

## **Do the Paradoxes Pose a Special Problem for Deflationism?**

Anil Gupta

University of Pittsburgh

The Liar and other semantic paradoxes pose a difficult problem for all theories of truth. Any theory that aims to improve our understanding of the concept of truth must, when fully stated, include an account of the paradoxes. Not only deflationism but also its competitors – for instance, correspondence and coherence – must ultimately address the paradoxes. The question that concerns me in this essay is whether it is especially urgent for deflationism to do so. Are the paradoxes a special threat, a special problem, for deflationism? I will argue that they are not.<sup>1</sup> Deflationists can leave the paradoxes to the specialists to puzzle over. It is the specialists who will be well served if they keep some insights of deflationism firmly in view.

### **I**

Deflationism rests on some claims about our ordinary concept of truth. The central one of these claims – and the one that is threatened by the paradoxes – is that sentences of the following form

(T) ‘-----’ is true iff -----,

---

<sup>1</sup>Keith Simmons (1999) and Elke Brendel (2000) have argued that the paradoxes undermine deflationism. Bradley Armour-Garb and J. C. Beall (2001) have responded in detail to Simmons’s argument.

the *T-biconditionals*, capture the meaning of ‘true’.<sup>2</sup> Thus W. V. Quine calls truth a device of “disquotation.” The effect of adding ‘is true’ to the quotation name “Snow is white”, Quine says, is to cancel the quotation marks and to yield the equivalent sentence ‘Snow is white’.<sup>3</sup> According to Paul Horwich, the T-biconditionals implicitly define ‘true’.<sup>4</sup> Similar ideas are expressed by other deflationists, including Hartry Field, Christopher Hill, and Michael Williams.<sup>5</sup>

Now, as is typical in philosophy, there is something very right in this central claim of the deflationists, and there is also something quite wrong. Let us try to isolate the element that is right. Let us begin by accepting that truth can, in one of its senses, be applied to sentences, and let

---

<sup>2</sup>This is a rough formulation of the claim. It will be sharpened a little in the following discussion.

<sup>3</sup>Quine writes in *Pursuit of Truth* (1992), “To ascribe truth to the sentence [‘Snow is white’] is to ascribe whiteness to snow . . . . Ascription of truth just cancels the quotation marks. Truth is disquotation (p. 80).” A little later, Quine adds that the disquotation account is “a full account: it explicates clearly the truth or falsity of every clear sentence (p. 93).” Quine’s writings have exercised a great influence on contemporary deflationism, but we should be cautious about attributing any full and unambiguous deflationism to him.

<sup>4</sup>*Meaning* (1998), p. 107. Horwich’s remark is actually directed to propositional truth. But he treats the other notions of truth in a parallel way; see Horwich 1990.

<sup>5</sup>See Field 2001, Hill 1987 & 2002, and Williams 1986. There are significant differences in the positions of the various deflationists but these can be neglected for the argument of this paper. I will work with disquotationalism as a representative deflationary theory. Let me stress, though, that I am not attributing disquotationalism to all deflationists.

us agree to focus on its application to the sentences of English. Further, let us agree to neglect the complications that indexicals, context sensitivity, and ambiguity create in any account of truth. The above rough formulation of the deflationists' claim does not get off the ground unless we make these concessions. Moreover, as far as our present interests are concerned, the concessions are inconsequential.

The deflationists' claim contains one principle that is undoubtedly true. This is the Closure principle:

*The Closure principle.* The following two rules of inference, TI and TE, hold for categorical affirmations:

(TI) A; therefore, 'A' is true

(TE) 'A' is true; therefore, A.<sup>6</sup>

Note that this principle is very weak: it licenses the interchange of A and "'A' is true' only in categorical contexts, not in contexts of hypothetical reasoning. For example, it does not entitle one to infer 'Snow is black' from the *supposition* that "'Snow is black' is true'. On the other hand, if one flat out asserts "'Snow is black' is true' then the Closure principle commits one to 'Snow is black'. Because of its weakness, the Closure principle does not yield inconsistencies even in the presence of paradoxical sentences. Set 'the Liar' to be a name of the sentence

The Liar is not true.

---

<sup>6</sup>Here and at many places below, quotes should be understood in the manner of Quine's corner quotes.

If TI and TE had unconditional validity then a contradiction would be immediate: the supposition ‘The Liar is not true’ would yield ‘The Liar is true’, and the supposition ‘The Liar is true’ would yield ‘The Liar is not true’. But such applications of TI and TE are not licensed by the Closure principle, for they occur within hypothetical contexts. The Closure principle, it is easy to show, remains consistent even when a language is enriched with all kinds of resources for expressing self- and cross-reference.<sup>7</sup>

The Closure principle is plainly correct even for paradoxical sentences. For example, a person who categorically affirms ‘The Liar is not true’ is thereby committed to ‘‘The Liar is not true’ is true’. If the person refuses to recognize the commitment, he is either confused or fails to fully grasp the meaning of ‘true’. The Closure principle ought, therefore, to be respected by all theories of truth, deflationist and non-deflationist alike. And, indeed, all the main approaches to the paradoxes validate the principle.<sup>8</sup>

---

<sup>7</sup>Friedman and Sheard 1987 is a rich study of the principles that can, and those that cannot, consistently be held in the context of self-referential truth. See McGee 1991 and Halbach 1994 for further illumination.

<sup>8</sup>See my paper “Truth” for a survey of the main approaches. The axiomatic theory KF articulated by Solomon Feferman (1984) does violate the Closure principle. In this theory, the inference rule “P, therefore ‘P’ is true” is not admissible: P can be a theorem, yet ‘‘P’ is true’ can fail to be one. (Actually, in KF, one can prove of some theorems that they are not true.) KF is elegant, but it is not a good account of the actual logic of truth, and Feferman has not proposed it as such. McGee 1991 contains a valuable discussion of KF.

The Closure principle is not respected by the approach Horwich takes to the paradoxes (Horwich 1990, p. 42). Horwich reacts to the paradoxes by excluding some of the T-

## II

Let us now turn to something stronger and more troublesome, namely, the claim that the T-biconditionals are correct. Is this claim right? Plainly, on some ways of understanding ‘true’, the claim is doubtful. It is sometimes said, following P. F. Strawson, that sentences whose presuppositions fail – e.g., ‘The king of France is bald’ – are neither true nor false. And it is also said that truth attributions to these sentences – e.g., ‘‘The king of France is bald’ is true’ – are simply false. If this is right, then the two sides of the T-biconditionals are not always equivalent, and we cannot unqualifiedly endorse the biconditionals. Nevertheless, there is a notion of truth on which the semantic value of a sentence is the same as that of its truth-attribution. On this notion, if A is neither true nor false, then ‘‘A’ is true’ is also neither true nor false. If A has a semantic value  $v$  then ‘‘A’ is true’ also has the semantic value  $v$ , and conversely. It is this notion – sometimes called the *weak* notion of truth (Yablo 1985) – that is of primary interest to the deflationists.

Deflationists do need to give some account of the other notions of truth. But let us not pause to reflect on what they might say about them. Let us work with the weak notion and return to the question whether the T-biconditionals can now be deemed to be correct. The restriction to the weak notion ensures that the T-biconditionals of ordinary, unproblematic sentences are correct.

---

biconditionals from the theory of truth. And he does not supplement the theory in any way to prevent the loss of Closure. The failure of Closure is, I think, a grave flaw in Horwich’s approach. One consequence of it is that Horwich’s theory fails to sustain his own claims about the concept of truth – for example, his claims about the explanatory power of his theory and about the generalization function of truth. Horwich recognizes the problem in the second edition of his book (p. 42, fn. 21), but I think he underestimates it.

The question is whether those of paradoxical sentences such as the Liar are also correct.

I suggest that we do not rush to give a definitive answer to this question. Let us recognize instead that a strong case can be made for both answers, that the T-biconditionals are not correct and that they are correct. The T-biconditional for the Liar,

‘The Liar is not true’ is true iff the Liar is not true,

is equivalent, by the substitutivity of identicals, to the biconditional,

The Liar is true iff the Liar is not true,

which is only a few short steps removed from an explicit contradiction. Further, it appears, ‘not’ can express a concept of negation on which the semantic values of P and not-P are always distinct. If so, there are readings on which the two sides of the above biconditional have different values; hence, on such readings, the biconditional is incorrect.

On the other hand, there are equally strong reasons to think that the biconditional is correct. The Liar exhibits a specific type of semantic instability: the idea that the Liar is true forces recognition of the idea that the Liar must then not be true, and the latter idea, in turn, forces recognition of the earlier one. This semantic behavior is quite special: it is quite different, for example, from that of the Truth Teller (“the Truth Teller is true”). The T-biconditional for the Liar seems to capture something right and important about the Liar.

Let us buttress this point in another way. The Liar remains puzzling even after we recognize that its T-biconditional is not true. Our attitude towards the Liar paradox is quite

different from that towards other popular puzzles and paradoxes. With the latter, our perplexity disappears completely once we concede that some crucial idea or presupposition that we brought to the puzzle is false. But with the Liar this is not so. Our perplexity remains even after we have been brought to concede that its T-biconditional is untrue. Furthermore – and this is a more important point –, we continue to use ‘true’ in ways that presuppose some sort of general equivalence between sentences and their truth attributions. These uses are not marginal, but central and important. Here is an example. Deflationists have observed that the truth predicate enables us to generalize over sentence positions. Consider the sentence ‘Self-conjunctions of truths are true’, or more formally,

(1) For all sentences  $x$ , if  $x$  is true then ‘ $x$  &  $x$ ’ is true.

Deflationists point out that this sentence serves as a universal generalization of

(2) If  $A$  then  $A$  &  $A$ .<sup>9</sup>

Observe that in (2) the variable ‘ $A$ ’ occupies sentence positions. In contrast, in (1) the variable ‘ $x$ ’ occupies name positions. The truth predicate in (1) enables us to gain the effect of generalizing over sentence positions while using variables that occupy name positions. Deflationists see in this fact the distinctive contribution of the truth predicate, and they view ‘true’ as having a purely

---

<sup>9</sup>I am not suggesting that all generalizations involving truth can be understood in this way. I am claiming only that in this example it is correct to take (1) as having the force of generalizing (2).

logical function. This last idea should certainly be subjected to a skeptical inquiry before it is accepted. But the earlier idea, that (1) expresses a generalization of (2), captures our ordinary understanding of (1). Observe, though, that (1) cannot serve as a generalization of (2) if we unequivocally reject the T-biconditionals for paradoxical sentences. For (1) yields only

(3) If ‘the Liar is not true’ is true then ‘the Liar is not true & the Liar is not true’ is true.

To serve as a generalization of (2), it should yield

(4) If the Liar is not true then the Liar is not true & the Liar is not true.

Without the T-biconditionals for the Liar and its self-conjunction, we lack the resources to move from (3) to (4), and back from (4) to (3).<sup>10</sup> The T-biconditionals, in some reading of them, are somehow in play in our understanding of (1).

Even if we are inclined to reject the T-biconditionals for the paradoxical sentences, we must recognize that the concept of truth imposes some constraints on the treatment of these sentences. The concept does not leave the status of paradoxical sentences completely open.<sup>11</sup>

Both the Liar and its self-conjunction are paradoxical. So, if the status of these sentences were

---

<sup>10</sup> Note that the Closure principle is insufficient to yield the desired equivalence.

<sup>11</sup>Scott Soames (1999) has proposed that ‘true’ is a “partially defined” predicate: the rules governing ‘true’ leave it open whether paradoxical sentences fall, or do not fall, under the concept of truth. See my “Partially defined predicates and semantic pathology” (2002) for some criticisms of Soames’s proposal.



completely open, then the supposition that the Liar is true and its self-conjunction untrue should be coherent. But plainly it is not: there is a conceptual constraint that requires that the Liar and its self-conjunction be treated in the same way. The source of this constraint lies in the truth conditions – however peculiar – that the T-biconditionals attribute to these sentences. Hence, an explanation of the constraint – as well as of numerous others like it – must appeal to the T-biconditionals.

In summary, the argument I am making is this. (i) The T-biconditionals explain the distinctive character of the Liar and other paradoxical sentences. These biconditionals explain how and why the behavior of paradoxical sentences differs from that of unproblematic sentences and also from other types of pathological sentences. (ii) The concept of truth imposes constraints even on the treatment of paradoxical sentences. An explanation of these constraints will need to appeal to the T-biconditionals. And, finally, (iii) *the T-biconditionals for paradoxical sentences play an indispensable role in ordinary uses of some unproblematic sentences*. In view of these three facts, it is unlikely that we can obtain a good account of the concept of truth if we unequivocally abandon the T-biconditionals for paradoxical sentences. Our account of truth must somehow work with these biconditionals.

The intuition that the T-biconditionals are correct is so strong that some theorists of truth and paradox have been moved to espouse the *Inconsistency View*, the view that the principles governing truth are inconsistent. Alfred Tarski wrote that the paradoxes “prove emphatically that the concept of truth . . . when applied to colloquial language in conjunction with the normal laws of logic leads inevitably to confusions and contradictions” (Tarski, 1935; p. 267 in Tarski 1956). Charles Chihara has defended Tarski’s suggestion at some length (1979). Graham Priest has gone a step further and has argued not only that the principles are inconsistent but that they prove some

contradictions to be true (1979 & 1987). I myself think that the Inconsistency View faces serious problems in explaining ordinary uses of 'true'. If the Inconsistency theorist accepts classical logic, as Tarski and Chihara do, he has difficulty explaining how even simple truth attributions (e.g., 'Snow is white' is true') manage to have coherent content.<sup>12</sup> If, on the other hand, the theorist abandons classical logic, as Priest does, then he has difficulty explaining our ordinary inferential practices. In any case, the point I want to stress is that the Inconsistency View, even if it misjudges the price of inconsistency, is based on an insight: none of the T-biconditionals can summarily be dismissed.<sup>13</sup>

I have been arguing that there is something correct about the T-biconditionals. I do not want to suggest, however, that this idea is a comfortable one, much less that the above considerations show us how to make sense of it. I hope the argument so far establishes this much: that we need a better understanding of the T-biconditionals and, in particular, of the connective 'iff' that occurs in them. It is not hard to supply a meaning for 'iff' under which the T-biconditional for the Liar sentence is false. The problem is to specify a meaning under which it is true. A sense of 'iff' under which the biconditional is false is as follows. 'A iff B' is true if and only if A and B have the same semantic value (i.e., they are both true, or both false, or both neither-true-nor-false, or both have the degree of truth *i* or . . . ); otherwise, it is false. Let us call 'iff' so understood the *material biconditional* and use the sign '≡' for it. Then, the very meaning

---

<sup>12</sup>See Belnap's and my book, *The Revision Theory of Truth* (1993), Chapter 1, for more on this point.

<sup>13</sup>It is this fact, I think, that has led some deflationists, and philosophers with deflationist sympathies, to opt for the Inconsistency View. See Armour-Garb and Beall 2001, Armour-Garb 200+, and Hill 2002.

of ‘not’ dictates that the material T-biconditional

‘The Liar is not true’ is true  $\equiv$  the Liar is not true

is not true. It is not so easy, however, to give a reading of ‘iff’ under which the biconditional is true.<sup>14</sup> This should not be surprising: the task of specifying such a sense for ‘iff’ is a crucial part of a theory of truth and paradox. Let us simply mark the problem for now by symbolizing the troublesome ‘iff’ as ‘ $\leftrightarrow^?$ ’. And let us accept it is a desideratum on a theory of truth that it supply an account of ‘ $\leftrightarrow^?$ ’ that sustains the following principle.

*The Tarski principle.* The biconditionals

(T?) ‘-----’ is true  $\leftrightarrow^?$  -----

are correct for the weak notion of truth.

From now on, I shall understand by ‘T-biconditionals’ sentences of the form (T?). Claims about T-biconditionals will remain schematic until a sense is given to ‘iff’ (‘ $\leftrightarrow^?$ ’).

Two minimal constraints may be imposed on an adequate treatment of ‘ $\leftrightarrow^?$ ’. First, the T-biconditionals, on the new reading, should sustain the Closure principle. Second, they should imply the material biconditionals for ordinary, non-pathological sentences. I myself would impose

---

<sup>14</sup>And only true. Priest’s dialetheism yields a reading of ‘iff’ on which the T-biconditional for the Liar is true. On this reading, however, the biconditional is also false.

a third, somewhat more controversial, constraint: the biconditionals should not imply contradictions. In particular, the T-biconditional for the Liar should not imply the corresponding material biconditional.

In summary: Deflationists accept the T-biconditionals and they are correct to do so. It is a burden on the specialist on paradox to tell a story that makes the deflationists' acceptance coherent and correct.

### III

Let us now turn to the full claim of the deflationists: that the T-biconditionals fix the meaning of 'true'. One minor doubt arises here immediately, namely, that the T-biconditionals do not guarantee that only sentences are true. They allow the possibility of, for example, the Moon being true. The doubt is easily addressed by modifying a little the deflationists' claim. Let us leave the modification tacit and move on to a more important issue: the meaning of 'meaning' in the claim.<sup>15</sup>

'Meaning', it is widely recognized, is ambiguous. It can be understood in a weak way as "extension," in an intermediate way as "intension," and in a strong way as "sense." (Further grades can also be distinguished but these three will suffice for our present purposes.) Meaning understood in a weak way captures extensional information about an expression. For example, in this way of understanding, the meaning of 'the number of planets' and 'the square of three' is the same and can be identified with the number nine. Meaning understood in the intermediate way, i.e.

---

<sup>15</sup>The condition that the theory of truth imply 'only sentences are true' is, of course, included in Tarski's Convention T (Tarski 1935).

as intension, captures extensional information across possible situations. (Different grades of “intension” can be distinguished on the basis of the range of possible worlds encompassed in an intension.) Thus, the intension of ‘the number of planets’ is different from that of ‘the square of three’, for in some possible situations the two expressions denote different numbers. Sense is a more fine-grained notion than intension and, it must be confessed, more murky. Sense is supposed to capture the cognitive content of an expression. The sense of ‘the square of three’ differs from that of ‘the number of planets’; it differs also from that of ‘nine’, though the two have the same intension.

There is terminological tangle about meaning as extension that we need to straighten out before we return to the deflationists’ claim. In one common use of ‘extension’, the extension of a one-place predicate is the set of objects of which the predicate is true. Thus, the extension of ‘river’ is the set of all rivers, the extension of ‘natural number between 1 and 5’ is the set {2, 3, 4}, and so on. ‘Extension’ so understood captures part of the extensional information about a (one-place) predicate, but it does not capture all of it. Extension does not tell us the objects of which the predicate is false. In a classical context, we can recover this information if the totality of all objects is given to us. This is not possible, however, if the predicate is gappy (say because of considerations of sortal correctness) or if it is fuzzy (because of vagueness) or if it is n-valued . . . . So, extension, in this way of understanding it, does not capture all the extensional information about a predicate. Let us use the term ‘signification’ for total extensional information about a term. Signification includes the information carried by extension, but it is richer. For example, the signification of a gappy predicate will yield the objects of which the predicate is true, those of which it is false, and those of which it is neither true nor false; the signification of a vague

predicate will yield the degree of applicability of the predicate to an object; and so on.<sup>16</sup>

On the weak reading of ‘meaning’, the deflationists’ claim yield this thesis:

*The Signification thesis.* The T-biconditionals fix the signification of the weak notion of truth. Or, more fully, given the non-semantic facts that obtain in the actual world, the T-biconditionals fix the actual signification of truth.

I think this thesis is better dubbed ‘principle’. It is true and captures an important constraint on theories of truth and paradox. Consider first the T-biconditionals for nonpathological sentences. Consider, for example, these three biconditionals:

‘Two plus two is five’ is true iff two plus two is five,

‘Two is sad’ is true iff two is sad,

‘Fred is bald’ is true iff Fred is bald,

Let us assume that, because of sortal incorrectness, ‘Two is sad’ is neither true nor false; and that Fred is bald to the degree  $r$ . Then, the three biconditionals dictate, correctly, that ‘true’ is false of ‘Two plus two is five’, that it is neither true nor false of ‘Two is sad’, and that it applies to the degree  $r$  to ‘Fred is bald’. The biconditionals fix properly the signification of ‘true’ with respect to these sentences. The point holds for all non-pathological sentences, because the T-biconditionals for these sentences imply the corresponding material biconditionals and, furthermore, our focus is

---

<sup>16</sup>This use of ‘signification’ was introduced in Belnap’s and my *Revision Theory*, pp. 30-31.

on the weak notion of truth.

Let us consider next the pathological sentences such as the Liar and the Truth Teller. Do they provide a counterexample to the Signification thesis? This question cannot be answered directly, for we do not have as yet a sufficiently rich account of the ‘iff’ in the T-biconditionals. Let us therefore reformulate the question thus. Should we require a theory of truth and paradox to preserve the Signification thesis over the pathological sentences? I think that the answer to this question is “Yes.” A good descriptive account of our notion of truth will not pin a particular semantic value (e.g., “false,” “neither true nor false,” and “applicable to degree  $r$ ”) on the Liar and the Truth Teller. Instead, it will simply indicate – and explain – the sort of pathological behavior that these sentences exhibit. Further, the special behavior of these sentences is rooted in the T-biconditionals. There is nothing over and above the T-biconditionals (and the contingent facts) that settles the signification of truth on the pathological sentences. There is a problem, I grant, in understanding how the T-biconditionals fix the behavior of pathological sentences. It is the task of a theory of truth and paradox to provide us with this understanding. The theory should provide us with a reading of the T-biconditionals that sustains the Signification thesis. Many current theories of truth and paradox, it seems to me, neglect this desideratum, with the result that time and effort are wasted on alternatives that are easily dismissed.<sup>17</sup>

---

<sup>17</sup>I mean the remark to be directed only to those theories that aim to address the *descriptive* problem posed by the paradoxes – i.e., the problem of giving a description of our ordinary concept of truth, as opposed to revising the concept to suit one or another purpose.

In the fixed-point approaches of Saul Kripke (1975) and Robert Martin and Peter Woodruff (1975), the Tarski principle is sustained if ‘iff’ is read as the Łukasiewicz biconditional and the signification of ‘true’ is taken to be the least fixed point of a particular monotone three-

Let us now turn briefly to the stronger readings of the deflationists' claim: that the T-biconditionals fix the intension of 'true' (the *Intension* thesis) and that they fix the sense of 'true' (the *Sense* thesis). These theses deserve an extended discussion, but for our purposes here it is sufficient to make just one observation about them: The paradoxes threaten these theses only to the extent that they threaten the Signification thesis. I have argued that the paradoxes do not threaten the Signification thesis. It follows that they do not threaten the Intension and the Sense theses.

It is the Sense thesis that is of critical importance to deflationism,<sup>18</sup> and we can reinforce the idea that the paradoxes do not threaten it as follows. Consider a fragment *L* of English that is weak in self-referential resources – so that pathological sentences cannot be expressed in it – but that is otherwise as rich as one cares to make it. Suppose we grant the deflationists that the Sense thesis holds for 'true' when it is restricted to *L*. Now, what arguments might the paradoxes provide to put the Sense thesis into doubt for languages with richer self-referential resources?

---

valued scheme (e.g., the Strong Kleene). But such a theory fails to sustain the Signification thesis. The thesis can be sustained, however, if keeping the same reading of 'iff', we take the signification to be the set of all fixed points of the scheme. Note that the Łukasiewicz biconditional cannot be expressed within Kripke's and Martin-Woodruff's fixed-point languages.

<sup>18</sup>The argument of my "A critique of deflationism" (1993), if sound, shows that the Intension thesis is insufficient to support deflationary conclusions about truth. The Intension thesis is a generalized version of the Signification thesis. It says that the signification of 'true' in a possible world *w* is fixed by the T-biconditionals and the facts in *w*. The Intension thesis can, therefore, be sustained (see *Revision Theory*, pp. 20-25). The Sense thesis is, however, more doubtful and it – or something like it – is needed to reach deflationary conclusions.



Suppose we enrich  $L$  with the name ‘the Liar’ and let its denotation be ‘The Liar is not true’. Do we have any grounds to doubt the Sense thesis – any grounds that are not at the same time grounds for doubting the Signification thesis? No! We can rehearse the Liar argument, we can deduce the contradictions, we can induce puzzlement. But we leave the Signification thesis intact. And we bring forth no doubts against the Sense thesis – at least none that do not arise even for the paradox-free language  $L$ .

In conclusion: the paradoxes do not pose any special threat to the deflationists’ claim that the T-biconditionals fix the meaning of ‘true’.

#### IV

Friends of deflationism have responded to the paradoxes in several ways. An especially popular way has been to abandon classical logic in favor of a non-classical one.<sup>19</sup> The popularity of this approach is understandable. It blocks the derivation of absurdities from paradoxical sentences. And, on some variants, it even allows one to affirm all the material T-biconditionals and, at the same time, to explain how the concept of truth manages to do substantive work. (In such variants, the classical principle *ex falso quodlibet* fails.) Furthermore, the work of Kripke, Martin and Woodruff, Albert Visser (1984), and others has powerfully illuminated the virtues of the non-classical approaches. Nevertheless, I want to argue that deflationists, as well as theorists of paradox, should take a neutral stance on logic. They should not let the Liar dictate to them one logic over another.

---

<sup>19</sup>See, for instance, Grover 1992, Weir 1996, Gauker 1999, Armour-Garb and Beall 2001, Armour-Garb 200+, and Field 2001.

Let us begin by observing that the principal claims of deflationism are neutral on logic. The plausibility of these claims – e.g., that the T-biconditionals fix the meaning of ‘true’ and that truth enables us to generalize on sentence positions – does not alter as we shift our attention from one part of our language to another. Whether we consider classical arithmetic, or discourse with vague terms, or biological discourse – where, because of considerations of sortal correctness, the semantics of predicates may be three-valued –, the deflationists’ claims remain equally plausible. The claims remain plausible even in those areas of language whose semantics has eluded us, e.g., hyper-intensional constructions. Moreover, if it were to turn out that we are wrong about the logic of even the simplest parts of our language – say, if it were to turn out that this logic is seven-valued (as the Jaina logicians thought) – the deflationists’ claims would remain unaffected. It appears, therefore, that these claims are not sensitive to the underlying logic and semantics of language.

A theorist of truth and paradox should also, I think, keep a neutral stance on logic. The paradoxes arise not just in classical logic but in other kinds of logics as well: three-, four-, n-valued, modal, intuitionist, relevant, etc. It is true that in some logics, for example, the Strong Kleene, the paradox appears to be less damaging. But this is so only because these logics are weak in their logical resources: some kinds of negation, for example, are inexpressible in them. Once the logics are enriched, the paradox appears with the full force it exhibits in classical logic.<sup>20</sup>

---

<sup>20</sup>The negation that is expressible in Strong Kleene is *choice* negation ( $\sim$ ):  $\sim A$  is true iff  $A$  is false, and  $\sim A$  is false iff  $A$  is true. If  $A$  is neither true nor false then  $\sim A$  is also neither true nor false. Exclusion negation ( $\neg$ ) is not expressible, however:  $\neg A$  is false iff  $A$  is true; otherwise,  $\neg A$  is true. The Liar sentence formed using exclusion negation is no less puzzling than the standard classical Liar.

The work on the paradoxes in the last thirty years has revealed a panoply of ways of weakening expressive power while gaining unproblematic self-referential truth. Broadly speaking, these ways fall into two groups: those that restrict syntactic resources and those that restrict logical resources. The effect of either type is the same: genuinely paradoxical sentences are eliminated from the language.<sup>21</sup> The work on expressively weak languages enriches our understanding of the concept of truth. But it does not by itself provide a solution to the descriptive problem of understanding truth. The project of seeking a solution to the descriptive problem is quite distinct from that of constructing expressively weak self-referential languages. A theorist engaged in the former project should aim to explain the working of the concept of truth in any logical environment: classical, as well as  $n$ -valued; intuitionist, as well as relevant; expressively weak, as well as expressively rich. Expressively weak logics provide no shelter to the descriptive theorists. Indeed, such theorists seek no shelter; they wish to face the full force of the paradox so that they can come to understand its nature and its source.

The behavior of paradoxes (and other pathological sentences) is so similar across different logics and semantics that it is fair to demand that any account of them be general, that it apply uniformly across the whole range of logics and semantics. No account of the paradoxes is likely to be plausible that makes an essential appeal to a particular semantics, e.g., the Strong Kleene. For then it would have no resources to explain similar paradoxes that arise under other semantics. If, following the Strong Kleene fixed-point route, it is said that the Liar is neither true nor false, then the question immediately arises what we should say about the variant Liar that is formulated using exclusion negation (see fn. 20). We cannot say that this Liar is neither true nor false, for that leads

---

<sup>21</sup>The Liar sentence formulated using choice negation in Strong Kleene languages is not, in my view, genuinely paradoxical. It exhibits no instability if it is deemed neither true nor false.

to absurdity.

Furthermore, in all cases, irrespective of the underlying logic, it is the T-biconditionals that are the source of the distinctive behavior of the Liar (and other pathological sentences). Hence, it is proper to demand that a theory of truth give an account of the biconditionals that is *logic neutral*. It should assign a reading to ‘iff’ (‘ $\leftrightarrow$ ’) that works – or at least can naturally be carried over – across the whole range of logics. That is, theorists of paradox should aim at an account of the T-biconditionals that enables them to meet the following challenge. Given any logic – classical, three-valued, n-valued, relevant, . . . – they can predict and explain which sentences behave pathologically in any given set of circumstances, and which sentences do not. Further, for each of the pathological sentences, they can predict and explain the particular kind of pathology that the sentence exhibits; and for each of the non-pathological ones, they can predict and explain the semantic value of the sentence. This requirement of generality, it seems to me, is the sound residue of “the universality of natural language” that has been the bugbear of theorists of paradox.<sup>22</sup>

There is one further requirement that I think we should impose on an account of the paradoxes: it should not attribute a special logic to sentences containing ‘true’. Logical resources (e.g., negation, conjunction, and quantification) should interact with ‘true’ in just the way that they do with the other predicates. In our ordinary reasoning with sentences containing ‘true’, we do not hold them to be above the usual logical laws. We are not worried – nor should we be – if in the course of an otherwise logically valid argument we take a detour through some sentences containing ‘true’. There is, to be sure, something very peculiar about ‘true’. The peculiarity, I am

---

<sup>22</sup>I have still to see a formulation of the thesis of “the universality of natural language” that is both clear and true.

suggesting, is all centered on the T-biconditionals and, in particular, on the little word ‘iff’.

## V

I have argued that a descriptive account of the concept of truth must provide a sense of ‘iff’ (‘ $\leftrightarrow$ ’) under which the following desiderata are satisfied.

*Desiderata A.* (a1) The T-biconditionals are correct and fix the signification of truth. (a2) The T-biconditionals for non-pathological sentences yield the corresponding material biconditionals. (a3) The Closure principle is sustained.

*Desiderata B.* (b1) The T-biconditionals do not imply contradictions. (b2) The T-biconditionals are logic neutral: they fix the signification of truth irrespective of the logic and semantics of the language. (b3) The logical rules of the language apply uniformly to sentences containing ‘true’.

There should not be any dispute about Desiderata A. Desiderata B, on the other hand, are bound to be somewhat controversial, since many existing theories of paradox fail to satisfy one or more of them. I myself think that a failure to satisfy any of the desiderata is evidence against a theory. A violation of desiderata A is, I think, a fatal flaw. Violations of desiderata B may be tolerable if it can be shown that they promote overriding virtues.

If the above desiderata are correct then the paradoxes pose no special problems for the deflationists. The deflationists can happily assert their favorite slogan, “the T-biconditionals fix the

meaning of ‘true’.” If a pesky objector invokes the paradoxes, the deflationists need only exercise a bit of caution. They should not rush – as they have been apt to do – to construct or to endorse a particular account of the paradoxes. Far too often this kind of move has led not only to a poor theory of paradox; it has led to an undermining of the deflationists’ principal claims. In response to the objector, the deflationists need say only that their claims, extensionally construed, constitute an important desideratum on any account of paradox. It is the task of the specialist to show how the desideratum can be satisfied.

It may be objected that the above desiderata are too strong, that they cannot possibly be satisfied. The response to the objection is that this just is not so. There is a natural and intuitive way of satisfying the desiderata, and it is this. Following Tarski, we view the T-biconditional for a sentence *S* as defining the conditions under which *S* falls under ‘true’. We observe – and this is a plain but striking fact – that the T-biconditionals sometimes define these conditions in circular ways. Often ‘true’ occurs in the defining conditions; sometimes the defining conditions for *S* contain as a constituent part the very thing being defined, namely, the truth attribution to *S*. The sense in which the T-biconditionals – including those of paradoxical sentences such as the Liar – are correct is one in which they are seen as laying down possibly circular defining conditions for the applicability of ‘true’. ‘Iff’ should therefore be viewed as representing “partial, possibly circular, definitional equivalence.”<sup>23</sup> This reading of ‘iff’ is logic-neutral in the desired way. Further, it is intuitively plausible that, on this reading, the T-conditionals fix the signification of

---

<sup>23</sup>Caution: This should not be taken to imply that the two sides of the T-biconditionals have the same sense or the same cognitive content. The biconditionals constitute at best an *intensionally* adequate definition of ‘true’, not a *sense* adequate definition. See *Revision Theory*, p. 130.

truth. They dictate (in light of the contingent facts) what sentences are non-pathological and what pathological. They dictate also the particular sort of pathology that the latter sentences display.

This reading of ‘iff’ rests on the idea that circular definitions are legitimate. Belnap and I have defended the legitimacy of circular definitions at length in our book (1993). We have tried to show that circular definitions are logically coherent and that they are in general contentful. I provide a quick overview of our theory in the survey paper “Truth” (2001). The key to meeting the above desiderata lies, I believe, in the recognition that truth is a circular concept.<sup>24</sup> It is not my aim here, however, to convince the reader of this claim. All I need do for my present purposes is to counter the suggestion that the above desiderata are too strong. Perhaps there are other ways of satisfying the desiderata than that provided by the theory of circular definitions. So long as the desiderata *can* be satisfied, they are reasonable. And if they are reasonable, then the paradoxes pose no special threat to deflationism.

The deflationists’ main claim – that the T-biconditionals fix the meaning of ‘true’ – is correct, I have argued, if it is read *extensionally*. This reading is insufficient, however, to establish deflationary conclusions about the role of truth in philosophy (e.g., in the theory of meaning). The deflationists need for their arguments much stronger readings – readings such as that the T-biconditionals fix the sense of ‘true’.<sup>25</sup> The debate over deflationism centers – and ought to center – on two points: (i) what precise reading of the deflationists’ claim is needed to establish deflationary conclusions about the role of ‘true’, and (ii) whether the claim is correct under this

---

<sup>24</sup>I know of only one serious objection to Belnap’s and my claim that truth is a circular concept. This was put forward by Robert Koons in his critical study (1994) of *Revision Theory*. Belnap’s and my reply (1994) accompanies Koons’s study.

<sup>25</sup>See my “A critique of deflationism” (1993).

reading. The paradoxes are only a distraction in this debate; they take attention away from the central issues.<sup>26</sup>

## References

Armour-Garb, Bradley 200+. Minimalism, the generalization problem and the liar. Forthcoming.

Armour-Garb, Bradley and J.C. Beall 2001. Can deflationists be dialetheists? *Journal of Philosophical Logic* 30, 593-608.

Armour-Garb, Bradley and J. C. Beall, eds. 2003. *Deflationary Truth*. La Salle IL: Open Court.

Brendel, Elke 2000. Circularity and the debate between deflationists and substantive theories of truth. In Chapuis and Gupta 2000, pp. 29-47.

Chapuis, André and Anil Gupta 2000. *Circularity, Definition and Truth*. New Delhi: Indian Council of Philosophical Research. Distributed outside India by Ridgeview Publishing Company, Atascadero CA.

Chihara, Charles S. 1979. The semantic paradoxes: a diagnostic investigation. *Philosophical*

---

<sup>26</sup>I am grateful to Professor Bradley Armour-Garb for helpful discussions, and for nudging me to write this paper.



*Review* 88, 590-618.

Feferman, Solomon 1984. Toward useful type-free theories I. *Journal of Symbolic Logic* 49, 75-111. Reprinted in Martin 1984, 237-287.

Field, Hartry 2001. *Truth and the Absence of Fact*. Oxford: Oxford University Press.

Friedman, Harvey, and Michael Sheard 1987. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic* 33, 1-21.

Gauker, Christopher 1999. Deflationism and logic. *Facta Philosophica* 1, 167-195.

Grover, Dorothy 1992. *A Prosentential Theory of Truth*. Princeton: Princeton University Press.

Gupta, Anil 1993. A critique of deflationism. *Philosophical Topics* 21, pp. 57-81. Reprinted, with a postscript, in Armour-Garb and Beall 2002, pp. xx-yy.

Gupta, Anil 2001. Truth. In Louis Goble, ed., *A Guide to Philosophical Logic* (Oxford: Blackwell Publishers), pp. 90-114.

Gupta, Anil 2002. Partially defined predicates and semantic pathology. *Philosophy and Phenomenological Research* 65, 402-409.

Gupta, Anil and Nuel Belnap 1993. *The Revision Theory of Truth*. Cambridge MA: The MIT Press.

Gupta, Anil and Nuel Belnap 1994. Reply to Robert Koons. *Notre Dame Journal of Formal Logic* 35, 632-636.

Halbach, Volker 1994. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic* 1994, 311-327.

Hill, Christopher S. 1987. Rudiments of a theory of reference. *Notre Dame Journal of Formal Logic* 28, 200-219.

Hill, Christopher S. 2002. *Thought and World: An Austere Portrayal of Truth, Reference, and Semantic Correspondence*. Cambridge UK: Cambridge University Press.

Horwich, Paul 1990. *Truth*. Oxford: Basil Blackwell. Second edition, 1998, Oxford University Press.

Horwich, Paul 1998. *Meaning*. Oxford: Oxford University Press.

Koons, Robert 1994. Book review: *The Revision Theory of Truth*. *Notre Dame Journal of Formal Logic* 35, 606-631.

Kripke, Saul A. 1975. Outline of a theory of truth. *Journal of Philosophy* 72, 690-716. Reprinted in Martin 1984, pp. 53-81.

Martin Robert L., ed. 1984. *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press.

Martin Robert L. and Peter W. Woodruff 1975. On representing 'true-in-L' in L. *Philosophia* 213-217. Reprinted in Martin 1984, pp. 47-51.

McGee, Vann 1991. *Truth, Vagueness, and Paradox: An Essay on the Logic of Truth*. Indianapolis, Hackett Publishing Company.

Priest, Graham 1979. The logic of paradox. *Journal of Philosophical Logic* 8, 219-241.

Priest, Graham 1987. *In Contradiction: A Study of the Transconsistent*. The Hague: Martinus Nijhoff.

Quine, W. V. 1992. *Pursuit of Truth*. Revised Edition. Cambridge MA: Harvard University Press.

Simmons, Keith 1999. Deflationary truth and the liar. *Journal of Philosophical Logic* 28, 455-488.

Soames, Scott 1999. *Understanding Truth*. Oxford: Oxford University Press.

Tarski, Alfred 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* 1, 261-405. Translated by J. H. Woodger, "The concept of truth in formalized languages," in Tarski 1956, pp. 152-278.

Tarski, Alfred 1956. *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Clarendon Press. Second edition, 1983, John Corcoran (ed.), Hackett Publishing Company.

Visser, Albert 1989. Four valued semantics and the liar. *Journal of Philosophical Logic* 13, 181-212.

Weir, Alan 1996. Ultra-maximalist minimalism! *Analysis* 56, 10-22.

Williams, Michael 1986. Do we (epistemologists) need a theory of truth? *Philosophical Topics* 14, 223-242.

Yablo, Steve 1985. Truth and reflection. *Journal of Philosophical Logic* 14, 297-349.