Maudlin's *Truth and Paradox*
Hartry Field

Tim Maudlin's *Truth and Paradox* is terrific. In some sense its solution to the paradoxes is familiar—the book advocates an extension of what's called the Kripke-Feferman theory (although the definition of validity it employs disguises this fact). Nonetheless, the perspective it casts on that solution is completely novel, and Maudlin uses this perspective to try to make the *prima facie* unattractive features of this solution seem palatable, indeed inescapable. Moreover, the book deals with many important issues that most writers on the paradoxes never deal with, including issues about the application of the Gödel theorems to powerful theories that include our theory of truth. The book includes intriguing excursions into general metaphysics, e.g. on the nature of logic, facts, vagueness, and much more; and it's lucid and lively, a pleasure to read. It will interest a wide range of philosophers.

1. A virtue of Maudlin's book is that it isn't concerned only with the *semantics* of languages with truth predicates, but also with their *inferential structure*. Let λ be a name of a sentence that asserts its own untruth: the sentence it names is ¬Tλ. ('T' abbreviates 'True'.) In the first chapter Maudlin considers the following argument (where <A> is the name of sentence A):

1. Tλ is T<¬Tλ>, so by "Downward T-inference" (the rule that allows the inference from T<A> to A) it implies ¬Tλ; i.e., Tλ implies its own negation.

2. So by classical reductio reasoning (the rule that when A implies its own negation, we can conclude ¬A), we conclude ¬Tλ.

3. But then by "Upward T-inference" (the rule that allows the inference from A to T<A>), we conclude T<¬Tλ>, i.e. Tλ.

The conclusion at stage 3 contradicts that at stage 2.

We have three choices: we can accept the contradiction, in a non-classical logic that blocks the derivation of arbitrary sentences from contradictions; we can disallow one or both T-inferences; or we can disallow reductio reasoning. Maudlin barely considers the first option,[1] and I won't either; this leaves the options of restricting the T-inferences and restricting reductio.

The reductio rule can be viewed as a product of two more basic rules: reasoning by cases (the rule that from 'A implies C' and 'B implies C' we can infer 'A∨B implies C') and the law of excluded middle. For suppose A implies ¬A. ¬A implies ¬A too, so using reasoning by cases, A∨¬A implies ¬A; but by excluded middle, A∨¬A is valid, so ¬A must be too. Many theorists (Gupta 1982, McGee 1990) see reasoning by cases as the culprit. Maudlin remarks that disallowing reasoning by cases makes disjunctions "inferentially impotent" (111), and I agree that that's objectionable; if so, keeping the T-inferences is best achieved by giving up excluded middle along with reductio reasoning. As Maudlin notes, if we keep both T-inferences we will also need to give up some other classical rules as well, such as the rule of conditional proof, because of Löb's (*aka* Curry's) Paradox.

Maudlin bills his theory as one on which the T-Inferences are valid and classical rules like reductio and conditional proof and excluded middle aren't. This is accurate *given what he means by 'valid'*, but it had me thoroughly misled about his theory through the early chapters. I think it would be more natural to describe his theory the other way around, as one where the classical rules are valid and Upward T-Inference isn't.

To explain this, I'll first describe an alternative type of solution (which I'll call a *Full T-Inference solution*) that unproblematically counts as one on which the T-inferences are valid and the classical rules just mentioned aren't. (Like Maudlin's, it doesn't allow accepting contradictions but does accept reasoning by cases). On a Full T-inference solution, the T-inferences are *inferentially valid*, in the sense that believing A commits you to believing that A is true (i.e. believing T<A>) and vice versa. (On *typical* Full T-inference solutions, believing ¬A also commits you to believing ¬T<A>, and vice versa. Indeed on many such solutions, we have full intersubstitutivity between T<A> and A: if X is any sentence in which A occurs only extensionally, and X* results from X by substituting T<A> for one or more occurrences of A, then one is committed to having the same attitude toward X and X*.) By the argument above, any Full T-inference solution that doesn't allow belief in contradictions and retains reasoning by cases cannot accept excluded middle: for it cannot accept that *either the Liar sentence is true or it isn't true*.[2]

Maudlin, by contrast, does believe that either the Liar sentence is true or it isn't true. Indeed, he believes the latter, i.e. he believes ¬T<¬Tλ>. But this is just ¬Tλ, so he simultaneously believes ¬Tλ and ¬T<¬Tλ>; i.e. *he believes ¬Tλ while believing that this belief of his is not true*. (He devotes considerable energy to trying to defuse the prima facie implausibility of this: more on that later.) Moreover, since he disallows accepting contradictions, he does not *also* believe that his belief *is* true; i.e. *he believes that ¬Tλ while not believing that <¬Tλ> is true*. So Upward T-Inference is not *inferentially* valid. When he calls it valid, what he means is that it's *semantically* valid in the sense of being (necessarily) truth-preserving. On his account we can have semantic validity without inferential validity. The reason: on that account we must believe certain sentences that we don't believe to be true (and indeed, believe *untrue*); when these believed sentences are the premises of a truth-preserving argument, there is no reason to believe the conclusion of the argument true, or to believe the conclusion itself.[3]

If we regard the T-inferences as inferentially valid, as in the naive theory of truth and in Full T-inference solutions to the paradoxes, then there is no real difference between

(i) the inferential validity of the inference from A to B and

(ii) the inferential validity of the inference from T<A> to T<B>.

But on Maudlin's theory, the fact that T<A> and T<B> aren't inferentially equivalent to A and B opens up the possibility of a difference between (i) and (ii); and the difference, for Maudlin, is huge. The semantic validity of the inference from A to B is equivalent in his theory to (ii), not to (i). So when Maudlin says that the Upward T-inference is "valid", what this amounts to is that for any A, if you accept T<A> then you should accept T<T<A>>; it doesn't mean that if you accept A then you should accept T<A>.

The upshot: while Maudlin's solution to the inferential Liar paradox has it that the T-inferences are semantically valid but classical rules like reductio and excluded middle aren't, it also has it that those classical rules are inferentially valid while upward T-inference isn't. (Downward T-inference is inferentially valid on Maudlin's view, though this doesn't follow from its semantic validity.) When Maudlin's theory is put in terms of inferential validity, it turns out to be (an extension of) what's often called the Kripke-Feferman theory (KF): a classical logic theory that allows Downward T-Inference but not Upward T-Inference (but contains the weaker rule that T<A> implies T<T<A>>).[4]

Turning from exposition, the substantive question I raise at this point is whether we shouldn't regard Upward T-Inference as inferentially valid rather than merely semantically valid. Some brief comments:

1. The one disappointing feature of the book is that it doesn't even mention any Full T-inference solutions: a discussion of why he thinks his solution preferable to them would have been extremely helpful.
2. Maudlin's initial motivation for the Upward and Downward T-Inference rules (pp. 9-10) seems based on the thought that T<A> is little more than a notational variant of A. But if that's so, it supports not only that the inference from the latter to the former must preserve truth, but also that we ought to have the same attitudes (belief, disbelief, etc.) toward each; i.e. it motivates the inferential validity of the rules, not just their semantic validity.
3. Denying the inferential validity of the Upward T-Inference cripples the ordinary use of 'true'. Suppose Jones disagrees with the overall body of theory that Maudlin espoused on Monday, but can't decide which specific claims of Maudlin's are faulty (or can't remember exactly what Maudlin said that was wrong). The standard story about the truth predicate (Quine 1970, Leeds 1978) is that it is precisely in circumstances like these that 'true' is useful: Jones can express his disagreement by saying "Something Maudlin said on Monday isn't true". But this loses its effect unless both Upward and Downward T-Inference (and related rules involving negation) are in force: indeed, since Maudlin himself thinks that important parts of his own theory aren't true, it is clear that Jones hasn't succeeded in expressing disagreement.[5]

But let's turn to Maudlin's views on the metaphysics of truth, which supply an intriguing and novel justification for rejecting the inferential validity of the Upward T-Inference.

2. Maudlin justifies his approach to the paradoxes by means of a fairly appealing picture of how semantics works. In this picture, truth and falsity are primarily properties only of *boundary sentences*: (parameterized) atomic sentences with predicates other than 'True'. (A *parameterized* atomic sentence is an atomic formula with objects assigned to its free variables.) And if we ignore vagueness (as Maudlin mostly does, saying it has nothing to do with the paradoxes), then all boundary sentences are either true or false. Other sentences can inherit truth or falsity from the boundary sentences, but this can happen only if these other sentences are *grounded* in a sense equivalent to Kripke's (i.e., members of the minimal Kleene-based fixed point).[6] So we get a threefold partition: the true, the false, and the ungrounded.

The claim that ungrounded sentences such as the Liar aren't true or false is not a claim that appears in the minimal fixed point. But it is part of Maudlin's theory, albeit a part of the theory that the theory itself asserts is untrue. A Full T-Inference theory, by contrast, contains only those sentences that appear in the fixed point (at least, this is so for sentences with no '6'). Such a theory thus agrees with Maudlin's in holding that you shouldn't assert that ungrounded sentences *are* true or false, but it says that you also shouldn't assert that they *aren't* true or false either (and also shouldn't assert that *either they are or they aren't*, so that the inability to answer either way is not a matter of *ignorance*). I don't see that Maudlin offers any argument for declaring that ungrounded sentences aren't true or false; he simply offers a *picture* on which they aren't (though admittedly one with some prima facie appeal).

Another important feature of Maudlin's theory is his account of what constitutes a legitimate logical connective. Maudlin takes connectives to be defined by their semantics (a wrong approach, I believe, but there's no space to pursue this[7]); but, motivated by his discussion of groundedness, he prohibits the "definition" of a connective from making a complex sentence true or false when this is not due to the truth and falsity of some of its simpler parts (49). As he illustrates, this excludes many connectives which, if added to the Kleene language, would lead to paradox. However, it also excludes many connectives that would *not* lead to paradox and that seem extremely natural. For instance, Kleene logic contains no serious conditional: e.g., no connective 6 for which all sentences of form A6A or AvB6A or T<A>6A or A6T<A> are true. Because of this, one cannot truly say in a theory whose logic is Kleene's that whenever a conjunction is true then so are the conjuncts, or that Downward or Upward T-Inference preserves truth. (The claim that they preserve truth is part of Maudlin's theory, but it is one of those parts that the theory declares untrue.) There are a number of ways of adding to Kleene logic a connective that validates these laws (and many other natural laws too), without engendering paradox (see note 2); on Maudlin's account, such connectives are automatically illegitimate, because a legitimate connective can't possibly be such that A6A is true when A is ungrounded. This position seems to me very costly, and I'd like to have seen a more serious argument for it.

Maudlin's position on this does have an interesting consequence: on his view, no *sentence* can be semantically valid, that is, there can be no semantically valid argument to it without undischarged premises. Equivalently, the view is that no sentence *of form T<A>* could be inferentially valid (there can be no inferentially valid argument to a sentence *of form T<A>* without undischarged premises). "Logic alone cannot establish the truth of any sentence" (117); the doctrine that it could is "mistaken, misleading and pernicious" (74), and it's as puzzling how logic could do that as how "pure reason" could give us knowledge of the structure of space. This is obviously a claim of great philosophical interest. But I'm skeptical: he grants that our inferential system could reasonably lead to conclusions of form Aw¬A, and he thinks truth is a logical notion, so it isn't obviously ridiculous to suppose that our inferential system lead to conclusions of form T(<Aw¬A>). Again, Maudlin does offer a picture on which the presence of truth in the latter conclusion makes premise-free inference to it much worse, but there isn't either a real argument for the picture or a serious assessment of its costs.

3. As emphasized, Maudlin's account has it that many claims that aren't true are nevertheless correct to believe—this includes not just the Liar sentence, but also central generalizations of Maudlin's theory, such as that if a conjunction is true then so are its conjuncts and that the T-Inferences are truth-preserving. It seems then that there is an important property in addition to truth, the property of being *correct to believe*. Maudlin accepts this, and calls it *permissibility*. He argues that it's very different from truth: unlike truth, permissibility is norm-dependent, and non-objective in that different legitimate norms are possible. It leads to its own paradoxes—e.g. sentences that assert their own impermissibility. But these are handled very differently from the paradoxes of truth—they are handled in a way somewhat reminiscent of the

Tarskian levels solution.[8] Maudlin justifies this by his claim that permissibility, unlike truth, is not objective, because dependent on norms of which none is uniquely best.

But it's unobvious that permissibility in his sense *should* be viewed as less objective than truth. Maudlin makes the non-objectivity of permissibility seem more obvious than it ought to seem by sometimes talking of permissibility as based on norms for what it's permissible to *assert* (e.g. 165, especially discussion of the Monastic Rule of Silence); but it's clear that he really needs to think of the norms as governing *belief*, e.g. as licensing the beliefs that the Liar sentence isn't true and that the T-Inferences preserve truth. Still, Maudlin could plausibly argue that rules governing beliefs, such as inductive rules, also lack objectivity, since there are multiple reasonable policies one might have for extrapolating from the observed to the unobserved. But that isn't enough for Maudlin's point. The rules for belief that Maudlin is concerned with aren't like inductive rules, for they don't take into account our epistemological situation, e.g. which evidence is available: they are rules governing a being who knows all the truths (but chooses to go beyond them in forming his beliefs). Indeed, Maudlin regards permissibility as completely objective when it comes to sentences that assert truth: T<A> is permissible when and only when A is true, and that is an entirely objective matter. It is only for other kinds of sentences that legitimate "norms of permissibility" can diverge from each other. Moreover, it's clear that Maudlin thinks all good norms agree *even about many things he regards as untrue*: for instance, he clearly thinks that all good norms agree in licensing belief in the untruth of the Liar, in the truth-preservingness of the T-Inferences, and in many other of the claims that Maudlin vigorously asserts in his book but regards as untrue.

In short, Maudlin's notion of permissibility is not nearly as dependent on arbitrary norms as his discussion suggests. Indeed, his notion of permissibility seems at least as much like the ordinary notion of truth as does his notion of truth. (For instance, in the ordinary sense of 'true', 'Not everything he said is true' expresses disagreement; in Maudlin's theory neither 'Not everything he said is true' nor 'Not everything he said is permissible' does that, but the latter comes somewhat closer to doing so.)

It seems best to adopt a view on which truth and permissibility don't diverge—a view on which one can assert the truth of A when and only when one can assert A. The Full T-Inference views mentioned earlier have this feature; indeed, the best of them allow for the intersubstitutivity of T<A> with A in all extensional contexts, and for a conditional which validates the Tarski conditionals T<A>6A and A6T<A>. Despite the ingenuity of Maudlin's discussion, I see little to favor his view over a view of that kind. Still, for anyone interested in the paradoxes this is a must-read book.

References

Brady 1989, "The Non-Triviality of Dialectical Set Theory". In Priest, Routley and Norman (eds.), *Paraconsistent Logic*. Philosophia Verlag.

Field 2003, "A Revenge-Immune Solution to the Semantic Paradoxes", *Journal of Philosophical Logic.*

Gupta 1982, "Truth and Paradox", *Journal of Philosophical Logic*.

Kripke 1975, "Outline of a Theory of Truth", *Journal of Philosophy*.

Leeds 1978, "Theories of Reference and Truth", *Erkenntnis*.

McGee 1990, *Truth, Vagueness and Paradox*. Hackett.

Quine 1970, *Philosophy of Logic*. Prentice-Hall.

Reinhardt 1986, "Some Remarks on Extending and Interpreting Theories with a Partial Predicate for Truth", *Journal of Philosophical Logic*.

Soames 1999, *Understanding Truth*. Oxford.

Yablo 2003, "New Grounds for Naive Truth Theory". In Beall (ed.), *Liars and Heaps*. Oxford.

Notes

1. He does note that it would need to be developed so as to deal with the Löb-Curry paradox, and that it's unobvious how best to do so.

2. On *one* way of reading Kripke's (1975) fixed point solutions based on the Kleene semantics, they are Full T-Inference solutions. (This reading is advocated in Soames 1999, though Soames' description of the theory as positing truth value gaps seems incorrect. See also the theory KFS in Reinhardt 1986.) But "Kripkean" Full T-Inference solutions aren't very satisfactory, since Kleene semantics doesn't allow for a serious conditional connective that would allow natural reasoning and the assertion of the biconditionals T<A> : A. But the Kleene logics can be extended to allow for a serious conditional in several different ways that permit that permit Full T-inference solutions: see for instance Field 2003 and Yablo 2003. And full T-inference solutions in weaker logics with serious conditionals date back to Brady 1989.

3. Incidentally, one of the things that can never be true, on Maudlin's theory, is the claim that Upward T-Inference is truth preserving, since that claim is a universal generalization over sentences some of which are "ungrounded". (As he shows with admirable clarity near the end of Chapter 2, there is no getting around such problems by appeal to restricted quantification.) On his account, Upward T-Inference is truth-preserving, but it isn't true that it is—an idea that those of us committed to Full T-Inference views will find hard to swallow.

4. Reinhardt 1986 distinguishes KF from the Full T-Inference theory KFS mentioned in note 2. How can two such different theories both be readings of Kripke 1975? Part of the answer is that Kripke focused on semantics rather than inferential rules, and there are different ways of getting the inferential rules from the semantics. If one takes the *acceptable inferences* to be those under which the minimal fixed point is closed, one gets a Full T-inference solution; if one takes the acceptable inferences to be the classical ones but the positive extension of the truth predicate to be the minimal fixed point, one gets KF.

5. One might try to solve the problem by bringing in a primitive notion of disagreement, but that wouldn't work for related problems in which 'true' is more deeply embedded, e.g. 'If everything Maudlin said on Monday is true then ...'.

6. It's the groundedness intuition that's crucial for Maudlin: fixed points other than the minimal Kleene-based one play no role in his account, unlike Kripke's.

7. I'll just say that truth-table "definitions" get bite only from the logic one uses to apply them; Maudlin implicitly assumes that one can use excluded middle for this (e.g. by assuming that the rows of a truth table are exhaustive), which begs important questions about how the "defined" connectives behave.

8. The view is that a sentence asserting its own impermissibility according to norms N is true, and that it's impermissible according to N but permissible according to other norms N*. Maudlin argues that unlike the levels solution, such N* shouldn't be viewed as *higher* norms, but simply as *different* norms.