

LIAR-LIKE PARADOX AND OBJECT LANGUAGE FEATURES

C. S. Jenkins and Daniel Nolan

Consider a classic Liar sentence, *s*:
s is not true

Straightforward reasoning seems to show that
s is true and *s* is not true:

- (1) *s* is true (assumption)
- (2) '*s* is not true' is true (meaning of '*s*')
- (3) *s* is not true (T-elimination)

Since (3) contradicts (1), (1) can be discharged:

- (4) *s* is not true (reductio ad absurdum)
- (5) '*s* is not true' is true (T-introduction)
- (6) *s* is true (meaning of '*s*')
- (7) *s* is true and *s* is not true (from (4) and (6))

The key rules, T-elimination and T-introduction, play an important role in this reasoning. These (or similar rules, and/or conditional statements) are often taken to be characterizing features of truth (or at least, of truth predicates). There are other ways to derive this contradiction from the Liar sentence given those key rules and standard logical principles: excluded middle provides another

way, and going from (3) via conditional proof to "if *s* is true then *s* is not true," and then to "*s* is not true" by *clavius* is another way, and there are others still. Notice that the proof of the contradiction rests on no undischarged assumptions: if the rules are allowed in full generality with a strong enough logic, the inconsistent theorems are generated.

Two features of object languages that tend to be blamed for Liar-like paradoxes are: (A) the object language's containing (some of) its own semantic predicate(s), and (B) the language's containing the means to refer to truth-bearers (sentences or propositions).

However, it turns out to be far from clear that either of these (or their interaction) is straightforwardly to blame for Liar-like paradox. It is the purpose of this paper to show that Liar-like paradox can occur without either feature being present.

Here is a classic motivating passage from Tarski (1944):

If we now analyze the assumptions which lead to the antinomy of the liar, we notice the following:

(I) We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "*true*" referring to sentences of the language; we have also assumed that all

sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "*semantically closed*."

(II) We have assumed that in this language the ordinary laws of logic hold.

(III) We have assumed that we can formulate and assert in our language an empirical premise such as the statement (2) which has occurred in our argument.

It turns out that the assumption (III) is not essential, for it is possible to reconstruct the antinomy of the liar without its help. But the assumptions (I) and (II) prove essential. Since every language which satisfies both of these assumptions is inconsistent, we must reject at least one of them.

It would be superfluous to stress here the consequences of rejecting the assumption (II), that is, of changing our logic (supposing this were possible) even in its more elementary and fundamental parts. We thus consider only the possibility of rejecting the assumption (I). Accordingly, we decide not to use any language which is semantically closed in the sense given.

This motivates the introduction of a hierarchy of languages, each containing a truth predicate for expressions of the language below it, but none containing its own truth predicate.

First of all, note that it is possible to create paradox without semantic closure. Consider a simple finite object language L1 which contains only the following:

- (i) a term for a concrete item—say 'A', which stands for a tennis ball—and a predicate of color—say 'is yellow', which means the same as it does in English;
- (ii) terms for some (but not all) of its own finitely many sentences;
- (iii) terms for some (but not all) of the sets of its sentences;

- (iv) a negation operator and an operator expressing set membership (for ease, these can be 'not' and 'is a member of').

Also stipulate that the set of true sentences of L1 exists and has a name in L1, and that the name in L1 of this set is 'S'. Now consider the following sentence of L1:

'p is not a member of S'

where 'p' is a term in L1 which refers to the very sentence just displayed. This is enough to generate a Liar-like paradox. To see this quickly and informally: if p is a member of S, p is true, and hence, looking at what p says, one can deduce that p is not a member of S. But if p is not a member of S, it is true, for that is what it says, and hence it is a member of S. In the manner of the above argument concerning the classic liar, one can reason (in the metalanguage, of course) as follows:

- (1) p is true-in-L1 (assumption)
- (2) 'p is not a member of S' is true-in-L1 (meaning of 'p')
- (3) p is not a member of S (T-elimination)
- (4) p is not true-in-L1 (definition of 'S')
- (5) p is not true-in-L1 (reductio from (4) and (1), assumption at (1) discharged)
- (6) p is not a member of S (definition of 'S')
- (7) 'p is not a member of S' is true-in-L1 (T-introduction)
- (8) p is true-in-L1 (meaning of 'p')
- (9) p is true-in-L1 and p is not true-in-L1 (5, 8, conjunction introduction)

Notice that the object language L1 does not contain any semantic predicates: that is, any predicates meaning 'true', 'false' or the like. The metalanguage contains a predicate 'true-in-L1', but no semantic predicate for its *own* sentences is used.

One might respond that in such a language as L1, 'is a member of S' is really a truth predicate. So this has not really shown that Liar-like paradox can be generated without an object-language truth predicate.

It seems implausible that 'is a member of S' is a truth predicate, but in any case, one can adjust the case so that no such response is available. Let L2 no longer contain a term for the set S. Consider, instead, the set S* of sentences of L2 containing all the true sentences of L2 except one: 'A is yellow'. Let S* also contain one untrue sentence, namely 'A is not yellow'. S* does not contain anything else. Now consider the following sentence of L2:

'q is not a member of S*'

where 'q' is a term in L2 that refers to the very sentence just displayed. Again, one has a Liar-like paradox on one's hands. If q is a member of S*, given that it is not identical to S*'s one untrue sentence, q is true, and hence, looking at what it says, one can deduce that it is not a member of S*. But if q is not a member of S*, it is true, for that is what it says, and hence, since S* contains all the true sentences apart from one which is not identical to q, q is a member of S*. It is very implausible to say that 'is a member of S*' is a truth predicate, given that the members of S* include an untruth and do not include all the truths. Nor is it straightforward to define a truth predicate in L2, or even something with the right extension, since it contains no term for a set like S and does not contain a name for each of its sentences. Matters can be made even harder by specifying that L2 does not contain a name for either the false sentence which is in S* or the true one which is not in S*.

It should now be clear that L2 lacks a truth predicate but is nevertheless susceptible to Liar-like paradox. If this is right, then banning a truth-predicate (and indeed various other resources a language might have to talk about itself) is not enough to prevent Liar-like sentences in that language. This means that such solutions cannot be enough on their own, and neither can solutions that rely simply on restricting principles involving a truth-predicate in the object language. Some approaches to the Liar restrict the T-rules and the associated T-schema to avoid paradox (see, e.g., Horwich 1990), but doing so at least for the object language will be of no help here: without a truth-predicate at all, a fortiori the object language will not incorporate rules for introducing and eliminating such a predicate.

One other question that might be raised is whether either 'true-in-L' predicate is really a truth predicate for (some of) the *meta*-language. If so, perhaps it could be argued that paradox has only arisen here because the metalanguage does not meet Tarski's strictures. 'True-in-L1' follows the right sort of introduction and elimination rules in the above reasoning, although it doesn't of course obey the truth introduction rule in full generality. Of course, any truth-in-the-object-language predicate will obey rules like that, if the metalanguage contains the object language. However, 'true-in-L1' only applies to any sentences in the metalanguage if it is assumed that the object language is literally a fragment of the metalanguage—if the object language is distinct, then 'true-in-L1' will be distinct from the 'true-in-ML' predicate that applies to the associated meta-language sentences.

Anyway, a case can be constructed without this feature, where the metalanguage does not contain the object language. Let 'SS' be the metalanguage name for S and 'is a member of' be the metalanguage expression for 'is a member of'. Provided you know how

to translate from object language to metalanguage, i.e., you know the metalanguage principle

TRANSLATION: 'p is not a member of S' is true-in-L1 iff p is not a member of SS

then you can reason as follows:

- (1) p is true-in-L (assumption)
- (2) 'p is not a member of S' is true-in-L1 (meaning of 'p')
- (3) p is not a member of SS (TRANSLATION)
- (4) p is not true-in-L1 (definition of 'SS')
- (5) p is not true-in-L1 (reductio from (4) and (1), assumption at (1) discharged)
- (6) p is not a member of SS (definition of 'SS')
- (7) 'p is not a member of S' is true-in-L1 (TRANSLATION)
- (8) p is true-in-L1 (meaning of 'p')
- (9) p is true-in-L1 and p is not true-in-L1 ((5), (8), conjunction introduction)

Once the idea is clear, other 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.

Some may be concerned that it should not be assumed that there will be a page of *The Big Book* where all and only the true sentences are written, and likewise for yellow sentences. But note that for small finite languages like the ones under consideration here, it can be specified that *The Big Book* is such that, for every set of sentences of the object language, there is some page of *The Big Book* such that all and only the members of that set appear on that page. Hence 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*.

It might be tempting to suggest at this point that it must be something to do with object-language reference to truth bearers (in this case, sentences of the object language) that is causing the problem. All the examples considered so far involve reference to sentences of the object language.

But notice, first, that *reference* to truth-bearers is not required; quantification over them will do just as well. Consider a language L3 which is just like L except that instead of terms for its own sentences it contains Russellian definite descriptions of them. Clearly, some sentence of the form:

'The F is not a member of S'

with S the set of true sentences of L3 and 'The F' a definite description of the sentence mentioned on the line above this one, will generate exactly the same problem as the sentences considered so far.

In fact, not even quantification over truth bearers is needed. One just needs to specify that the object language contains terms for the singletons of its sentences, and a primi-

tive subethood relation. One then considers the sentence:

'T is not a subset of S'

with 'T' a term referring to the singleton of the sentence just displayed, and S the set of true sentences of the object language.

Some might still feel that talk of sets is (in some sense) talk about those sets' members. They might therefore think that talk of sets of truth-bearers does not really get away from talk of truth-bearers. Again, this does not seem very plausible; one can, for instance, talk knowledgeably about a set while having no idea what its members are. (E.g., 'The set Ross is thinking of has a transfinite cardinality'.) But in any case, one can get even further away from such talk, if desired, without losing the ability to generate Liar-like paradoxes.

Consider a finite object language L4 which is like the original language L1 except that each of the sentences has a unique number associated with it. This might be a Gödel number, or a number in some canonical ordering of the sentences of L4. In addition, one of the sentences of L4 is:

'n is not a member of N'

where 'n' is a name for the number of the sentence just displayed, and 'N' is a name for the set of numbers of true sentences of L4.

This sentence generates a Liar-like paradox. If it is true, then n is not a member of N, for that is what the sentence says. But if n is not a member of N, then the sentence is not true, from the specifications of 'n' and 'N'. If on the other hand the sentence is not true, then n is not in N, given the specifications of 'n' and 'N'. But then the sentence is true, for what it says is that n is not in N.

However, paradox is generated here without object language reference to truth bearers, or any object language semantic terminology. Although meta-language semantic terminology and reference to object language truth

bearers is used to fix the reference of 'n' and 'N', this does not mean the object language terms themselves are semantic or refer to truth bearers. Both 'n' and 'N' are names, the former for a number, the latter for a set of numbers.

It would moreover be strange to claim that 'is a member of N' is somehow a truth predicate, given that it is a predicate applicable to numbers and not to the kinds of things which can be true or false (in L4). But in any case, one can further obviate any possible concern that paradox is only created here because 'is a member of N' is really a truth predicate, by constructing a variant case along lines familiar from an earlier example. So consider a language L5 with its sentences associated as before with numbers, containing the following sentence:

'm is not a member of M'

where m is a name for the number of the sentence just displayed and M is a name for the set which contains the numbers of all the true sentences of L5 except for one, which is a sentence about a tennis ball. M also contains the number of exactly one false sentence, also having to do with tennis balls. The same paradox is generated as before, and 'is a member M' certainly is not a truth predicate, for one of the numbers in M is a number of a false sentence, and one of the numbers of true sentences is not in M.

One thing left to blame for creating Liar-like paradox is object language negation. But notice that it would be quite easy to construct a paradox without object language negation, by introducing an object language name 'O' for the set of all numbers of sentences of the object language which numbers are not members of M. Then consider the sentence

'o is a member of O'

where 'o' is a name in the object language for the canonical number of the above sentence.

Apart from the obvious consequences of the points made in this paper for views which attempt to block paradox merely by rejecting semantic closure, there are some interesting consequences for other views as well. Field (forthcoming) defends a view which rejects excluded middle and the *reductio* rule, which he considers closely related to excluded middle, to defend against Liar-like paradoxes. He tries to make this palatable by saying that excluded middle still holds *most* of the time. In particular, he says, according to the view under consideration, "excluded middle [is] perfectly acceptable within standard mathematics, physics and so forth" (Field forthcoming, p. 4). On that view,

paradox can only be avoided by rejecting the application of excluded middle to . . . Liar-like sentences formed from truth-like predicates. (Since excluded middle is to hold within ordinary mathematics and physics, this means that no truth-like predicates can be constructed within their vocabulary.) (p. 10)

It is hard to see how this can be right, given that in the last two paradoxes described above only mathematical vocabulary was used in the object language. It seems that the existence of such paradoxes must sometimes enforce the restriction of excluded middle even when using purely mathematical vocabulary. Moreover, there is no 'truth-like predicate' in Field's sense involved in these last two examples. A 'truth-like predicate' for Field is a predicate 'T' that obeys:

(T-Elim): A follows from T(<A>)

and

(T-Intro): T(<A>) follows from A,

or at least obeys (T-Elim) and

(T-Incoherence): A and $\neg T(<A>)$ are jointly inconsistent.

Note that 'is a member of N' does not obey any of these rules, on the assumption that 'follows from' means 'is logically derivable

from', and 'inconsistent' means 'logically inconsistent'. And 'is a member of M' clearly does not obey any of them even if one does not make any such assumption. Hence it is a mistake to blame truth-like object-language predicates for the Liar paradox and conclude that one is safe as long as one sticks to the language of mathematics and physics where no such predicates arise.

Some approaches to the classic Liar face no additional pressure from the paradoxes discussed here; for instance, paraconsistent approaches (see, e.g., Priest 1987) and certain paracomplete approaches (though that of Kripke 1975, with its appeal to a Tarskian hierarchy to avoid semantic closure when discussing the status of his "grounded" predicate, is not off the hook). Other approaches face, if not additional pressure, at least a slightly different challenge from the ones they have addressed so far. Contextualism about 'true', for instance (see, e.g., Burge 1979), may be able to handle the cases discussed here but the explanation of what brings about the context shift will need to be tailored to suit the fact that 'true' does not appear in the target sentence.

The negative moral of the story is that it seems unlikely that what features the object language has matters much at all for whether there are Liar-like paradoxes around. It appears that there is no single object language terminological device which can be blamed for the creation of paradox, nor any interesting collection of such devices whose interaction can be so blamed. One possible positive moral is that, in the cases described in this paper, it is somehow the use of *metalanguage* semantic terminology, truth-rules, negation, and/or means to refer to or quantify over object language truth bearers that are generating the problem.

University of Nottingham

ACKNOWLEDGEMENTS

The authors would like to thank Michael Glanzberg and JC Beall, and audiences at the Australian National University and the University of Melbourne, for helpful discussion of this material. Special thanks are due to John Hawthorne for helping to spark some of the ideas in this paper. Daniel Nolan's work on this paper was supported by a Philip Leverhulme Prize from the Leverhulme Trust.

REFERENCES

- Burge, T. 1979. "Semantical Paradox," *Journal of Philosophy*, vol. 76, pp. 169–198.
 Field, H. Forthcoming. "Solving the Paradoxes, Escaping Revenge," in *The Revenge of the Liar: New Essays in Paradox*, ed. J. Beall (Oxford: Oxford University Press). (Page numbers in this paper refer to the online draft, available at <http://philosophy.fas.nyu.edu/docs/IO/1158/revengetex3.pdf>.)
 Horwich, P. 1990. *Truth* (Oxford: Basil Blackwell).
 Kripke, S. 1975. "Outline of a Theory of Truth," *Journal of Philosophy*, vol. 72, pp. 690–716.
 Priest, G. 1987. *In Contradiction* (Dordrecht: M. Nijhoff).
 Tarski, A. 1944. "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research*, vol. 4, pp. 341–376.