SOUTHWEST PHILOSOPHY REVIEW

Vol. 29, No. 1 **January, 2013**

ARTICLES

Liar-Like Paradoxes and Metalanguage Features

Klaus Ladstaetter
Washburn University

1. Introduction

In their (2008) C. S. Jenkins and Daniel Nolan (henceforth, JN) argue that it is possible to construct Liar-like paradox in a metalanguage, even though its *object language* is not semantically closed. I do not take issue with this claim; however, I shall criticize several further points (more or less explicitly) contained in JN's article.

First, I take issue with the view that JN's examples (the ones on which I focus) show that it is possible to construct Liar-like paradox in a metalanguage, even though this *metalanguage* is not semantically closed. Instead I argue that there is a good sense in which it is a mistake to say that "... no semantic predicate for its [the metalanguage's] *own* sentences is used" (JN, 2008, p. 69; my insertion).

Second, I am critical of the view that JN's examples of Liar-like paradox present *counterexamples* to Tarski's diagnosis of the classic Liar paradox. Instead I argue that Tarski's diagnosis, when modified in a suitable way and applied to Liar-like paradox, is confirmed (rather than undermined) by JN's examples.

Third, I find fault with JN's failure to notice *Tarski's postulate*, the requirement that a semantic term must not be introduced into the metalanguage except by definition. I believe that the neglect and violation of *Tarski's postulate* is the root cause of the semantic closure of the *metalanguage* considered by JN and thus of the possibility of constructing Liar-like paradox *within it*.

Finally, I suggest a resolution of the discussed paradox in Tarski's spirit. Moreover, I raise doubts whether there is a possible world about which we can reason in the way suggested by JN.

2. The Argument

JN claim that

> ... ways of generating Liar-like paradox with object languages that contain no semantic terminology are easy to come up with. For instance, one can specify a possible situation in which all the true sentences of the language are tokened in yellow, and consider:

> ‘q is not tokened in yellow’
>
> where ‘q’ refers to the sentence just mentioned.
>
> Or one can specify that all and only the true sentences of the object language are written on page 2226 of *The Big Book*, and that the language contains a sentence:
>
> ‘p is not written on page 2226 of *The Big Book*’
>
> where ‘p’ refers to the very sentence just displayed. (2008, p. 70)

The examples presented by JN are structurally identical. Rather than reconstructing both arguments I will reconstruct the argument *form* that they share. Here is the reconstruction.

Imagine a world which contains a language L such that all and only those sentences that are true-in-L are φ (e.g. are tokened in yellow, are written on page 2226 of *The Big Book*). Imagine also that in this possible world the language L contains the expression:

q is not φ

where ‘q’ names the very expression just displayed. We, in the actual world, can then reason in the metalanguage of L about this possible world as follows:

1	(1)	(∀x)(‘q’ names x ↔ q = x)	A
2	(2)	‘q’ names ‘q is not φ’	A
1	(3)	‘q’ names ‘q is not φ’ ↔ q = ‘q is not φ’	UI, 1
1,2	(4)	q = ‘q is not φ’	↔E, 2,3
5	(5)	(∀y)L-sentence (y is φ ↔ y is true-in-L)	A (“correlation”)
5	(6)	‘q is not φ’ is φ ↔ ‘q is not φ’ is true-in-L	UI, 5
5	(7)	‘q is not φ’ is not φ ↔ ‘q is not φ’ is not true-in-L	↔Neg, 6
1,2,5	(8)	q is not φ ↔ ‘q is not φ’ is not true-in-L	=, 4,7
9	(9)	‘q is not φ’ is true-in-L ↔ q is not φ	A
1,2,5,9	(10)	‘q is not φ’ is true-in-L ↔ ‘q is not φ’ is not true-in-L	↔Sub, 8,9

Contradiction. All ingredients of a paradox (or of an antinomy) have been assembled. As Tarski (1969, p. 66) remarks:

> Starting with premises that seem intuitively obvious, using forms of reasoning that seem intuitively certain, an antinomy leads us to nonsense, a contradiction. Whenever this happens, we have to submit our ways of thinking to a thorough revision, to reject some premises in which we believed or to improve some forms of argument which we used.[1]

3. Comments on the Argument

Line (1) of the argument is obtained by applying the schema:

(REF) $(\forall x)$(‘a’ refers to x $\leftrightarrow$ a = x)

to the L-expression ‘q’. Line (1) may, in Tarski’s spirit, be regarded as a partial definition of reference (while line (9) may be regarded as a partial definition of truth). Line (2) formally captures the reference-fixing stipulation informally being made by JN while setting the stage of the argument. Line (3) follows from line (1) by universal instantiation. Line (4) is derived from lines (2) and (3) by biconditional elimination. The assumption on line (5) *correlates* the true L-sentences with those that are φ. Line (7) follows from line (6) by negating both sides of the biconditional which is obtained from line (5) by universal instantiation.

Line (8) is obtained from lines (4) and (7) by substituting identicals for each other. Let us call the entire process by which line (8) has been derived from lines (1), (2), and (5) the *process of diagonalization*.

On the left side of the biconditional on line (8) the L-expression ‘q is not φ’, and not just its name, occurs in the metalanguage of L for the first time. Notice that as soon as we let ‘q’ be the name of the expression ‘q is not φ’, the expression at once becomes meaningful. That is, as soon as the reference of ‘q’ is fixed this way, ‘q is not φ’ says of ‘q is not φ’ (i.e. of *itself*) that it is not φ (e.g. that it is not tokened in yellow, that it is not written on page 2226 of *The Big Book*). “Direct” self-reference has been achieved here in the way outlined by Kripke:

> Let ‘Jack’ be a name of the sentence ‘Jack is short’, and we have a sentence that says of itself that it is short. I can see nothing wrong with “direct” self-reference of this type. ... There is no vicious circle in our procedure, since we need not *interpret* the sequence of marks ‘Jack is short’ before we name it. Yet if we name it “Jack,” it at once becomes meaningful and true. (1975, p. 693)

On the right side of the biconditional on line (8) the sentence "'q is not φ' is not true-in-L" occurs – a sentence which says of 'q is not φ' that it is not true-in-L. Line (8) thus tells us that the sentence 'q is not φ' – which says of itself that it is not φ – and the sentence which says of this very sentence that it is not true-in-L are *materially equivalent* (i.e. that they *logically say the same* or have the *same truth value*).

The T-biconditional on line (9) is an instance of the schema:

(T) 'p' is true-in-L ↔ p.

On the left side of the biconditional on line (9) the predicate 'is true-in-L' is applied to the (now fully interpreted) L-sentence 'q is not φ', i.e. to a sentence which logically says the same as the sentence which says of this very sentence that it is not true-in-L, as is witnessed by line (8).

The result of the process of diagonalization, i.e. the assertion on line (8), together with the assertion of the T-biconditional on line (9) immediately lead to the contradiction on line (10), by biconditional substitution.[2] More informally, the contradiction has been derived from line (8) which says:

This sentence is not φ ↔ 'This sentence is not φ' is not true-in-L,

and from line (9) which says:

'This sentence is not φ' is true-in-L ↔ this sentence is not φ.

4. Semantic Closure and *Tarski's Postulate*

Tarski's *diagnosis* of the classic Liar paradox is well-known and will not be repeated here.[3] It boils down to the claim that *if* a language £ is semantically closed, *then* it is possible to construct the Liar antinomy *within this* language, where:

£ *is semantically closed* ↔

(i) in addition to £-expressions, £ contains the names of these £-expressions, and

(ii) £ contains semantic terms that apply to £-expressions that already contain these semantic terms,[4] and

(iii) all sentences that determine the adequate usage of the semantic terms can be asserted in £.

In reaction to the classic Liar paradox, Tarski decides not to use *any* semantically closed languages. His decision necessitates the distinction of an entire *hierarchy* of object- and metalanguages (cf. Tarski, 1944, p. 67). The starting point of the hierarchy is a *base* language $£_0$. Restricting the vocabulary of $£_0$ to non-semantic terms guarantees that $£_0$ is not semantically closed. Each language $£_{k+1}$ contains the language $£_k$ on the level below it (or its translation into $£_{k+1}$) as a part. $£_{k+1}$ also contains a truth predicate, 'is true$_{k+1}$', *for* sentences of the language $£_k$; but each truth predicate is subject to the *restricted T-schema*:

$$\text{(TR)}\ \ `p_k\text{' is true}_{k+1} \leftrightarrow p_k,$$

where the resulting T-biconditionals belong to $£_{k+1}$. No language $£_{k+1}$ can thus contain a truth predicate 'is true$_{k+1}$' which applies to its *own* sentences, i.e. to sentences of $£_{k+1}$.

While JN correctly contend that no language of the Tarski-hierarchy can contain its *own* truth predicate (cf. JN, 2008, p. 68), they fail to notice that Tarski imposes further conditions on the object- and metalanguages of the hierarchy. The condition relevant for our discussion is that

> ... we desire *semantic terms* (referring to the object language) *to be introduced into the metalanguage only by definition*. For, if this postulate is satisfied, the definition of truth, or of any other semantic concept, will fulfill what we intuitively expect from every definition; that is, it will explain the meaning of the term being defined in terms whose meaning appears to be completely clear and unequivocal. And, moreover, we have then a guarantee that the use of semantic concepts will not involve us in any contradiction. (Tarski, 1996, p. 67)

It can hardly be denied, in light of this passage, that Tarski thinks that the *mere* distinction between object- and metalanguages and the restriction of the application of semantic (e.g. truth) predicates to the sentences of the language in the hierarchy below is *not sufficient* for excluding the possibility of constructing a semantic paradox. In addition, the fulfillment of what I'll henceforth call *Tarski's postulate* is needed. Tarski's postulate is the requirement that a semantic term be introduced into a metalanguage only as *definiendum*.

As I see it, Tarski justifies this postulate *quite independently* of considerations regarding paradox. We are interested in clarifying the meanings of semantic terms, and we can satisfactorily do so only if we define them by other terms whose meanings are already completely clear. This is why we require semantic terms to be introduced into the metalanguage only by definition. As a *consequence* we will have the guarantee that the defined semantic terms will not involve us in paradoxes.[5]

5. Modification of Tarksi's Definition

It is worth noting that Tarski's initial explication of the semantic closure of a language is *intra-linguistic*. That is, it does not take into account the object/metalanguage-distinction. Once the distinction is drawn though, Tarski's initial explication should be adapted to it. In Tarski's spirit, I suggest the following modification:

> *A metalanguage* $£_{k+1}$ *is semantically closed* $\leftrightarrow$
> (i) in addition to $£_k$-expressions, $£_{k+1}$ contains the names of these $£_k$-expressions, and
> (ii) $£_{k+1}$ contains semantic terms (such as 'is true$_{k+1}$') that *either* apply to $£_{k+1}$-expressions that already contain these semantic terms, *or* to $£_k$-expressions that are materially equivalent to $£_{k+1}$-expressions that contain these semantic terms,[6] and
> (iii) all REF-, SAT-, and T-biconditionals for $£_k$-expressions *or* for $£_{k+1}$-expressions can be asserted in $£_{k+1}$.[7]

Given this modification, the *metalanguage* of JN's examples clearly is *semantically closed*. Not only does it contain the L-sentence 'q is not φ' (see line 8), but it also contains its name. Moreover, on line (9) the predicate 'is true-in-L' is applied to the L-sentence; and even though the L-sentence itself does not contain the predicate 'is true-in-L', we know from line (8) that the sentence is materially equivalent to a sentence of the metalanguage of L which *does* contain this very predicate. The result, then, is the same as if the predicate 'is true-in-L' were directly applied to the metalanguage sentence "'q is not φ' is not true-in-L". So, in this sense, it is false to claim, as JN do, that no semantic predicate is used in the metalanguage *for* one of its *own* sentences; in this sense, the metalanguage of L *does* contain its *own* truth predicate.

The obtained result confirms Tarski's diagnosis, according to which the semantic closure of a language is *sufficient* for the possibility of constructing semantic paradox *within it*. What ultimately gives rise to the semantic closure of the metalanguage is the violation of *Tarski's postulate*. The violation occurs on line (5), for the predicate 'is true-in-L' is here introduced into the metalanguage *not* as definiendum, but as part of an assumption which *correlates* the true L-sentences with those that are φ.

6. Towards a Resolution of the Paradox

If we do not want to be compelled to accept a contradiction, some assumption of the argument – line (1), (2), (5), (9), or a combination thereof – has to be given up. It is, of course, the nature of a paradox that each of its assumptions, taken individually, is apparently acceptable, while their joint assertion leads to a contradiction.

Since the REF-biconditional on line (1) is presumably as "untouchable" as is the T-biconditional on line (9), and since there is nothing wrong per se with letting 'q' be the name of the partially uninterpreted sequence of marks 'q is not φ' (see line 2), the problem is created by the assertion on line (5) – in conjunction with the other assumptions, of course.

In light of these other assumptions we are simply *not* free to assume that the *imagined* world contains a language whose sentences are true just in case they are φ. Constituting a violation of *Tarski's postulate*, the assumption on line (5) should thus be rejected. If it is decided to do so, the paradox cannot arise.

7. A Possible World?

The question emerges whether we can even *imagine* a world that *both* contains a sentence that says of itself that it is not φ *and* at the same time belongs to a language whose sentences are true just in case they are φ. To lower the abstraction-level of the discussion, let 'written on page 2226 of *The Big Book*' be a stand-in for 'φ'.

Clearly I can imagine a world with a language L containing a sentence that says of itself that it is not written on page 2226 of *The Big Book*. And clearly I can imagine a world with *The Big Book*, where on page 2226 all and only the true L-sentences appear. But can I imagine a world in which both are the case?

JN obviously think that we can imagine such a world. For the theorists argue that (for the small, finite language L they have in mind) it can be specified that *The Big Book* contains *all* sets of L-sentences. Thus *The Big Book* will also contain the set of all and only those sentences that are true-in-L. If there is a worry at all,

> ... the worry must really be that none of these sets is the set of all and only the true sentences of the language. But that claim is extremely puzzling, and does not constitute a satisfactory response on its own to the paradox of *The Big Book*. (JN, 2008, p. 70)

Contra JN I believe that there can be no set of *all and only* those sentences that are true-in-L *and* that are written on page 2226 of *The Big Book*. That is, there can be no world with the language L in which *The Big Book* is written, where L also contains a sentence that says of itself that it is not written on page 2226 of *The Big Book*.

For if the Liar-sentence under consideration *is* true-in-L, then, given what it says, it is *not* written on page 2226 of *The Big Book*; and so the page will *not* contain *all* sentences that are true-in-L. But if it is *not* true-in-L, then, given what it says, it *is* written on page 2226 of *The Big Book*; and so the page will *not* contain *only* sentences that are true-in-L.

Here is another way of thinking about the problem. The presented argument allows us, via obvious fiddling with lines (8) and (9), to draw the conclusion:

> q is written on page 2226 of *The Big Book* ↔
> q is not written on page 2226 of *The Big Book*.

The question, then, is whether it is possible for the sentence 'q is not written on page 2226 of *The Big Book*' to be *both* written *and* not be written on page 2226 of *The Big Book*. The answer, to my mind, is negative.[8]

JN, on the other hand, seem to be able to imagine that the sentence (a slice of *reality*, albeit of *linguistic* reality) both has this property and does not have it. The theorists thus seem to be able to imagine that *reality itself* (in this alleged possible world) is *contradictory*. But this defies the very notion of a (normal) possible world in my view.[99]

8. Conclusion

In the words of Hilary Putnam (who uses them in a different context though), my assessment of JN's example is as follows:

> This story may seem intelligible to us at first blush, at least as an amusing possibility. On reflection, however, we come to see that a logical contradiction is involved. (2002, p. 50)

Notes

[1] Cf. also Sainsbury (2007, p. 1) whose notion of a paradox is closely related to Tarski's.

[2] Given classical logic, the derivation of an "explicit" contradiction in the form of a conjunction is routine and will not be presented here.

[3] Cf. Tarski (1996, p. 66; 1983, pp. 158 and 164f.).

[4] In regard to the truth predicate, condition (ii) is frequently stated by the slogan: £ contains its *own* truth predicate.

[5] For an informal characterization of definitions and their rules, cf. Tarski (1965, pp. 33-36).

[6] In a slogan: $£_{k+1}$ contains its *own* truth predicate.

[7] SAT-biconditionals are obtained from the schema:

(SAT) (∀x)('F' is satisfied by x ↔ x is F)

[8] Cf. Eldridge-Smith, where the authors discuss the Pinocchio paradox – which is structurally similar to JN's example – and where they conclude:

> It seems that there could be a logically possible world in which Pinocchio's nose grows if and only if he is saying something that is not true. However, there cannot be such a logically possible world wherein he makes the statement 'My nose is growing'. (2010, p. 214)
>
> For in such an alleged possible world Pinocchio's nose would have to be *both* growing *and* not growing when he makes this statement.

[9] Some dialetheists are, of course, not impressed by arguments that aim to show that reality is not contradictory. While Priest does not endorse the correspondence theory of truth, he argues for its compatibility with the existence of true contradictions. This requires that the world contain *both* an atomic, positive *and* its corresponding negative fact which, given a suitable theory of facts, is at least *prima facie* 'perfectly possible' in Priest's view (cf. Priest, 2006, pp. 51-54).

Works Cited

Eldridge-Smith, Peter and Veronique Eldridge-Smith. (2010) The Pinocchio Paradox. *Analysis* 70(2): 212-215.

Jenkins, Carrie S. and Daniel Nolan. (2008) Liar-Like Paradox and Object Language Features. *American Philosophical Quarterly* 45(1): 67-73.

Kripke, Saul. (1975) Outline of a Theory of Truth. *The Journal of Philosophy* 72(19): 690-716.

Priest, Graham. (2006) *Doubt Truth to be a Liar*. Oxford: Clarendon Press.

Putnam, Hilary. (2002) Brains and Behaviour. In David J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 45-54). New York: Oxford University Press. (Original work published 1968.)

Sainsbury, Richard, M. (2007) *Paradoxes* (2nd ed.). Cambridge: Cambridge University Press.

Tarski, Alfred. (1965) *Introduction to Logic*, 3rd ed. New York: Oxford University Press.

———. (1969) Truth and Proof. *Scientific American* 220: 63-77.

———. (1983) The Concept of Truth in Formalized Languages. In John Corcoran (ed.), Joseph H. Woodger (trans.), *Logic, Semantics, Metamathematics*, 2nd ed. (pp. 152-278). Indianapolis: Hackett Publishing Company. (Original work published 1935.)

———. (1996) The Semantic Conception of Truth and the Foundations of Semantics. In Aloysius P. Martinich (ed.), *The Philosophy of Language*, 3rd ed. (pp. 61-84). Oxford: Oxford University Press. (Original work published 1944.)