

Prosentence, Revision, Truth, and Paradox

NUEL BELNAP

University of Pittsburgh

Tim Maudlin's *Truth and Paradox* (Maudlin 2004, cited here as T&P), a book that is richly endowed with interesting analyses and original theses, chooses to ignore both the prosentential theory of truth from Grover, Camp and Belnap 1975 and the revision theory in its book form, Gupta and Belnap 1993 (*The Revision Theory of Truth*, henceforth RTT).¹ There is no discussion of either theory, nor even any mention of them in the list of references. I offer a pair of quotes chosen from among a number of T&P generalizations that Maudlin would doubtless have modified if RTT had been on his mind at the time of composition of T&P. (1) "...every acceptable account of truth seems to imply that the TInferences must be valid" (p. 15). My response is that the revision theory of truth is built on an explicit denial of this. Rather than taking them as "valid," RTT takes the T-Inferences as stage-of-revision-shifting revision principles in the context of a definitional account of truth. (2) "...most discussions of the Liar paradox and related paradoxes...do not address [such questions as]...where [T&P's] Proof Lambda and Proof Gamma go wrong" (p. 20). In fact, RTT is not open to this criticism. It's simple natural-deduction calculus C_0 addresses exactly such questions.

1 Tarski's strategy as normative

Now, given the thousands of pages on truth, that T&P omits discussion of two of my favorite theories is thoroughly understandable. I wish, however, to restore some balance by, instead of directly discussing various parts of T&P, taking most of my time to discuss the impact that *prosentence* and *revision* might and I think should have had on T&P. I begin, however, with a small defense of Tarski (as if he needed it). The background is the observation that although much philosophical logic is offered with no sharply defined purpose, some studies are intended as definitely *descriptive* and others as definitely *normative*. Tarski's Convention T speaks to the *descriptive* side of

¹ T&P indeed cites the 1984 reprinting of the earlier essay, Gupta 1982.

his enterprise, in which role it has been roundly criticized, including complaints by both T&P and RTT. Rather than throw out a marvelous theory, however, I think that we should adopt "Tarski's strategy" as a *normative* guide. As a crude characterization of *Tarski's strategy*, I mean that when you wish to do serious work on the use or meaning of a semantic predicate, for goodness sakes avoid situations in which exactly the same semantic vocabulary is both *mentioned* as an object of your study and is *used* by you in carrying out your investigations. You will save yourself (and your readers) endless confusion.

2 The prosentential theory as descriptive

Whereas I endorse adopting the Tarski strategy as explicitly normative advice, I endorse the prosentential theory for its descriptive power. T&P does not mention the prosentential theory at all, unless Maudlin means indirectly to include it when discussing theories that claim that "*n* is true" is a notational variant of the sentence denoted by *n*. T&P is surely right in saying that such a claim endorses both Tarski's T-biconditionals and the T-inferences that take us between "*n* is true" and the sentence denoted by *n* (p. 9). That is, however, a misdescription of what I take to be the heart of the prosentential theory, namely, *in ordinary ordinary language, such as we use in conversation, there is no serious T-predicate, and hence there are neither T-biconditionals nor TInferences.*²

If Penelope says, "The seat of all learning is in the coccyx," and then Kit says "That's true," it's not a bad theory to suggest that Kit's utterance serves not to *characterize* what Penelope said. Instead, Kit's use of "That's true" is as a prosentence that makes *anaphoric* reference to what Penelope said. The prosentence gives Kit a way of asserting what Penelope said, while giving her the credit. I don't see how *this* thought is improved by tacking on a theory of ungrounded sentences, nor is there a reason to think of Kit as speaking in a metalanguage, nor even as using a serious truth predicate.

3 T&P as descriptive and normative

So I endorse Tarski's strategy as good advice, and the prosentential theory as good description. What about the aims of T&P? As I understand it, T&P divides the theory of truth into two parts: The semantic part is descriptive, while the proof-theoretical part is normative. Here is an unambiguous statement of T&P's descriptive intent.

...what is on offer is a complete analysis of truth in a natural language, an analysis both of the truth predicate and of truth itself, in so far as it admits a general analysis. (p. 177)

The most conspicuous features of T&P's semantic description are these.

² I express a point of view that is not necessarily shared by Grover and Camp.

- The subject language must carry truth linguistically as a *predicate of sentences*.
- The language must be (at least) three-valued, including not only the values **T** and **F**, but also the value **U**. In T&P, “ungrounded” doesn’t mean merely “neither true nor false”; rather, T&P’s innovation is that **U** is a value reserved for those sentences for which an inductive definition of truth—that is, a definition that proceeds from the values of subformulas to the value of the compound—would be circular. Take a truth predication

t is true.

with, for this example, *t* an atomic term. Suppose *t* denotes a complex sentence *S*. Then we must first have a semantic value for *S* before we can obtain a semantic value for “*t* is true.” Usually there is no problem, but suppose *S* has as a subformula the very sentence “*t* is true.” Then we must have a value for “*t* is true” before we can obtain a value for *S*. Voilà, instantly there is a circle calling for the third value **U**. (This is a good place to note that RTT does not invoke a third value of any kind in building a truth theory for a given two-valued “ground language,” as RTT calls the truth-free part. In other words, circle or not, adding a truth predicate in no way disturbs the ground language.)

- Standard connectives and quantifiers must be given a truth-functional reading in the three values, **T**, **F**, and **U**.
- And most striking, according to T&P, no language has any connective whose semantics would lead to *troublesome* semantic paradox by exploiting the threat of circularity. Here I am using “troublesome” loosely, as signifying something like “If Penelope utters the Liar in an assertive tone of voice, then, so long as our best logicians are restricted to the ground language plus the truth predicate, *they would not have even the most nebulous idea of what to say* that would serve as a suitable semantic characterization of what Penelope asserted. Since in T&P the Liar and “All truths are true,” for instance, simply take the value **U**, T&P aspires to put to rest all logical anxiety, for there is certainly no paradox in saying that a certain sentence takes the value **U**.

T&P does, however, worry a bit the possibility of “revenge” in the form of a language containing both the standard self-applying truth predicate together with Strong Negation ($\sim\mathbf{U}=\mathbf{F}$ instead of $\sim\mathbf{U}=\mathbf{U}$). In

the hands of most students of truth riddles, a three-valued self-referential language with truth and Strong Negation is unavoidably saddled with the so-called Strengthened Liar Paradox, a semantic cousin of the Liar using Strong Negation instead of plain three-valued negation. T&P argues, however, that there is no such thing as Strong Negation. To imagine the existence of such a language is to indulge in an illusion (see p. 34, where Strong Negation is labeled “incoherent.”) Thus, T&P “solves” (as in T&P’s subtitle) or “resolves” (p. 48) or “dissolves” (p. 49) all semantic paradoxes by concluding that there can’t be any such languages, and therefore there can’t be any revengeful semantic paradoxes. At this point it seems to me that T&P has switched to its own artificial concept of truth, and in the process has become unfaithful to our concept of truth. In contrast, RTT says that our concept of truth is exactly what gives rise to troublesome semantic paradoxes—not merely to some sentences that are neither true nor false. RTT defends this by taking as the ground language, not some fragment of our sloppily characterized ordinary language, but a formally described language that is unarguably two-valued.

- That discussion was semantic. T&P, with careful subtlety, does allow for *inferential* paradoxes. For example, as we noted above, speaking *semantically*, in T&P, the Liar and a “logical” generalization such as “All truths are true” have exactly the same status: Each is semantically ungrounded. There is, however, a set of proof-rules more or less approved by T&P, and dignified by the rubric, “logical consequence” (p. 155), that can distinguish the “acceptability” of “All truths are true,” and “The Liar is true,” agreeing with intuition by rating the first acceptable and the latter not. The trouble is that “logical consequence” and “acceptability,” like all inferential concepts not determined by semantics, can be as variable (T&P says “fickle”) as the weather in Pittsburgh. In other words, reliance on the notion of “logical consequence” as defined in T&P is apparently, according to T&P itself (pp. 173-175), a strictly optional matter. If this is so, it would seem that T&P is not in a position to defend its concept of “logical consequence” as one that can guide us in our reasoning.

(In contrast, the revision theory distinguishes “The Liar is true” and “All truths are true” *semantically*, opposing them as a semantic paradox vs. a semantic logical truth. I note, not so incidentally, that these judgments are not made on intuitive grounds, nor are they special to truth, but flow directly out of the revision theory of circular concepts, as indicated below.)

I do not mean to give the impression that T&P just states the no-paradox thesis without warrant; indeed, one of the most interesting contributions of T&P is the argument that the language must be three-valued, with T, F, and U. It is a subtle argument, and one might want to quarrel with it; but here I only wish to emphasize T&P's conclusion: There are *no troublesome semantic paradoxes* surrounding truth. Adding U as a third value, concludes T&P, removes all the semantic trouble. The contrast between T&P and RTT on this descriptive point is absolute. RTT urges that the Liar is in fact paradoxical in a *troublesome* sense (it confounds our best logicians). One does not understand our concept of truth if one does not see both that such sentences as the Liar, the Truth Teller, Curry's paradox, and so forth are pathological, that the pathology arises from the very nature of truth, and that such a thesis does not at all impugn either ourselves or the concept of truth. RTT argues that you should reject any *descriptive theory* that tries to "solve" the paradoxes by removing them from the barrel of permanently troublesome problems.

4 The centrality of circularity

Often when folks think of Gupta's work on truth they fasten on his early articles, and more or less ignore the later book, RTT. Maudlin's T&P is a striking example of what seems to me a thoroughly understandable but sad mistake. The reason for this judgment is that Gupta's early articles tended to emphasize the details of the transinitely defined revision process, including the delicate matter of how, technically, one handles the process at limit ordinals. This approach to Gupta treats his work, like that of Herzberger, as a minor (if interesting) variant of Kripke's ideas. T&P takes this line. In my judgment, however, the single most significant output of Gupta's work is the thesis, stated already in the first paragraph of RTT, that "truth is a circular concept." To my doubtless thoroughly biased mind, the observation that truth is circular stands alone in the post-Tarski era as near-equal in importance to Convention T.³ One does not need the Liar or the Truth Teller to be struck by the often *non*-pathological circularity of truth. One simply has to consider an absolutely harmless T-biconditional such as

(The sentence) " 'Fred smokes' is true" is (itself) true (that was the *definiendum*) if and only if 'Fred smokes' is true (and that was the *definiens*).

You need to take this as—in Tarski's words—a partial definition of the truth predicate in order to note that "is true" appears both on the left as part of the

³ So as not to mislead, I note that RTT rejects that portion of Convention T which requires that the definition of truth should imply all the T-biconditionals (p. 29 of RTT), while at the same time endorsing the claim that the T-biconditionals, *when read as partial definitions*, fix the signification of truth.

definiendum and on the right as part of the *definiens*. Nothing could be more obviously circular than that.

Accordingly, RTT is motivated to give a *universally applicable* account of circular definitions. For this one cannot look to the classical tradition, which, beginning long before the modern story about definitions was first told by Frege and the Polish logicians, has categorized circular definitions as nonsense lying beyond the pale. These considerations drive (or should drive) one to a *non-classical* theory of definitions, which is exactly what RTT supplies.⁴ What is striking is that *nothing else needs to be non-classical* in explaining how circular concepts work. The logic of the application of RTT to a classical two-valued language is itself classical and two-valued. Nor is there anything in the RTT account of circular definitions that *requires* classical logic; RTT makes sense of circular definitions against the background of any logic whatsoever. No wonder that RTT is in a position to suggest—and argue—that as a consequence, you cannot understand the behavior of the truth predicate without careful consideration of those features of its signification that derive from its circularity. Partly what I am emphasizing is that paradox and the like is not to be blamed on self-reference, except to the extent that self-reference leads to definitional circularity. RTT gave evidence for that when it demonstrated how easy it is to cook up truth-like paradoxes and pathologies from circularity without even a hint of self-reference.

Suppose that's right. What must a competing theory of truth do to meet the challenge? It seems to me that just about all the post-Tarski theories of truth self-indulgently call into play the awesome metaphysical nature of truth, or at least involve some heavy breathing. A small example from T&P: "If the Truth Teller could possibly be either true or false, where could either of those truth values have come *from*?" (p. 50). To compete with the revision theory, however, which involves no heavy breathing, you must be prepared to give an account of *any* concept based on a circular definition to the extent that its behavior rests on that circularity. This is likely to be a challenge—and a healthy one—for T&P. The only way out for T&P that I can see is to deal directly with the RTT thesis that *both* the ordinary *and* the peculiar behavior of the truth predicate arise from the circularity of its definition.

5 Paradox everywhere

I want to extend the claim that circularity is central. The idea is that one does not obtain a good theory of truth by treating it in philosophical isolation.

⁴ Here "non-classical" does not refer to a non-classical logic (many-valued, intuitionistic, and so forth). I simply mean an account of (formal) definitions that, because it finds room for circular definitions, steps outside the boundaries laid down by Frege and the Polish logicians.

For example, circular definitions do not need to be only of predicates. Any part of speech can carry a circular concept. Here is an easy one concerning denotation and arithmetic.

Let δ symbolize the denotation function, and let us baptize the symbol “ $\delta a + 1$ ”, using the letter “ a ” for its name; that is, we declare that $a = \delta a + 1$.

So by applying the denotation-function, δ , to both sides, we readily obtain $\delta(a) = \delta(\delta a + 1)$. But $\delta(\delta a + 1) = \delta a + 1$ (because δ is denotation), so $\delta(a) = \delta a + 1$.

As a result, there is a paradox. Ordinary Peano arithmetic says that the adding-one function has no fixed point: $n + 1 \neq n$, all n . But this stands in contradiction to what we have just shown, that $\delta(a) + 1 = \delta(a)$, i.e., that $\delta(a)$ is a fixed point of adding one. Call this the Paradox of the Adder.”

We may “solve” the Adder paradox in the Tarski way, by classifying it as due to bad grammar ($a = \delta a + 1$ both uses “ a ” on the left, and mentions it on the right). We may also “solve” the Adder in analogy to the three-valued “solution” to the Liar paradox. That is, we may let \mathbf{u} be “the ungrounded number,” following Kleene arithmetic by declaring that \mathbf{u} is a fixed point for the adding-one function: $\mathbf{u} + 1 = \mathbf{u}$. Then we can let the denotation of a be \mathbf{u} , which “solves” the paradox. So the Adder presents us with a three-way dilemma: if we stick with ordinary arithmetic with the paradox unsolved, we have a contradiction in semantics. If we use the Tarski “solution,” we are likely starting an endless climb up a hierarchy. If we use the T&P “solution” of the Liar paradox to model a solution to the Adder paradox, we are no longer talking of ordinary arithmetic. The bite of the Adder is painful if not deadly. I

in any case, I trust that you see that the Adder is exactly the same as the Liar, *mutatis mutandis*: Given the Liar and no hierarchy, either negation has a fixed point, in which case we are not doing (ordinary two-valued) semantics, or else, if negation has no fixed point, we have a contradiction in our semantic theory.

It is evident how the revision theory handles a language-fragment in which the paradox of the Adder arises. The same old doctrine of circular definitions works *without further tinkering* to put the Adder back into its cage. It is other with T&P, which is concerned with the paradoxes of truth as an isolated phenomenon. If I am right Maudlin must face up to the venomous Adder as well—in some way, of course, that succeeds in avoiding its fangs.

Let me confess at once that “must” is much too strong. There is no doubt that Maudlin works with great profit in many areas, and that there is no intellectual *necessity* that he think about the Adder. Still, there is a point there,

namely, that the general-purpose revision theory naturally informs a great many, if not all, circular concepts. In the last chapter of RTT the following are among those listed: standard semantic concepts such as reference and satisfaction; setmembership and identity of sets; modal, doxastic, and epistemological notions such as necessity, belief, and knowledge; and value notions such as “best move” in a game with two or more players.⁵

Finally, let me go back to the beginning. I say again that Maudlin’s T&P is chock full of truly interesting theses, analyses, perspectives, and arguments, hardly any of which have here been surveyed. Instead I have said at least as much about the prosentential and revision theories of truth, and how they interact with or challenge the truth theory of T&P. My justification has been that Maudlin (understandably) chose not to mention these theories that I think so important, and that I think you should think important, and each of which has something to say about key parts of T&P.

References

- Chapuis, A. 2000, “Rationality and circularity”, in Chapuis and Gupta 2000, pp. 49-78.
- Chapuis, A. and Gupta, A., eds 2000, *Circularity, definition and truth*, Indian Council of Philosophical Research.
- Grover, D. 1992, *A prosentential theory of truth*, Princeton University Press, Princeton.
- Grover, D., Camp, J. and Belnap, N. 1975, “A prosentential theory of truth”, *Philosophical studies* 27, 73-125. Reprinted in Grover 1992.
- Gupta, A. 1982, “Truth and paradox”, *Journal of philosophical logic* 11, 1-60. Reprinted in R. L. Martin (ed.), *Recent essays on truth and the Liar Paradox*, Oxford University Press, Oxford, 1984, pp. 175-235.
- Gupta, A. 2000, “On circular concepts”, in Chapuis and Gupta 2000, pp. 123-153.
- Gupta, A. and Belnap, N. 1993, *The revision theory of truth*, MIT Press, Cambridge, MA.
- Maudlin, T. 2004, *Truth and paradox: solving the riddles*, Clarendon Press, Oxford.

⁵ This last entry is not listed in RTT. Only later did Gupta 2000 and Chapuis 2000 make their brilliant applications of revision theory to decision problems.