Remarks on a Foundationalist Theory of Truth

Anil Gupta
University of Pittsburgh

I

Tim Maudlin’s *Truth and Paradox* offers a theory of truth that arises from a foundationalist picture of language.¹ The picture is attractive, and Maudlin builds on it courageously (indeed, fearlessly). From the formal point of view, the theory of truth that emerges is, as Maudlin observes, nothing other than the least-fixed-point theory of Saul Kripke.² From the philosophical point of view, however, the differences between Maudlin’s and Kripke’s theories are large. It is these differences that lead Maudlin to claim advantages that Kripke did not (and, I think, would not) claim for his own theory. Maudlin says that his theory demands no object-language/metalanguage distinction, that he has “developed a theory of truth for a language that can serve as its own metalanguage (191).” He promises early on (p. 4) that his theory will be “more adequate to our actual practice of reasoning about truth [than revision and other fixed-point theories].” And he claims that the language he has constructed is expressively complete (168).

The foundationalist picture that underlies these striking claims is as follows. Consider a formal language $L$ that has, apart from the usual logical resources (e.g., conjunction, negation, and quantifiers), a truth predicate $T$ for itself. Plainly, the truth values of the sentences of $L$ depend immediately upon the truth values of certain other sentences. Thus, the truth value of a conjunction ($A \& B$) depends immediately upon the truth values of the conjuncts $A$ and $B$, and the

---


A formal difference between Maudlin’s and Kripke’s constructions is that Maudlin’s language has both a truth and a falsity predicate. I shall neglect this difference below, since it has no substantive consequences for my argument.
truth value of $T(\text{“}C\text{”})$ depends on $C$. In Maudlin’s terminology, $A$ and $B$ are immediate semantic constituents of $(A \& B)$, and $C$ is an immediate semantic constituent of $T(\text{“}C\text{”})$. Picture the resulting semantic dependencies as a directed graph for $L$, one that contains an arrow from each immediate semantic constituent of a sentence $X$ to $X$. Sentences that have no immediate semantic constituents are the boundary sentences; they lie on the boundary of the graph. Let us assume that the boundary sentences behave classically, and so have exactly one of the truth values true and false. (This simplifying assumption is not required by Maudlin’s theory, and it is easily relaxed.) Maudlin expresses his core foundationalist idea thus:

(*)  [T]ruth and falsity are always ultimately rooted in the state of the world. That is: if a sentence is either true or false, then either it is a boundary sentence, made true or false by the world of non-semantic facts, or it is semantically connected to at least one boundary sentence, from which its truth value can be traced (49).

Starting at the boundary, one can calculate the truth values of certain sentences in the “interior” of the graph of the language [examples: $Gb \& \sim Gb$ and $T(\text{“}T(\text{“}Gb\text{”})\text{”})$, where $Gb$ is a boundary sentence]. Certain other sentences receive no truth value in the course of this calculation (examples: the Liar sentence that says of itself that it is not true, and the Truth Teller that says of itself that it is true). The truth values of the former sentences can be traced to the boundary; these are precisely the sentences that are grounded in Kripke’s sense. The Liar and the Truth Teller have no connection to the boundary, and they are ungrounded.

Kripke’s least-fixed-point construction embodies a foundationalist idea similar to that expressed in (*). However, Maudlin’s reading of (*) is strong and goes beyond anything found in Kripke. On Maudlin’s reading, (*) serves a role similar to the Vicious Circle Principle in Russell: it rules certain seemingly meaningful resources as illegitimate. The point can be appreciated through an example. After giving his least-fixed-point construction, Kripke notes that there is a sense in which the Liar sentence is not true, for it does not belong to the extension of truth under

3Maudlin gives a substitutional interpretation for first-order quantifiers. The immediate semantic constituents of a quantified sentence are its instances. Further, if $n$ is an individual term that denotes a sentence $A$, then $A$ is the sole immediate semantic constituent of $T(n)$.
the least fixed point. But this sense can be expressed only in a metalanguage. One may, for instance, express it by using a “strong” notion of truth, $T_s$, under which a truth predication $T_s('A')$ receives the value false iff $A$ receives a value distinct from true. However, this strong notion cannot be present in the object language; it must be relegated to a metalanguage. Maudlin regards the need for a metalanguage as a serious flaw in Kripke’s theory. His own response to the problem is to invoke (*) to rule out the strong notion of truth. If there were such a notion, then some sentences would receive a classical truth value even though they are not connected to the boundary (e.g., $T_s('L$')). Maudlin writes, “There can be only one truth predicate, and if it is predicated of an ungrounded sentence the result is an ungrounded sentence, for inescapable graph-theoretic reasons (52).” The absence of strong truth from Maudlin’s language can create the appearance that the language is expressively incomplete but, Maudlin holds, this appearance is due to an illusory notion. Similar arguments serve to rule out non-monotone connectives, and also other fixed points as possible interpretations of truth. Whereas Kripke highlights the multiplicity of fixed points and evaluation schemes, Maudlin cuts this multiplicity down by invoking (*).

There is a difficulty here, however. Maudlin says, on the basis of (*), that the Liar is not true (and also not false), for it is not connected to the boundary. But if we apply (*) to Maudlin’s own statement then we must rule that it too is not true (and not false), for it too is not connected to the boundary. Maudlin fearlessly accepts the result, and maintains that the notion of truth must be separated from, what he calls, permissible assertion. His statement about the Liar – indeed his whole theory – is not true but, he insists, it is permissible to assert. This, as he sees it, is the key to overcoming the object-language/metalanguage distinction. “To talk about semantics,” he writes, “one does not need to enlarge the language, one rather needs to relax one’s standards (154).” Maudlin counsels us that, when doing semantics in the presence of the Liar, we should turn into liars: we should say things that we know to be untrue.

II

Two of Maudlin’s claims can, I believe, be quickly set aside. The first is that he has “developed a theory of truth for a language that can serve as its own metalanguage (191).” The theory he offers (in the chapter titled “A Language That Can Express Its own Truth Theory”) is not a theory of
truth for the language in which it is expressed. The theory employs a special second-order quantifier but contains no explanation of it. And any attempt to extend the theory to this quantifier faces serious problems. Maudlin’s theory is at best a theory for a fragment of the language in which it is formulated. It does not establish the dispensability of the object-language/metalanguage distinction.

The second claim is that Maudlin’s theory is “more adequate to our actual practice of reasoning about truth [than revision and other fixed-point theories] (4).” Whatever other virtues Maudlin’s theory may have, a greater fidelity to our actual practice is not one of them. For example, on Maudlin’s theory, not only the Liar but also the Truth Teller (and even “This very sentence is either true or untrue”) imply outright contradictions. For another example, the theory rules that it is permissible to assert “every true sentence is true but ‘every true sentence is true’ is not true.” These are by any measure highly serious departures from our actual practice.

4The second-order quantifier ranges over functions from sentences to ordinals – call these, for the space of this note, rank functions. The theory Maudlin offers (on p. 85) contains no separate or explicit clause for this quantifier. If, however, the substitutional interpretation is extended to it, then, plainly, the quantifier is not ranging over all the rank functions and we are owed an account of its range. Moreover, we need a proof that the following sort of closure obtains: the rank functions quantified over suffice for the semantic analysis of the language quantifying over them. Otherwise, we have no assurance that the theory is “permissible” in Maudlin’s sense. Generally, attempts to construct semantically self-sufficient languages founder in their attempts to establish this kind of closure.

A small point: the base cases in the theory, as formulated on p. 85, seem to me to be ungrammatical. If Maudlin wishes to avoid Tarski’s procedure of explicitly listing all the base cases, perhaps he should use the device of quantifying into quotes, which is made perfectly clear by Nuel Belnap and Dorothy Grover. See their “Quantifying in and out of quotes,” in Hugues Leblanc (ed.), Truth, Syntax and Modality (Amsterdam: North Holland, 1973), pp. 17-47.

5This is so because the Truth Teller is never true in the least-fixed-point interpretation.

6Maudlin recognizes that some of the consequences of his theory are “extremely unintuitive (58).” I should also note that the implications stated above hold in the semantic system endorsed by Maudlin, not in the logical calculus he offers. Maudlin’s calculus seems to me to be exceedingly weak in power; even alphabetic variants seem not to be inter-derivable in it.

An elegant formalization of Kripke’s theory is provided by Michael Kremer in “Kripke and the Logic of Truth,” Journal of Philosophical Logic 17 (1988), 225-278. Kremer’s calculus is complete for the Strong Kleene fixed-point semantics when validity is defined using all fixed points (not, as in Maudlin, just the least fixed point). Kremer’s calculus is, however, sound for Maudlin’s semantics.

Maudlin’s Truth and Paradox – Page 4
no comparable departures in revision theories or in (plausible versions of) fixed-point theories. Maudlin would claim, I think, that the dispensability of the object-language/metalanguage distinction is one crucial point where his theory is more in accord with actual practice. But, as I have just noted, Maudlin has not established the dispensability of the distinction. Moreover, the dispensability is a highly theoretical idea, one that is debatable and that requires an extensive argument. It is not something that can be read-off from our actual use of the concept of truth.\(^7\)

Maudlin says that the interpretation of the truth predicate \(T\) is partial; it is given by the least fixed point \(<E, A>\), where the extension \(E\) and the anti-extension \(A\) do not exhaust the domain of sentences. The intuitive rules for truth \([C, \therefore T(‘C’); \text{and conversely}]\) are valid under this semantics, and the paradoxical argument is blocked because not all classical laws hold (in particular, \textit{reductio ad absurdum} fails). As Maudlin sees it, one important lesson of the Liar is that the logic of our language is not classical; not all classical laws are truth preserving.

Maudlin says also that the norms of assertion permit us to assert some sentences that are not true, and they render it impermissible to assert some sentences that are not false. That is, \(E\) is a proper subset of the set of permissible sentences \(P\), and \(A\) is a proper subset of the set of impermissible sentences \(I\). On the particular account of permissibility Maudlin offers, permissible sentences are those that are true under the closure of the least fixed point. That is, \(P\) is the set of sentences that are true, and \(I\) the set of sentences that are false, when \(E\) is made the classical interpretation of \(T\). Thus, \(P\) and \(I\) together exhaust the set of sentences \(S\): \(P \cup I = S\). Maudlin offers a calculus for reasoning with permissible sentences. This calculus is classical; if \(A\) is permissible and it classically implies \(B\) then \(B\) is permissible also. On Maudlin’s theory, then, permissibility-preserving inferences are not always truth preserving. Furthermore, truth-preserving inferences are not always permissibility preserving. \(T\)-Introduction – “\(C, \therefore T(‘C’)\)” – is an example of an inference that is truth preserving but not permissibility preserving by Maudlin’s theory.

----

\(^7\)In \textit{The Revision Theory of Truth} (Cambridge MA: MIT Press, 1993), Belnap and I argued for the separation of the project of constructing a good descriptive account of the concept of truth from that of constructing semantically self-sufficient languages. I have argued further for this in “Definition and revision: A response to McGee and Martin,” \textit{Philosophical Issues} 8 (1997), 419-443.
The separation of truth from permissibility is a key element in Maudlin’s theory. But there is reason to doubt that the separation is coherent. For, if the norms of assertion are as Maudlin states them to be, then, one wonders, why isn’t the interpretation of \( T \) classical? Why isn’t it simply the closure of the fixed point? Imagine we encounter a community that uses the predicate \( T \) in accordance with Maudlin’s norms. Members of the community assert \( Tb \) when and only when the denotation of \( b \) falls in \( E \); they deny \( Tb \), otherwise.\(^8\) As they “play the assertion game” they proceed perfectly classically. They freely invoke Excluded Middle, Conditional Proof, \textit{Reductio Ad Absurdum}, etc.\(^9\) For every sentence \( X \) in their language \( L \), they assert either \( X \) or the negation of \( X \), and they do so in perfect accordance with the norms of assertion. Even from an absolute, external viewpoint – we can imagine – there is no reason for the community to change its pattern of assertions and denials. If all this were to obtain, then it would appear that \( T \) is a perfectly classical predicate in the community’s language, one whose extension is \( E \). And if this is right then \( T \) does not really express “true in \( L \).” The set of sentences that are true in \( L \) is \( P \), and \( E \) is a proper subset of \( P \). So, \( T \) expresses at best a restricted truth predicate.\(^10\) Maudlin would have us imagine that \( T \) might be a partial predicate in \( L \), one with the interpretation \(<E, A>\). But how can this be when in the “assertion game” \( T \) behaves completely classically? How can the interpretation of a unary predicate be so divorced from idealized assertion and denial?

Maudlin’s core foundationalist idea, as it is expressed in the first part of (*), is undeniably attractive:

\[(*) \text{[T]ruth and falsity are always ultimately rooted in the state of the world.}\]

\(^8\)Imagine that the non-semantical part of the language is exceedingly simple. Alternatively, imagine that the community has access to an oracle.

\(^9\)Recall that Maudlin’s permissibility-preserving calculus is completely classical. The rule that is restricted here is \( T \)-Introduction.

\(^10\)We can allow the community to use Maudlin’s “truth-preserving” calculus. But the rules of this calculus, on the present interpretation, are \( T \)-preserving, not truth preserving. The calculus captures (partly) the logic of \( T \), not the logic of the language. \( T \) is somewhat like the provability predicate. A rule of inference, e.g., “\( A \), therefore \( \neg A \) is provable,” may be provability preserving, but may fail to be truth preserving. Similarly, for the rule of \( T \)-Introduction in Maudlin’s calculus.
However, the way that this idea is made precise in the second part of (*) gives rise to doubts:

\begin{quote}
\textbf{(*)B} \quad \text{If a sentence is either true or false, then either it is a boundary sentence, made true or false by the world of non-semantic facts, or it is semantically connected to at least one boundary sentence, from which its truth value can be traced (49).}
\end{quote}

Suppose a language \( L \) has arithmetical resources and can express such claims as (1) and (2).

\begin{enumerate}
\item \( \text{‘0’ is a meaningful expression of } L \).
\item \( \text{‘0’ is synonymous with ‘0’ in } L \).
\end{enumerate}

Can the truth and falsity of, respectively, (1) and (2) be traced – as required by (*B) – to the “world of non-semantic facts”? If (*B) is to be at all plausible, the world of non-semantic facts had better include facts to render (1) true and (2) false. If so, then there should also be facts to render the likes of (3) true.

\begin{enumerate}
\item \( \text{The Truth Teller has a cyclic graph in } L \).
\end{enumerate}

Claims not only about meanings but also about graphs are rooted in the world; language does not stand over against the world, it is a part of the world. Now consider (4):

\begin{enumerate}
\item \( \text{The Truth Teller is ungrounded in } L \).
\end{enumerate}

If the truth of (3) can be traced to the world of non-semantic facts, then the truth of (4) must also be traceable to this world, since (4) is equivalent to (3).\footnote{The definition of “grounded” should make “\( x \) is ungrounded in \( L \)” equivalent to \( (x \text{ is the Truth Teller and } x \text{ has a cyclic graph in } L) \) or (\( x \text{ is not the Truth Teller } \) and \ldots \).} But now (*B) loses its critical edge; it

11
cannot be used to rule out strong truth and other notions.\textsuperscript{12} Only on an implausibly narrow reading can (*B) serve the critical role that Maudlin assigns to it.

It is true that the presence of “ungrounded in $L$” in $L$ is liable to cause nasty paradoxes. However, these paradoxes are not inevitable. It can be shown that a Weak Kleene language retains the fixed-point property even if it is supplemented with a “neither-true-nor-false” operator (“$!$”).\textsuperscript{13} Speakers of this extended language can say truly that the Liar and the Truth Teller are neither true nor false (or, equivalently, ungrounded) without running the risk of paradox. There is nothing inherently problematic or ineffable about “$!$”. It is true that in the presence of certain other resources this concept is liable to generate paradoxes. But the same is true of other logical concepts. The conceptual situation here is that certain combinations of syntactic, semantic, and logical richness are paradox free, while others generate paradox. “$!$” leads to paradox if it is present with the Strong Kleene conjunction (“$\&$”), and certain semantic and syntactic resources. We can avoid paradox by removing “$!$” or “$\&$” from the language, or by restricting semantic or syntactic resources. But there are no \textit{a priori} logical or metaphysical principles that require us to adopt one set of restrictions over others, and it is folly to seek such principles. Indeed, I think there are no logical or metaphysical principles that dictate that we \textit{must} restrict resources. It is a remarkable fact that in our ordinary thinking we impose no restrictions. We somehow manage to work with resources that are rich – so rich that they generate paradoxes. To understand why paradoxes arise and how we manage to do conceptual work \textit{even in their presence} seems to me to be a worthy logical goal, one that is philosophically fruitful.\textsuperscript{14} 

\textsuperscript{12}Strong truth can be defined as “grounded and true.”

\textsuperscript{13}See §2E of \textit{Revision Theory}. The result here is closely connected to a theorem I proved jointly with Robert L. Martin.

\textsuperscript{14}Thanks to Nuel Belnap, Jiahong Guo, Katarzyna Kijania-Placek, and José Martínez-Fernández for helpful discussions on Maudlin’s book.

On Maudlin’s conception of its graph, (4) ungrounded. However, once this graph is accepted, one must reject (*B) as capturing the foundationalist intuition expressed in (*A). The truth of (4) is rooted in the world but, on Maudlin’s graph, it cannot be traced to the truth or falsity of boundary sentences.

Maudlin insists that “ungrounded” be treated as the third truth value. But this does nothing to render plausible the idea that (4) is ungrounded and thus untrue.