Chapter 7
Knights, Knaves, Truth, Truthfulness, Grounding, Tethering, Aboutness, and Paradox

Stephen Yablo

Abstract Knights always tell the truth; Knaves always lie. Knaves for familiar reasons cannot coherently describe themselves as liars. That would be like Epimenides the Cretan accusing all Cretans of lying. Knights do not *intuitively* run into the same problem. What could prevent a Knight from truly reporting that s/he always tells the truth? Standard theories of truth DO prevent this, however, for such a report is self-referentially ungrounded. Standard theories have a problem, then! We try to fix it.

7.1 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* speaks the truth, for knights and knaves *both* say, "I am a knight"—knights because "I am a knight" is true in their mouths,

S. Yablo (✉)
Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: yablo@mit.edu

M. Fitting and B. Rayman (eds.), *Raymond Smullyan on Self Reference*,
Outstanding Contributions to Logic 14, https://doi.org/10.1007/978-3-319-68732-2_7

and knaves because it is false in theirs. Since his description of *A* is true, *B* must be a knight. *B* might equivalently ~~had~~ have replied that *A* said he always told the truth, for that is the kind of speech behavior that is definitive of a knight.

have

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.

7.2 Russell and Moore

The Smullyan puzzle recalls a remark of Kripke's about Russell's sense of, or radar for, paradox. Russell 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, 691–692)

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."

7.3 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![1] 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.

7.4 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:

> 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* (overdetermined) but *indeterminate* (underdetermined).

Return now to *Everything I say is true*. I will call it the Truthfulness-Teller, because the speaker (Moore, we suppose) is declaring himself to be generally truthful, and

[1] Self-identified knights are the group Smullyan admires the most. If they were talking nonsense, he would have noticed it.

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.[2]

7.5 Kripke and Dependence Trees

There are really two puzzles here. One, the comparative puzzle, asks why the Truthfulness-Teller should seem 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.[3]

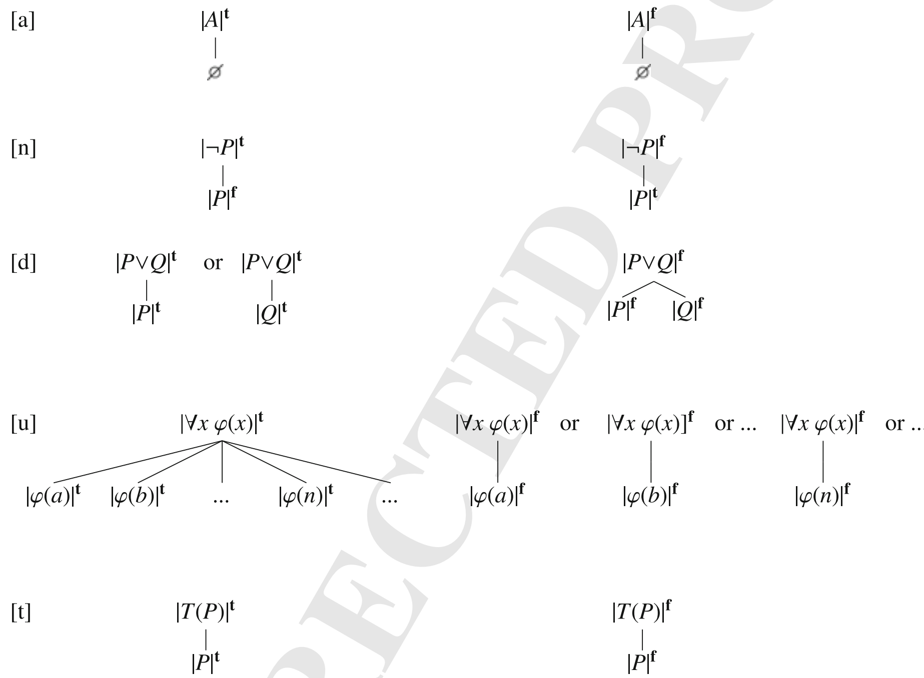
To appreciate how the theory works, let's associate with each sentence P two "attributions" $|P|^{\mathbf{t}}$ and $|P|^{\mathbf{f}}$, one assigning truth to P, the other falsity. A relation Δ on the set of attributions is a *dependence* relation iff it satisfies these conditions:

[2] 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.

[3] 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 (2006).

(a) if A is atomic, $|A|^{\mathbf{t}}$ and $|A|^{\mathbf{f}}$ bear Δ to nothing (written $\varnothing$)
(n) $|\neg P|^{\mathbf{t}}$ bears Δ to $|P|^{\mathbf{f}}$; $|\neg P|^{\mathbf{f}}$ bears Δ to $|P|^{\mathbf{t}}$
(d) $|P \vee Q|^{\mathbf{t}}$ bears Δ either to $|P|^{\mathbf{t}}$ or $|Q|^{\mathbf{t}}$; $|P \vee Q|^{\mathbf{f}}$ bears Δ both to $|P|^{\mathbf{f}}$ and $|Q|^{\mathbf{f}}$
(u) $|\forall x\ \varphi(x)|^{\mathbf{t}}$ bears Δ to $|\varphi(n)|^{\mathbf{t}}$ for each name n; $|\forall x\ \varphi(x)|^{\mathbf{f}}$ bears Δ to $|\varphi(n)|^{\mathbf{f}}$ for some particular name n
(t) $|T(A)|^{\mathbf{t}}$ bears Δ to $|A|^{\mathbf{t}}$; $|T(A)|^{\mathbf{f}}$ bears Δ to $|A|^{\mathbf{f}}$

P is *grounded-true* iff there is a dependence relation Δ such that every Δ-path starting from $|P|^{\mathbf{t}}$ leads to a fact—an atomic attribution $|A|^{\mathbf{t}}$ ($|A|^{\mathbf{f}}$) such that A really is true (false) in reality, as represented by the underlying model. Equivalently, $|P|^{\mathbf{t}}$ sits atop a factual Δ 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 ✓ if they're factual and ✗ if they conflict with the facts. To get a decorated tree from a plain one, one starts by tagging terminal nodes with ✓ s and ✗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 $|P|^{\mathbf{t}}$✓ 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.

447426_1_En_7_Chapter ✓ TYPESET DISK LE ✓ CP Disp.:16/10/2017 Pages: 18 Layout: T1-Standard

$$
\begin{array}{c}
|(\exists x,y\text{: } Brx \ \& \ Bry)\ T(x) \ \& \ F(y)|^{\mathbf{t}} \\
\begin{array}{cc}
|T(I)|^{\mathbf{t}} & |F(J)|^{\mathbf{t}} \\
| & | \\
|I|^{\mathbf{t}} & |J|^{\mathbf{f}}
\end{array}
\end{array}
$$

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

$$
\begin{array}{c}
|(\exists x,y\text{: } Brx \ \& \ Bry)\ T(x) \ \& \ F(y)|^{\mathbf{t}} \\
\begin{array}{cc}
|T(I)|^{\mathbf{t}} & |F(J)|^{\mathbf{t}} \\
| & | \\
|I|^{\mathbf{t}}\checkmark & |J|^{\mathbf{f}}\checkmark
\end{array}
\end{array}
$$

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,

$$
\begin{array}{c}
|(\exists x,y\text{: } Brx \ \& \ Bry)\ T(x) \ \& \ F(y)|^{\mathbf{t}} \\
\begin{array}{cc}
|T(I)|^{\mathbf{t}}\checkmark & |F(J)|^{\mathbf{t}}\checkmark \\
| & | \\
|I|^{\mathbf{f}}\checkmark & |J|^{\mathbf{f}}\checkmark
\end{array}
\end{array}
$$

and at stage 3,

$$
\begin{array}{c}
|\ (\exists x,y\text{: } Brx \ \& \ Bry)\ T(x) \ \& \ F(y)|^{\mathbf{t}}\checkmark \\
\begin{array}{cc}
|T(I)|^{\mathbf{t}}\checkmark & |F(J)|^{\mathbf{t}}\checkmark \\
| & | \\
|I|^{\mathbf{t}}\checkmark & |J|^{\mathbf{f}}\checkmark
\end{array}
\end{array}
$$

A decorated tree headed by $|\varphi|^{\mathbf{t}}$✓ means that φ 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.

$$
\begin{array}{c}
|L|^{\mathbf{t}}\ (= \ |\neg T(L)|^{\mathbf{t}}) \\
| \\
|T(L)|^{\mathbf{f}} \\
| \\
|L|^{\mathbf{f}} \\
| \\
|T(L)|^{\mathbf{t}} \\
| \\
|L|^{\mathbf{t}} \\
| \\
\text{etc}
\end{array}
\qquad\qquad
\begin{array}{c}
|L|^{\mathbf{f}}\ (= |\neg T(L)|^{\mathbf{f}}) \\
| \\
|T(L)|^{\mathbf{t}} \\
| \\
|L|^{\mathbf{t}} \\
| \\
|T(L)|^{\mathbf{f}} \\
| \\
|L|^{\mathbf{f}} \\
| \\
\text{etc}
\end{array}
$$

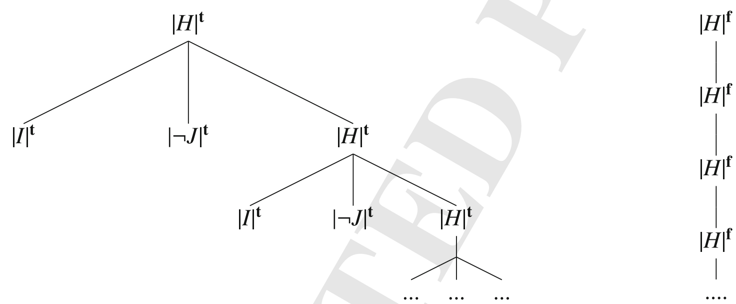
(1) Liar

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 $|P|^{\mathbf{t}}$ and $|P|^{\mathbf{f}}$ 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 = *Ice is cold* and $\neg J$ = *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 $|H|^{\mathbf{t}}$ 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 something very like true. There might be some support for this idea in the fact that $|H|^{\mathbf{t}}$'s tree is "better"—more grounded in nonsemantic facts—than $|H|^{\mathbf{f}}$'s. We'll return to this theme in a moment.

7.6 Immodesty

The Truthfulness-Teller is ungrounded, on Kripke's theory, because it immodestly extends to itself the compliment (truth) that it pays to other sentences. I can think of two ways to make it less immodest, so as to give it a better shot at truth. We could muck with the subject term, so that it covered fewer sentences. Or we could scale back the predicate, so that it attributed a weaker property.

On the first strategy, we mistake *All my statements are true* for H_1, which attributes truth only to Moore's *other* statements. That we were taking it for H_1 nicely explains

447426_1_En_7_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:16/10/2017 Pages: 18 Layout: T1-Standard

why H would strike us as true. H_1 really *is* true; Moore's other statements really do have the property (truth) that's attributed to them. This approach also explains why the Truth-Teller seems worse off than the Truthfulness-Teller. If we cut back K's subject term ("this very sentence"), then nothing is left; there are no other statements that K describes inter alia as true. K is worse off than H because there is no worthwhile K_1 standing to it as H_1 stands to H.

These results are obtained, however, by twisting H's intuitive content out of recognition. *All my statements are true...with the possible exception of this one* is the statement of some kind of trickster, not a George Edward Moore. To exempt his declaration of honesty from its own extension is the last thing Moore wants. Here then is our first condition on a satisfactory solution: *All my statements are true* should not make an exception of itself.

Doesn't this make the problem unsolvable, though? For Moore's declaration 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 meant is H_2: Everything I say is true-to-the-extent-evaluable.

This again does violence to the content. Suppose Moore had on other occasions uttered a bunch of ungrounded nonsense: Liars and Truth-Tellers and whatever other semantic pathologies you like. His statements are true to the extent evaluable, just because they are not evaluable. 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, as it might be. This is what we suggest below (Sect. 7.11): even if H is not true full stop, still it is true to the facts. The problem of course is to identify this new subject matter. I propose to creep up on it slowly, by way of liberalized dependence trees.

7.7 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.

447426_1_En_7_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:16/10/2017 Pages: 18 Layout: T1-Standard

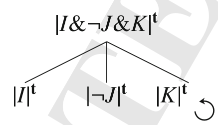
0. P is true iff $|P|^{\mathbf{t}}$ sits atop a dependence tree all of whose branches terminate in facts.

It is grounded-false iff $|P|^{\mathbf{f}}$ 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 $|P|^{\mathbf{t}}$ 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 $|P|^{\mathbf{t}}$ ✓ heads a decorated dependence tree constructed to slightly weaker specifications: a parent node is marked ✓ iff (i) some of its children are marked ✓, and (ii) none if its (other) children are marked ✗. 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 that way in certain constructions. An example is $I\&\neg J\&K$, where I and $\neg J$ are plain truths, and K is again the Truth-Teller. The tree is



(4) Truth-Teller Plus

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

7.8 TRUTH and Tethering

I want to go back now to an idea from Sect. 7.7: some ungrounded attributions are closer to being grounded than others. A glance at their trees makes clear that $|H|^{\{\mathbf{t}}$, for instance, is less ungrounded than $|I\&\neg\, J\&\, K\,|^{\mathbf{t}}$, which is less ungrounded than $|K|^{\mathbf{t}}$, and also less ungrounded than $|H|^{\mathbf{f}}$. In what sense, though?

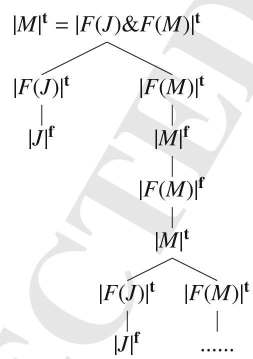
A node is *tethered*, let us say, if it has a finite path to the facts—a fact, recall, is a non-semantic atomic attribution $|A|^{\mathbf{t}}$ ($|A|^{\mathbf{f}}$) such that A is true (false) in the underlying model. A branch or tree is tethered if all its nodes are. Looking back now at the trees provided for $|H|^{\mathbf{t}}$ and $|H|^{\mathbf{f}}$, 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 $|P|^{\mathbf{t}}$ has a tethered dependence tree.

The Truthfulness-Teller is TRUE by this strengthened standard too. Each occurrence of $|H|^{\mathbf{t}}$ has *two* paths to the ground, ending in $|I|^{\mathbf{t}}$ and $|\neg J|^{\mathbf{t}}$ respectively. K's conjunction with I and $\neg J$ is not TRUE according to [**2.**], since the tree has a branch $|K|^{\mathbf{t}} \rightarrow |K|^{\mathbf{t}} \rightarrow |K|^{\mathbf{t}} \rightarrow$ ... 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 hooks up with the facts— every node on its tree depends on them—whereas K is floating around absolutely untethered, depending only on itself.[4]

A problem emerges when we consider the *Un*truthfulness 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 $|M|^{\mathbf{t}}$ has the following as one of its trees.



(5) Mendacity-Teller

Every node here has a finite path to $|J|^{\mathbf{f}}$; $|J|^{\mathbf{f}}$ 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 $|M|^{\mathbf{t}}$ on top and $|M|^{\mathbf{f}}$ further down, making the tree as a whole inconsistent. Perhaps

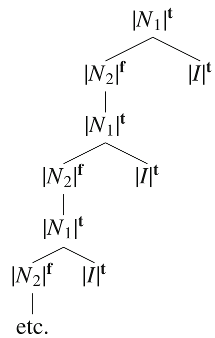
3. P is TRUE (third try) iff $|P|^{\mathbf{t}}$ has a *consistent* tethered dependence tree.

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

[4]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).

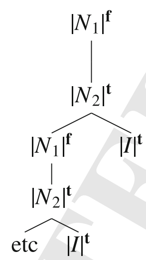
447426_1_En_7_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:16/10/2017 Pages: 18 Layout: T1-Standard

should be the FALSE one rather than N_1. Yet here is a consistent tethered tree for $|N_1|^{\mathbf{t}}$[5]:



(6) McGee Tree

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



(7) McGee's Other Tree

The McGee trees show that [**3**.] needs to be tightened up a bit:

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

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

(8) Truthfulness-Teller False[6]

[5]Compressed for readability.

This[6] again is untethered, containing not even one node with a finite route to the ground. $|H|^{\mathbf{t}}$ is thus the only one of $|H|^{\mathbf{t}}$, $|H|^{\mathbf{f}}$, to have a consistent tethered tree, which justifies our preference for *I (Moore) never lie* over *I (Moore) do sometimes lie.*

7.9 Fixed Points

Subject matters as we are going to be conceiving them (following Lewis) are equivalence relations on worlds. What plays the world role in this application are fixed points. These are much better known than dependence trees, but require a bit of explanation as they haven't been mentioned yet in this paper.

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

(A) if A is atomic, $|A|^{\mathbf{t}} \in \mathcal{P}$ ($|A|^{\mathbf{f}} \in \mathcal{P}$) iff A is true (false) in the underlying model.[7]
(N) $|\neg S|^{\mathbf{t}} \in \mathcal{P}$ iff $|S|^{\mathbf{f}} \in \mathcal{P}$; $|\neg S|^{\mathbf{f}} \in \mathcal{P}$ iff $|S|^{\mathbf{t}} \in \mathcal{P}$
(D) $|S \vee S'|^{\mathbf{t}} \in \mathcal{P}$ iff $|S|^{\mathbf{t}} \in \mathcal{P}$ or $|S'|^{\mathbf{t}} \in \mathcal{P}$; $|S \vee S'|^{\mathbf{f}} \in \mathcal{P}$ iff $|S|^{\mathbf{f}} \in \mathcal{P}$ and $|S'|^{\mathbf{f}} \in \mathcal{P}$
(U) $|\forall x\, \varphi(x)|^{\mathbf{t}} \in \mathcal{P}$ iff $|\varphi(n)|^{\mathbf{t}} \in \mathcal{P}$ for each n; $|\forall x\, \varphi(x|^{\mathbf{f}} \in \mathcal{P}$ iff $|\varphi(n)|^{\mathbf{f}} \in \mathcal{P}$ for some n
(T) $|T(S)|^{\mathbf{t}} \in \mathcal{P}$ iff $|S|^{\mathbf{t}} \in \mathcal{P}$; $|T(S)|^{\mathbf{f}} \in \mathcal{P}$ iff $|S|^{\mathbf{f}} \in \mathcal{P}$

If these rules look familiar, and they should, it's because the left-to-right directions of **(A)**-**(T)** are the same as the tree rules **(a)**-**(t)** laid down in Sect. 7.5.

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 $\mathcal{A}$. Kripke finds work for lots of fixed points, but the one he particularly emphasizes is

$$\mathcal{G}_{\mathcal{A}} = \text{the least fixed point over } \mathcal{A}.$$

A sentence S is grounded-true (-false), for Kripke, given nonsemantic facts $\mathcal{A}$, iff $|S|^{\mathbf{t}}$ ($|S|^{\mathbf{f}}$) belongs to $\mathcal{G}_{\mathcal{A}}$—or, what is really no different, $|S|^{\mathbf{t}}$ ($|S|^{\mathbf{f}}$) belongs to *every* fixed point over $\mathcal{A}$. This conforms to our tree-based definition, since $\mathcal{G}_{\mathcal{A}}$ turns out (unsurprisingly) to be precisely the set of attributions with trees that terminate in the facts, as represented by $\mathcal{A}$.[8]

A prima facie advantage of fixed points over trees is that they make for a richer taxonomy. P is *paradoxical*, for instance, if no fixed point contains either $|P|^{\mathbf{t}}$ or $|P|^{\mathbf{f}}$.[9] It is *unstable* iff it is true in some consistent fixed points and false in others. It

[6]Taken from (3) above.
[7]The underlying model M is a model, possibly partial, of the T-free part of the language.
[8]Yablo (1982).
[9]No consistent fixed point that is; but we have defined fixed points so that all of them are consistent.

447426_1_En_7_Chapter ☑ TYPESET ☐ DISK ☐ LE ☑ CP Disp.:16/10/2017 Pages: 18 Layout: T1-Standard

is *stable* iff it is true in some fixed points and false in none (or vice versa); it receives in other words the same truth-value in every 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.

Both H and K—-the Truthfulness Teller and the Truth Teller—-are unstable; they are true in some fixed points, clearly, 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 intrinsic truth in the same neighborhood 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 of E’~~S~~s truth. The intrinsic attributions can be joined together into a single compendious fixed point $\mathcal{I}$, Kripke shows, the “maximal intrinsic fixed point.” P is intrinsically true (false) just if it is true (false) in $\mathcal{I}$.

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.) It’s a compliment that cannot be paid to H, however. The Truthfulness-Teller patterns with the Truth Teller in being not even stably true, much less intrinsically so. If we stipulate that ~~TRUTH~~H is false in a fixed point, we then provide a reason for its falsity; it is a counterexample to the generalization that everything Moore says is true. If we stipulate that it is true, we eliminate the one possible counterexample to its truth, namely, itself.

7.10 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.

If we gather together all the attributions $|\varphi|^{\mathbf{v}}$ on a consistent tree, we get a partial valuation $\mathcal{V}$ that is closed under the left-to-right directions of (**A**)-(**T**); a valuation like that is called *sound*. Sound valuations generate fixed points $\mathcal{V}*$ under repeated application of the right-to-left directions of (**A**)-(**T**). Every tree is in that sense 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 is the key to distinguishing H from K in fixed point terms. K has no factual prerequisites and faces no factual threats. 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 *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 making Moore’s other statements true: ice has got to be cold and Jello cannot be hot. H is only *conditionally* possible. The result $\neg H$ of negating it is, however, unconditionally possible just like K; whatever the ground-level facts may be, we can consistently treat H as false by virtue of its own falsity. H is more beholden to the actual facts than its negation, and than K and its negation. Of the four, it is the only one that owes its construability as true to the way things actually turned out.

Now this is not quite enough for TRUTH, for it holds of $K\&I$—*This sentence (up to the ampersand) is true & Snow is white*—as well that (i) it owes its construability as true to the way things turned out, while (ii) 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, to repeat an observation made earlier.

Suppose we use *fact-dependent* for the property of being construable as true in *these* factual circumstances—$\mathcal{A}_@$—but not in *all* factual circumstances. The problem with $K\&I$ is that while it is fact-dependent 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-dependent, not an amalgam of something fact-dependent 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-dependent, all of its component attributions should be fact-dependent; it should contain nothing that is unconditionally possible, nothing that is construable as true no matter what. An attribution is *thoroughly fact-dependent* iff it belongs to a fixed point *all* of whose attributions are fact-dependent,

This is reminiscent of what we said about tethered trees; the attributions on them may not all be grounded, but they all have finite paths to the ground. The two notions—tethered tree and thoroughly fact-dependent fixed point—are connected, it turns out. $|\varphi|^t$ heads a consistent tethered tree just if φ is true in at least one thoroughly fact-dependent fixed point. (For short, one ~~a~~ fact-dependent fixed point.)

If a tree is untethered, it has a node n with no finite path to the non-semantic atomic facts. The subtree that n heads must therefore be free of such facts. Let $\mathcal{N}$ be the subtree's contents = the set of all attributions on it. These attributions form a sound set (the contents of any tree make a sound set) that is consistent with any $\mathcal{A}$ (because $\mathcal{A}$ is made up of ground-level attributions and $\mathcal{N}$ is free of such attributions). $\mathcal{N} \cup \mathcal{A}$ generates a fixed point containing the attribution in n ($|\varphi|^{\mathbf{v}}$, let's say) by application of the right-to-left directions of closure rules (**A**)-(**T**). $|\varphi|^{\mathbf{v}}$ is fact-independent because $\mathcal{A}$ was arbitrary. *An untethered tree must therefore contain elements that are fact-independent.*

Suppose conversely that an attribution $|\varphi|^{\mathbf{v}}$ is fact-independent, that is, $|\varphi|^{\mathbf{v}}$ is unconditionally possible. Then for every $\mathcal{A}$ whatsoever there is a fixed point above $\mathcal{A}$ that assigns $\mathbf{v}$ to φ. 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 $|\varphi|^{\mathbf{v}}$. 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

Lemma $|\varphi|^{\mathbf{v}}$ *has a consistent tethered dependence tree iff it belongs to a fact-dependent fixed point.*

From this it follows that

Theorem *φ is TRUE iff it is true in at least one fact-dependent fixed point and false in no such fixed points.*[10]

The theorem bears on a problem posed above: can a subject matter be identified such that φ is TRUE iff it is true about that subject matter?

7.11 True to the FACTS

A *subject matter*, for Lewis, is an equivalence relation on worlds.[11] Sentence S is *wholly about subject matter* M just if S's truth-value never varies between M-equivalent worlds. *The number of stars is prime* is wholly about how many stars there are, since worlds with equally many stars cannot disagree on whether their stars are prime in number.[12] *The number of stars exceeds the number of planets* is *not* wholly about the number of stars, since its truth-value can change though the number of stars holds fixed. Now we define the notion of truth at a world where a given subject matter is concerned: we define

(TAM) S is true about M in a world w iff it is true ~~(period)~~ in a world M-equivalent to w.

Can (TAM) be all one one line please, with s\pace above and below?

Worlds for these purposes can be fixed points, as indicated earlier.[13] The facts in two worlds are the same if, although they may evaluate T-sentences differently, regular old non-semantic atomic sentences have the same truth-value in both of them. Suppose that w and w' are fact-dependent. They agree on subject matter F, short for FACTS, just if the same facts obtain in both of them.

(SMF) Two worlds are ~~F-equivalent iff (i) both are fact-dependent, and (ii) their facts are the same.~~

Can (SMF) be written as indicated please, with space above and below?

φ By (TAM), φ is true about F in w iff φ is true in a fact-dependent fixed point w' agreeing with w in its non-semantic atomic facts. Consider now truth about F in F the actual world $w_{@}$, defined as the least fixed point based on the actual facts. ~~; Our Theorem above can be restated as follows.~~

Theorem* all on one line please

Theorem* *φ is TRUE iff it is true in the actual world where the FACTS are concerned.*

Can Theorem* be un-indented like Theorem on the top of the page?

This is a shorter way of saying, as we did above, that to be TRUE is to be true in at least one fact-dependent fixed point whose facts are the actual ones. The paper could end right here, but I have a parting speculation I'd like to get on the table.

The notion of aboutness we get from Lewis is important and interesting. But it is not the only one possible. We saw for instance that *The number of stars exceeds*

[10] If φ is TRUE, then $|\varphi|^{\mathbf{t}}$ has a consistent tethered dependence tree and $|\varphi|^{\mathbf{f}}$ doesn't. By the Lemma, φ is true in a fact-dependent fixed point but not false in any fact-dependent fixed points. The converse is similar.

[11] Lewis (1988).

[12] Lewis (1988).

[13] We will be interested only in fact-dependent fixed points, more carefully, fixed points that are fact-dependent relative to some choice $\mathcal{A}$ of non-semantic atomic facts.

 Text

(SMF) Worlds are F-equivalent iff
(i) both are fact-dependent, and
(ii) the same facts obtain in each.

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 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; there can't be zero stars, for instance, compatibly with the stars outnumbering the planets.

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 $\mathcal{S}$-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).

~~The claim so far is that w~~ W here Lewis's *supervenience*-based notion of aboutness focuses on whether φ $\mathcal{S}$'s semantic properties hold fixed when you hold the state of things wrt M fixed, there is another notion, the *differential* notion, that looks rather at how φ $\mathcal{S}$'s semantic properties are apt to *change* if you vary the state of things with respect to M. *H* and *K* may be equally about the facts (or not) in the supervenience sense, but they are not equally about the facts in the difference-making sense. What do I mean by this?

Changing the 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 Jello is hot. The Truthfulness Teller outdoes the Truth Teller differential-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-dependence and fact-freedom, and it ~~would be~~ might be possible ~~interesting to try (at some point)~~ to reformulate ~~these~~ those earlier notions in differential aboutness terms.

7.12 Conclusion

The way is now clear for Moore to call himself honest without falling afoul of the strictures imposed by the best known theory of truth. Knights are encouraged to avail themselves of this opportunity, too.

References

Barwise, J., & Etchemendy, J. (1989). *The liar: An essay on truth and circularity*. USA: Oxford University Press.

Boolos, G. (1971). The iterative conception of set. *The Journal of Philosophy,* 215–231.

Cook, R. T. (2014). *The Yablo paradox: An essay on circularity*. Oxford: Oxford University Press.

Davis, L. (1979). An alternate formulation of Kripke's theory of truth. *Journal of Philosophical Logic*, *8*(1), 289–296.

Hazen, A. (1981). Davis's formulation of Kripke's theory of truth: A correction. *Journal of Philosophical Logic*, *10*(3), 309–311.

Herzberger, H. G. (1970). Paradoxes of grounding in semantics. *The Journal of Philosophy*, 145–167.

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

Lewis, D. (1988). Statements partly about observation. In *Papers in philosophical logic*. Cambridge: Cambridge University Press.

McLarty, C. (1993). Anti-Foundation and Self-Reference. *Journal of Philosophical Logic*, *22*(1), 19–28.

Mirimanoff, D. (1917). *Les antinomies de Russell et de Burali-Forti: et le problème fondamental de la théorie des ensembles*. Enseignement mathématique.

Schlenker, P. (2007). The elimination of self-reference: Generalized yablo-series and the theory of truth. *Journal of philosophical logic*, *36*(3), 251–307.

Smullyan, R. (1986). *What is the name of this book?* Touchstone Books.

Yablo, S. (1982). Grounding, dependence, and paradox. *Journal of Philosophical Logic*, *11*, 117–138.

Yablo, S. (1993). Hop, skip and jump: The agonistic conception of truth. *Philosophical Perspectives*, *7*, 371–396.

Yablo, S. (2006). Circularity and paradox. In T. Bolander, V. F. Hendricks, & S. A. Pedersen (Eds.), *Self-reference* (pp. 139–157). Stanford: CSLI Publications.

Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.

Yuting, S. (1953). Paradox of the class of all grounded classes. *Journal of Symbolic Logic*, *18*(2), 114.

Author Queries

Chapter 7

Query Refs.	Details Required	Author's response
AQ1	Please check and confirm if the author and respective affiliation has been correctly identified. Amend if necessary.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	
Insert in text the matter indicated in the margin		New matter followed by or
Delete	/ through single character, rule or underline or through all characters to be deleted	or
Substitute character or substitute part of one or more word(s)	/ through letter or through characters	new character / or new characters /
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	/ through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	,
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	/ through character or where required	
Reduce space between characters or words	between characters or words affected	