  $P$ .

The Truth-Teller  $K (= T(K))$  again has two trees, each with a single infinite branch. The difference is that  $K$ 's trees are, taken individually, consistent; neither assigns truth and falsity to any sentence  $P$ . There is to that extent a consistent scenario where  $K$  is true, and another where  $K$  is false. Still, that neither tree terminates means that  $K$  is ungrounded, that is, neither grounded-true nor grounded-false.



(2) Truth-Teller

Now the truthfulness-teller  $H$ . Assume that Moore's other statements (other than  $H$ ) are  $I = \text{Ice is cold}$  and  $\neg J = \text{Jello isn't hot}$ ; then  $H = T(I) \ \& \ T(\neg J) \ \& \ T(H)$ . The trees of interest are



(3) Truthfulness-Teller

From the right-hand tree we see that  $H$  is not grounded-false. The tree for  $|H|^t$  has an infinite branch too, though, so  $H$  is not grounded-true either. Both of the trees are consistent, as with  $K$ . Officially then,  $H$  is underdetermined, just like the Truth-Teller. But that is not how it strikes us. It strikes us as true—or, at any rate, closer to true than false. There might be some support for this idea in the fact that  $|H|^t$ 's tree is “better”—more grounded in nonsemantic facts—than  $|H|^f$ 's. We'll return to this theme in a moment.

## 6 Modesty

I can think of two ways of making the Truthfulness-Teller less immodest, so as to give it a better shot at truth. We could muck with the subject term so that it covers fewer sentences. Or we could scale back the predicate, so that it attributed a weaker property.

The obvious way of cutting back the subject term is to let  $H$  attribute truth only to Moore's *other* statements. That would explain why it strikes us as true. Those other statements do indeed have the property (truth) that's attributed to them. It would also explain why the Truth-Teller seems worse off than the Truthfulness-Teller. If we cut back the subject term “this very sentence,” then nothing is left. There are no other statements that  $K$  describes *inter alia* as true.

These results are obtained, however, by changing the claim's intuitive content. *Everything I say is true...with the possible exception of this very claim*, sounds like the statement of a trickster, not a Moore. Moore does not want to exempt his declaration of honesty from the compliment it pays to his other statements. So here is our first condition on a satisfactory solution: *All my statements are true* should not make an exception of itself.

This makes the problem unsolvable. For Moore's declaration of honesty not to make an exception of itself would seem to mean that it is one of the statements that it describes as true. But then it has a Truth-Teller inside it, with the truth-destroying ungroundedness that that entails.

But there's a second thing we could try—targeting not the subject term but the predicate. Perhaps what Moore really meant is: Everything I say is true-to-the-extent-evaluable.

That again seems to distort the content. Suppose you utter a bunch of Liars and Truth-Tellers and other pathological sentences. *Your statements are true to the extent evaluable* is then true! The Truthfulness-Teller is not so easily saved. If I say, *All my statements are true*, when in fact NONE have this property, my claim may be many things, but “true” is not one of them. *All my statements are true* should attribute *truth*, not something weaker like truth-where-evaluable. And now we are back in trouble, because if  $H$  calls itself true, then it is NOT true, on account of being ungrounded; to be true, it must have been true already.

What other way of modifying the Truthfulness-Teller is there, though, if we are not allowed to make the subject term more demanding, or the predicate less so? Maybe it is not  $H$  that needs to be modified, but the claim we make on its behalf. Rather than calling it true, period, perhaps we can call it true about a certain subject matter: THE FACTS. This is roughly what we propose below:  $H$  IS FACTUAL—true TO THE FACTS—without being grounded. First though we explore another way of bringing out what *I (Moore) never lie* has that *I am not now lying* lacks.

## 7 TRUTH and grounding

For Kripke, in the first instance anyway, a sentence is true (false) only if it's grounded-true (-false). The Truthfulness-Teller seems to cast doubt on this idea. Let's remind ourselves of what it means for a sentence to be grounded-true.

0.  $P$  is true iff  $|P|^t$  sits atop a dependence tree all of whose branches terminate in facts.

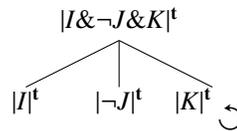
It is grounded-false iff  $|P|^f$  sits atop a dependence tree all of whose branches terminate in a fact. (Or, what comes to the same,  $\neg P$  is grounded-true.) Could the lesson of  $H$  be that grounding is too strict a condition? As a first stab at something looser, consider

1.  $P$  is TRUE (first stab) iff  $|P|^t$  sits at the top of a dependence tree

- (i) *some* of whose branches terminate in facts, and
- (ii) all of whose terminating branches terminate in facts.

(I write TRUE so as not to beg any questions about the identity of this truth-like property with the one Kripke is attempting to analyze.) A sentence is true, in other words, if  $|P|^t$  heads a decorated dependence tree constructed to slightly weaker specifications: a parent node is marked  $\checkmark$  iff (i) some of its children are marked  $\checkmark$ , and (ii) none if its (other) children are marked  $\times$ . The earlier requirement was that a parent node is validated iff *all* its children are validated.

This makes the Truthfulness-Teller TRUE, which is good, but it also makes the Truth-Teller true, or at least, treats it as true in certain constructions. Take for instance  $I \& \neg J \& K$ , where  $I$  and  $\neg J$  are plain truths, and  $K$  is again the Truth-Teller. The tree is



(4) Truth-Teller Plus

Note, the  $\curvearrowright$  notation is to indicate that a node depends on itself; the tree fully spelled out puts  $|K|^t$  on top of an infinite descending chain of  $|K|^t$ 's.  $I \& \neg J \& K$  meets the condition [1.] lays down for TRUE: some branches terminate in facts, the others don't terminate. This seems just wrong, however. How can  $I \& \neg J \& K$  be TRUE, when  $K$ , its third conjunct, lacks this property?

## 8 TRUTH and tethering

I want to pick up now on the suggestion in section 7 that some ungrounded attributions are closer to being grounded than others. A glance at their trees makes clear that  $|H|^t$  is less ungrounded than  $|I \& \neg J \& K|^t$ , which is less ungrounded than  $|K|^t$ , and also less ungrounded than  $|H|^f$ . In what sense, though?

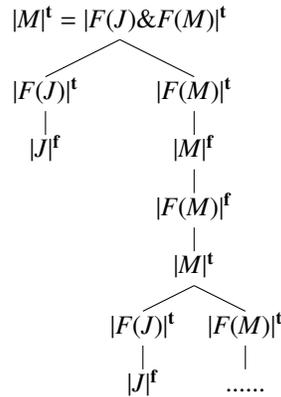
A node is *tethered*, let us say, if it has a finite path to the facts—the facts, recall, are the non-semantic atomic attributions  $|A|^t$  ( $|A|^f$ ) such that  $A$  is true (false) in the underlying model. A branch or tree is tethered if all its nodes are. Looking back now at the trees provided for  $|H|^t$  and  $|H|^f$ , we see that they greatly differ in this respect. In the first, every node is tethered; every node has a finite path to the facts. In the second, *no* node has this property. Maybe the requirement ought to be, not no *infinite* branches, but no *untethered* branches, where a branch is tethered iff every node is tethered; every node has a (finite) path to the facts.

2.  $P$  is TRUE (second stab) iff  $|P|^t$  has a tethered dependence tree.

The Truthfulness-Teller is TRUE by this strengthened standard too. Each occurrence of  $|H|^t$  has *two* paths to the ground, ending in  $|I|^t$  and  $|\neg J|^t$  respectively.  $K$ 's conjunction with  $I$  and  $\neg J$  is not TRUE according to [2.], since the tree has a branch  $|K|^t \rightarrow |K|^t \rightarrow |K|^t \rightarrow \dots$  all of whose elements are untethered.

This idea of tethering speaks to the “comparative” problem of how  $H$  can be better off than  $K$ , even though it in some sense includes  $K$ , or an analogue of  $K$ .  $H$ 's advantage is that every last bit of it is about the facts, in that every node depends on them, whereas  $K$  is floating around absolutely untethered, depending only on itself.<sup>4</sup>

A problem emerges when we consider the Untruthfulness or Mendacity-Teller, *Everything I say is false* (henceforth  $M$ ). Suppose that my only other statement is  $J$  (*Jello is hot*), which is false. Then  $M = F(J)\&F(M)$ ; whence  $|M|^t$  has the following as one of its trees.



(5) Mendacity-Teller

Every node here has a finite path to  $|J|^f$ ;  $|J|^f$  is factual; so every node here is tethered. “Everything I say is false” ought, then, according to [3.], to be TRUE. But it is in reality paradoxical, since if  $M$  is true, then, given that it has  $F(M)$  as a conjunct, it is FALSE. (Whereupon it is TRUE after all, and so on.)

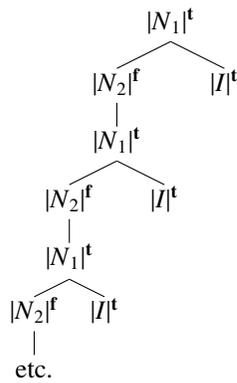
Notice something objectionable about tree (5), however; it has  $|M|^t$  on top and  $|M|^f$  further down, making the tree as a whole inconsistent. Perhaps

**3.  $P$  is TRUE (third try) iff  $|P|^t$  has a *consistent* tethered dependence tree.**

This is better, but even a consistent tethered tree is not enough, as we see from an example of Vann McGee's. Let  $N_1$  be  *$N_2$  is false and ice is cold*, while  $N_2$  is  *$N_1$  is false and ice is cold*. Surely  $N_1$  cannot be TRUE, for then  $N_2$  would have to be FALSE, which is ruled out by symmetry considerations; there is no reason why  $N_2$  should be the FALSE one rather than  $N_1$ . Yet here is a consistent tethered tree for  $|N_1|^t$ :<sup>5</sup>

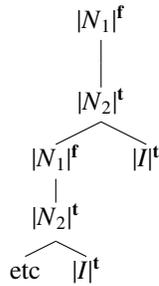
<sup>4</sup>There could be an “unwinding” of  $K$  that does not depend on itself, yet is equally untethered. Kripke notes the possibility of “an infinite sequence of sentences  $P_i$ , where  $P_i$  says that  $P_{i+1}$  is true” ([Kripke(1975)], 693). For unwindings more generally see [Schlenker(2007)] and [Cook(2014)].

<sup>5</sup>Compressed for readability.



(6) McGee Tree

What is interesting is that such a tree is *also* constructible for  $|N_1|^f$ ; it mirrors the tree for  $|N_2|^f$  that is embedded in tree (6).

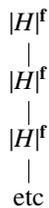


(7) McGee's Other Tree

All right, well maybe [3.] needs to be tightened up a bit:

**4 .**  $P$  is TRUE (fourth and final stab) iff  $|P|^t$  has, while  $|P|^f$  lacks, a consistent tethered dependence tree.

The Truthfulness-Teller  $H$  is TRUE, according to [4.], given that  $|H|^t$  has a tethered tree, if no consistent tethered tree can be constructed for  $|H|^f$ . The only possible tree for  $|H|^f$ , assuming as usual that Moore's other statements are  $I$  and  $\neg J$  (both true), is



(8) Truthfulness-Teller False<sup>6</sup>

This again is untethered, containing not even one node with a finite route to the ground.  $|H|^t$  is thus the only one of  $|H|^t$ ,  $|H|^f$ , to have a consistent tethered tree, which justifies our preference for  $I$  (*Moore never lie* over  $I$  (*Moore do sometimes lie*)).

## 9 Fixed Points

A sentence is grounded-true (-false), we said, iff the corresponding attribution  $|S|^t$  ( $|S|^f$ ) has a dependence tree all of whose branches terminate in facts—atomic attributions  $|A|^t$  ( $|A|^f$ ) such that  $A$  really is true (false) in the underlying model. Kripke’s definition is a bit different; he uses not trees but so-called fixed points—sets of attributions satisfying certain conditions. A fixed point is a set of attributions  $\mathcal{P}$  such that

(A) if  $A$  is atomic,  $|A|^t \in \mathcal{P}$  ( $|A|^f \in \mathcal{P}$ ) iff  $A$  is true (false) in the underlying model.<sup>7</sup>

(N)  $|\neg S|^t \in \mathcal{P}$  iff  $|S|^f \in \mathcal{P}$ ;  $|\neg S|^f \in \mathcal{P}$  iff  $|S|^t \in \mathcal{P}$

(D)  $|S \vee S'|^t \in \mathcal{P}$  iff  $|S|^t \in \mathcal{P}$  or  $|S'|^t \in \mathcal{P}$ ;  $|S \vee S'|^f \in \mathcal{P}$  iff  $|S|^f \in \mathcal{P}$  and  $|S'|^f \in \mathcal{P}$

(U)  $|\forall x \varphi(x)|^t \in \mathcal{P}$  iff  $|\varphi(n)|^t \in \mathcal{P}$  for each  $n$ ;  $|\forall x \varphi(x)|^f \in \mathcal{P}$  iff  $|\varphi(n)|^f \in \mathcal{P}$  for some  $n$

(T)  $|T(S)|^t \in \mathcal{P}$  iff  $|S|^t \in \mathcal{P}$ ;  $|T(S)|^f \in \mathcal{P}$  iff  $|S|^f \in \mathcal{P}$

Suppose that  $\mathcal{A}$  is a set of nonsemantic atomic attributions.  $\mathcal{P}$  is a fixed point over  $\mathcal{A}$  iff it is a fixed point whose nonsemantic atomic attributions are precisely those in  $\mathcal{A}$ . A fixed point of particular interest is  $\mathcal{G}_{\mathcal{A}}$  = the least fixed point over  $\mathcal{A}$ . A sentence  $P$  is grounded-true (-false), for Kripke, assuming  $\mathcal{A}$  sums up the nonsemantic facts, iff  $|P|^t$  ( $|P|^f$ ) belongs to  $\mathcal{G}_{\mathcal{A}}$ , or what comes to the same,  $|P|^t$  ( $|P|^f$ ) belongs to *every* fixed point over  $\mathcal{A}$ . This conforms to our tree-based definition, since  $\mathcal{G}_{\mathcal{A}}$  turns out to be precisely the set of attributions with trees that terminate in the facts, as represented by  $\mathcal{A}$ .<sup>8</sup>

A prima facie advantage of fixed points over trees is that there are lots of them, which makes for a richer taxonomy.  $P$  is *paradoxical*, for instance, if no consistent fixed point contains either  $|P|^t$  or  $|P|^f$ . It is *unstable* iff it is true in some consistent fixed points and false in others. It is *stable* iff it is true in some consistent fixed points and false in none (or vice versa); it receives in other words the same truth-value in every consistent fixed point that’s defined on it, and there are some.  $P$  is *intrinsically* true iff it is true in a thoroughly stable fixed point, meaning, one defined only on stable sentences.

So, for instance,  $H$  and  $K$ —the Truthfulness Teller and the Truth Teller—are both unstable. They are true in some consistent fixed points, and false in others.  $T(K) \supset T(K)$ , however, is stably true: true in those consistent fixed points where it has a truth-value at all. Is it intrinsically true? No, for  $T(K) \supset T(K)$  is evaluable only in fixed points that assign a value to  $K$ , and  $K$  is unstable. An example of an intrinsic truth in the same area is  $E = \neg(T(E) \& \neg T(E))$ —“This very sentence is not both true and untrue.” Its one potential truth-value is *true*, and the one unchanging basis for that truth-value is the fact just mentioned, the fact that it is true. The intrinsic attributions can be joined together into a single compendious fixed point  $\mathcal{I}$ , Kripke shows:  $\mathcal{I}$  is the “maximal intrinsic fixed point.”  $P$  is intrinsically true (false) just if it is true (false) in  $\mathcal{I}$ .

<sup>6</sup>Taken from (3) above.

<sup>7</sup>The underlying model  $M$  is a model, possibly partial, of the T-free part of the language.

<sup>8</sup>[Yablo(1982)]

Now, if one is looking for a compliment that can be paid to ungrounded sentences—which we are, given the true-seemingness of the Truthfulness-Teller— intrinsic truth is a Kripkean’s first thought. (“This sentence is not both true and false” is intrinsically true, as just noted.) Unfortunately it’s a compliment that cannot be paid to  $H$ . The Truthfulness Teller patterns with the Truth Teller in being not even *stably* true, much less intrinsically so. If we stipulate that  $H$  is false in a fixed point, we then provide a reason for its falsity, viz. that it is a counterexample to the generalization that everything one says is true. If we stipulate that it is true, we then eliminate the one possible barrier to its truth, viz, failure to conform to that generalization.

## 10 TRUTH in fixed points

Can *This speaker is truthful* really be no better than *This sentence is true*, from a fixed point perspective? That would be surprising, given the close connection between fixed points and trees. The attributions  $|\varphi|^v$  on a consistent tree can be shown to form a sound partial valuation  $\mathcal{V}$ , meaning a valuation contained in its Kripke Jump  $J(\mathcal{V})$ . Sound valuations generate fixed points  $\mathcal{V}^*$  under repeated application of  $J$  (since  $J$  is monotonic), so every tree can be considered the seed of a fixed point. And of course there will be other fixed points above  $\mathcal{V}^*$  involving attributions not forced by  $\mathcal{V}$  but allowed by it.

This forced/allowed distinction will be useful in trying to distinguish  $H$  from  $K$  in fixed point terms.  $K$  has no factual prerequisites and faces no factual obstacles. No matter what the ground-level facts  $\mathcal{A}$  may be,  $K$  is true in some fixed points above  $\mathcal{A}$  and false in others.  $K$  and  $\neg K$  are both in a sense necessarily, or unconditionally, possible: each holds in *some* fixed point above every factual ground.

The Truthfulness-Teller is different in this respect.  $H$  can be true only in fixed points above  $\mathcal{A}$ ’s with facts like ice being cold and Joad not being sneaky.  $H$  is only *conditionally* possible.  $H$ ’s negation however is unconditionally possible just like  $K$ ; whatever the ground-level facts may be, we can consistently treat  $H$  as false in virtue of its own falsity Here then is one way in which  $H$  is more factual than its negation, and then  $K$  and its negation. Of the four, it is the only that owes its construability as true to the way things actually turned out.

Now this is not quite enough because it holds of  $K\&I$ —*This sentence (the first conjunct) is true & Snow is white*—as well that it owes its construability as true to the way things turned out. while its negation is construable as true no matter what (by letting  $K$  be false). And yet  $K\&I$  certainly does not strike us as TRUE, as discussed earlier in this pper.

Suppose we use *fact-bound* for the property of being construable as true in *these* factual circumstances— $\mathcal{A}_@$ — but not in all factual circumstances. The the problem with  $K\&I$  is that while it is fact-bound taken as a whole, its first conjunct is unconditionally possible or fact-free. What is special about the Truthfulness-Teller is that it is *thoroughly* fact-bound, not an amalgam of something fact-bound with something fact-free.

How to define this in fixed point terms? Consider the fixed points above  $\mathcal{A}_@$ . For one of these to be fact-bound, all of its component attributions should be fact-bound; it should contain nothing that is unconditionally possible, nothing that is construable as true no matter what. A statement is thoroughly fact-bound, I propose, iff it is true in a fact-bound fixed point.

I’m hoping this reminds you of what we said about tethered trees. Tethered trees can have infinite paths but every node has got to depend on some fact or other. The two notions are connected as follows. Suppose a node has no route to the ground in a given tree. The subtree that it stands at the top of is free of ground-level attributions. But then the subtree’s contents can be added to any  $\mathcal{A}$  whatsoever to obtain the seed or a fixed point where the sentence has the truth-value assigned it in the tree. *An untethered tree must therefore contain elements that are fact-free.*

Suppose conversely that an attribution  $|\varphi|^v$  is fact-free, that is, unconditionally possible. Then for every

$\mathcal{A}$  whatsoever there is a fixed point above  $\mathcal{A}$  that assigns  $v$  to  $\varphi$ . This is so in particular if  $\mathcal{A}$  is the empty set.

Fixed points by definition satisfy conditions **(N)** for negation, **(D)** for disjunction, **(U)** for quantification, and **(T)** for truth. The left-to-right directions of these rules give us all we need to construct a tree for  $|\varphi|^v$ . The tree is going to be untethered because there were no ground-level attributions in the fixed point:  $\mathcal{A}$  is the empty set. We have shown that

**Theorem 1**  $|\varphi|^v$  has a consistent tethered tree based in actual facts iff it belongs to a fact-bound fixed point.

**Corollary 1.1**  $\varphi$  is TRUE iff it is true in at least one fact-bound fixed point and false in no such fixed points.

This second way of putting it is interesting because it links up with some existing work on aboutness and partial truth. A statement that is not true full stop may still be true about a certain subject matter. The challenge is to identify a subject matter  $M$  such that  $\varphi$  is TRUE iff it is true about  $M$ , or true where  $M$  is concerned.

## 11 Truth about M

... is a tricky notion. David Lewis's theory of aboutness in "Statements Partly About Observation" gives us a good place to start ([Lewis(1988)]).

Sentence  $S$  is about subject matter  $M$ , for Lewis, just if worlds alike where  $M$  is concerned are alike where  $S$ 's semantic properties are concerned. So, for instance, *The number of stars is prime* is about how many stars there are, since worlds with equally many stars agree in whether the number of stars in them is prime.<sup>9</sup> *The number of stars exceeds the number of planets* is not in Lewis's sense about the number of stars, for Lewis, since its truth-value can change though the number of stars remains what it is. A statement is true about  $M$  in a world  $w$  iff it is true in some world  $u$  that is just like  $w$  where  $M$  is concerned.

Worlds for these purposes can be fact-bound fixed points, meaning, fixed points that are fact-bound relative to some choice  $\mathcal{A}$  of non-semantic facts. Worlds are equivalent iff their  $\mathcal{A}$ 's—their non-semantic atomic facts—are exactly the same. Calling the subject matter  $G$ ,  $\varphi$  is true about  $G$  in  $w$  iff it is true in some fact-bound fixed point agreeing with  $w$  in its non-semantic atomic facts. The actual world @ is the least fixed point based on the actual facts. TRUTH can now be defined as follows:

$\varphi$  is TRUE iff it is true about  $G$  in @, that is, true in a fact-bound fixed point whose facts are the actual ones.

This is fair enough but we can bring subject matter to bear in a deeper and more revealing way. Because Lewis's notion of aboutness is not the only one we can imagine.

*The number of stars exceeds the number of planets* is not in Lewis's sense about the number of stars, since its truth-value can change though the number of stars remains what it is. But, of course, there is another sense in which *The number of stars exceeds the number of planets* IS about the number of stars; its truth-value is *sensitive* to how many stars there are, in that, for instance, there can't be zero stars compatibly with the stars outnumbering the planets. And one can imagine conversely a sentence that is about the number of stars in the supervenience sense but not the sensitivity or difference-making sense. *The number of stars is positive* is supervenience-about how many stars there are, in that worlds  $M$ -alike are always  $S$ -alike, but not differentially about how many stars there are, in that  $M$ -different worlds—worlds with unequally many stars—do not thereby differ in whether the number of stars in them is positive ([Yablo(2014)]).

So, where Lewis's *supervenience*-based notion of aboutness focuses on whether  $S$ 's semantic properties hold fixed when you hold the state of things wrt  $m$  fixed, another notion, the *differential* notion, looks rather

---

<sup>9</sup>[Lewis(1988)]

at how  $S$ 's semantic properties are apt to *change* if you vary the state of things with respect to  $m$ . And although  $H$  and  $K$  may be equally about the facts in the supervenience sense, they are not equally about the facts in the difference-making sense.

Changing the (worldly) facts  $\mathcal{A}$  has no effect on the Truth Teller whatever—it can be true or false as you please—but the Truthfulness Teller loses its shot at truth if we move to a world where Joad is sneaky. The Truthfulness Teller outdoes the Truth Teller aboutness-wise because *changing* the facts has the potential to *change*  $H$ 's semantic properties, but not the potential to change the semantic properties of  $K$ . This obviously links up with our talk earlier of fact-boundness and fact-freedom, and it would be interesting to try (at some point—not today!) to reformulate these earlier notions in differential aboutness terms.

## 12 Conclusion

### References

- [Barwise and Etchemendy(1989)] J. Barwise and J. Etchemendy. *The liar: An essay on truth and circularity*. Oxford University Press, USA, 1989.
- [Boolos(1971)] George Boolos. The iterative conception of set. *The Journal of Philosophy*, pages 215–231, 1971.
- [Cook(2014)] Roy T Cook. *The Yablo Paradox: An Essay on Circularity*. Oxford University Press, 2014.
- [Davis(1979)] Lawrence Davis. An alternate formulation of Kripke's theory of truth. *Journal of Philosophical Logic*, 8(1):289–296, 1979.
- [Hazen(1981)] Allen Hazen. Davis's formulation of Kripke's theory of truth: A correction. *Journal of Philosophical Logic*, pages 309–311, 1981.
- [Herzberger(1970)] Hans G Herzberger. Paradoxes of grounding in semantics. *The Journal of Philosophy*, pages 145–167, 1970.
- [Kripke(1975)] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.
- [Lewis(1988)] David Lewis. Statements partly about observation. In *Papers in philosophical logic*. Cambridge University Press, 1988.
- [McLarty(1993)] Colin McLarty. Anti-Foundation and Self-Reference. *Journal of Philosophical Logic*, 22(1):19–28, 1993.
- [Mirimanoff(1917)] Dimitry Mirimanoff. *Les antinomies de Russell et de Burali-Forti: et le problème fondamental de la théorie des ensembles*. Enseignement mathématique, 1917.
- [Schlenker(2007)] Philippe Schlenker. The elimination of self-reference: Generalized yablo-series and the theory of truth. *Journal of philosophical logic*, 36(3):251–307, 2007.
- [Smullyan(1986)] Raymond Smullyan. *What is the name of this book?* Touchstone Books, 1986.
- [Yablo(1982)] Stephen Yablo. Grounding, dependence, and paradox. *Journal of Philosophical Logic*, 11: 117–138, 1982.

- [Yablo(1993)] Stephen Yablo. Hop, skip and jump: The agonistic conception of truth. *Philosophical Perspectives*, pages 371–396, 1993.
- [Yablo(2004)] Stephen Yablo. Circularity and paradox. *Self-reference*, pages 139–157, 2004.
- [Yablo(2014)] Stephen Yablo. *Aboutness*. Princeton University Press, 2014.
- [Yuting(1953)] Shen Yuting. Paradox of the Class of All Grounded Classes. *Journal of Symbolic Logic*, 18 (2):114, 1953.