

R. M. Sainsbury

Paradoxes

3rd. edition

Cambridge 2009: Cambridge UP

6.2 The Liar: semantic defects

The material of the next four sections is hazardous. (Recall the fare of Philotas, mentioned in the introduction.)

A relatively recent version of the Liar paradox appears in St. Paul's Epistle to Titus (1, 12-13) (see question 6.2). This version involves the island of Crete and the notion of lying, and lying involves an intention to deceive. These features are irrelevant to the paradox. Eliminating such irrelevancies, we get something like this:

What I am now saying is false.

6.2

It is not clear that the saint sees any logical, as opposed to moral, problems. The relevant text is as follows:

12. One of themselves, even a prophet of their own, said, The Cretans are always liars, evil beasts, slow bellies.
13. This witness is true. Wherefore rebuke them sharply, that they may be sound in the faith.

St. Paul's version depends on the assumption that all the other Cretans are liars. Construct an explicit argument for the contradiction (perhaps modeled on that given below for L_1) that makes this dependence plain.

The simplest version of all, which will be the starting point of the discussion, is

L_1 : L_1 is false.

Here we have a sentence, called L_1 , that is supposed to say of itself that it is false. One can derive something apparently paradoxical as follows. Suppose it is true; then it is as it says it is – false. So it is false. Suppose that it is false. Well, *false* is just what it says it is, and a sentence that tells it the way it is is true. So it is true. To sum up: if L_1 is true, it is false; and if it is false, it is true. Is this paradoxical? Perhaps it sounds as if it is, but let us look more carefully. We have two conditional claims:

- (a) If L_1 is true, then it is false.
- (b) If L_1 is false, then it is true.

We assume that anything that is false is not true, and anything that is true is not false; so (a) and (b) yield:

- (a) If L_1 is true, then it is not true.
 (b) If L_1 is false, then it is not false.

If a sentence implies its own negation, then we can infer that negation. (This principle is called *consequenzia mirabilis*. It amounts to the validity of the sequent: $A \rightarrow \neg A \vdash \neg A$.) Both (a) and (b) offer inputs to this principle. The first assures us that " L_1 is true" implies its negation, so the principle tells us that we can infer that L_1 is not true. The second, in an exactly parallel way, enables us to infer that L_1 is not false. So standard reasoning guarantees that L_1 is not true and is also not false. Let us summarize this as follows:

G: L_1 is neither true nor false.

Is *this* paradoxical? Not unless we have some independent reason to suppose that L_1 is either true or false. For example, we might be able to justify some *principle of bivalence*, perhaps to the effect that *every* sentence, and so in particular L_1 , is either true or false. Otherwise we might simply *accept* G, saying that L_1 lies in a *gap* between truth and falsehood (hence the "G"). This would not in itself offer a complete account of the paradox, for it would remain to discover general principles to explain why L_1 should fail to be true and fail to be false. But accepting G would at least fix the general approach.

We could not accept G if there were some irresistible reason for supposing that L_1 had to be either true or false. Might there be an irresistible reason for accepting some principle of bivalence? The version given in the preceding paragraph is certainly not true. Questions are expressed in sentences, but no question is either true or false. Suppose then we restrict the principle to declarative, indicative sentences. Still, there are putative counterexamples, for example:

You have stopped beating your wife.

If you have never beaten your wife, the sentence is certainly not true; but to say it is false, or to say that you have not stopped beating your wife, arguably suggests that you are still beating her. Again, consider a case in which someone says

That elephant is about to charge

when there is no elephant in the offing. We certainly cannot count the sentence as true; but can we count it as false? If we did this, should not the following sentence be true?

That elephant is *not* about to charge.

Yet, if there is no elephant, this seems as poor a candidate for truth as the previous one.

Despite the apparent counterexamples, it is hard not to feel the pull of the thought that some principle of bivalence, no doubt suitably refined, ought to be correct. The underlying idea might be expressed like this: any non-defective representation of how things are in the world must be either accurate or inaccurate, true or false. Some sentences, like questions and commands, are not designed to represent the world, so there is no question of them representing it correctly or incorrectly. Other sentences, though designed to represent the world, fail to count as representations at all, correct or incorrect, in virtue of some semantic defect. The case of the missing elephant is a putative example. For that sentence to represent the world at all, it must refer to an elephant (or so it is plausible to suppose). Since it fails to refer, it counts as semantically defective, and so counts as neither true nor false.

In sum, one natural and immediate response to the Liar is to accept the reasoning which leads to the conclusion that the paradoxical sentence, L_1 , is neither true nor false (see question 6.3). Since it is hard to suppose that a semantically non-defective sentence could fail to be either true or false, this approach carries with it the obligation to explain wherein the defectiveness of L_1 consists. It would be totally unexplanatory to say that the defect consists in the sentence's potential for paradox; for that potential is precisely what we need to understand.

6.3

* Show how the principles of reasoning in this section can apparently be used to derive that L_1 is both true and false. This derivation shows that one should not regard L_1 as a basis for a straightforward proof of G.

Most accounts of the Liar paradox endeavor to establish plausible general principles upon which the Liar sentences are defective. We will consider some of these in detail in subsequent sections.

6.3 Grounding and truth

One approach to identifying a semantic defect in L_1 starts from the idea that the truth of a sentence must be grounded in something outside the sentence itself. We can make the thought vivid by imagining how one might introduce someone to the notion of truth.

We take the learner to understand most of English, but not the word "true." We could try to explain the notion of truth using the following recipe:

You should call a sentence true iff you are willing to assert it.

("If" abbreviates "if and only if.") The learner could use this explanation to respond to, for example, "Snow is white" by saying "True!" and to respond to "Grass is red" by saying "Not true!" However, he could not in the first instance use the explanation to find out how to respond to a sentence like:

(1) "Snow is white" is true.

Until he has already understood "true," he cannot know what it would be to be willing to use this sentence to assert something. At a later stage, once he becomes aware that he is to respond with "True!" to "Snow is white," he will be able to see that he should assent to (1), and hence see that (1) is itself something to which "true" applies. The picture is of someone climbing a ladder. At the base there are sentences not containing the word "true," to which he can learn to apply the word. As he does so, he can thereby come to see how to apply the word to sentences such as (1) on the next rung up; sentences applying "true" to sentences at the base. He can work his way up this ladder indefinitely. Where S is a sentence not containing "true," he can use this process to understand any sentence of the form

... "S is true" ... is true,

where the second ellipsis stands in for any number of further occurrences of "is true," (see question 6.4).

6.4
What does the first ellipsis indicate?

Learning how to apply the concept of truth requires there to be sentences that do not themselves invoke the concept: these are the base sentences. The learning situation mirrors a putative metaphysical fact: *truth depends on something outside itself*. One might defend the claim that L_1 is neither true nor false on the grounds that L_1 does not respect this fact. Accordingly, it is semantically defective. Let me try to explain this.

Whether or not "Snow is white" is true depends on whether or not snow is white. In this simple case, whether or not something is true depends quite directly on a fact that can be expressed without invoking the concept of truth: on whether or not snow is white. This is an example of how truth depends upon something outside itself. In more complex cases, the

dependence is less direct. For example, consider (1) again ("Snow is white" is true"). Whether (1) is true depends on whether or not "Snow is white" is true. This in turn depends on whether or not snow is white. So whether or not (1) is true depends, but at one remove, on whether or not snow is white. In the end, we get back to a non-truth-involving question. In this reflection, we travel down the ladder toward the base. In considering learning, we were traveling upward from the base – same ladder, different direction.

To reinforce the suggestion, consider this series of sentences:

- (S2) (S1) is true
- (S3) (S2) is true
- (S4) (S3) is true
- ⋮

Can we genuinely make sense of such a series? Everything depends upon what (S1) is. If it is, for example, "Snow is white," then there is no problem: we reach base. However, we would never reach base if (S1) were, for example,

(S1) (S4) is true.

Here truth wanders in a circle, without ever touching the ground. In this case we need to say that none of the sentences is true – and also, for the same reason, none is false (Kripke 1975).

This line of thinking gives a general reason for accepting $G(L_1)$ is neither true nor false). L_1 can never reach base: there is no getting to a non-truth-involving fact on which the truth or falsehood of L_1 could depend. We come back always to L_1 itself, which is not a base sentence. To summarize: the trouble with L_1 is that it is *ungrounded*. The same account also applies well to

T_1 : T_1 is true.

Here is a sentence that seems to say of itself that it is true. It is not paradoxical. The supposition that it is true does not lead to the conclusion that it is not; the supposition that it is not true does not lead to the conclusion that it is. Still, intuitively there is something wrong with T_1 , and L_1 shares the defect. The account just given purports to identify this defect: T_1 is ungrounded. Like L_1 , it does not make contact with a non-truth-involving base, so both sentences are neither true nor false.

Thus G can be defended. We can provide reasons, independently of threat of paradox, for thinking that L_1 is neither true nor false: it is ungrounded. But even if all this is accepted, paradox remains.

6.4 The Strengthened Liar

G says that L_1 is neither true nor false, and thus accepts the reasoning we considered at the beginning of section 6.2. However, G itself appears to support a paradox.

G entails that L_1 is not false. This is the negation of L_1 itself. So G entails **not- L_1** : L_1 is not false.

So not- L_1 is true (using the principle that anything that entails a sentence entails the truth of that sentence). This in turn entails that L_1 is false (using the principle that any sentence whose negation is true is false). So G appears to entail a contradiction: that L_1 is not false and L_1 is false (see question 6.5). Hence it cannot constitute a resolution of the paradox.

6.5

* Show how one can (apparently) derive that L_1 is not false without appeal to G.

A related difficulty is that G is unable to deal with a related paradoxical sentence:

L_G : L_G is either false, or else neither true nor false.

We can reason as follows: suppose L_G is neither true nor false; then it is true (since it is an or-statement one of whose alternatives is true), and so it is either true or false. We can then reason as we did with L_1 to show that it is neither true nor false. Combining results, we show that it is both neither true nor false, and also either true or false.

The easiest way to see what is going on in reasoning of this sort is to consider yet another paradoxical sentence:

L_2 : L_2 is not true.

Suppose L_2 is true. Then it is as it says it is, namely, not true; so it is not true. Suppose that it is not true. Well, *not true* is just what it says it is, and a sentence that tells it the way it is is true; so it is true. To sum up: if L_2 is true, it is not true; if it is not true, it is true.

This appears to be a genuine contradiction, and one which cannot be assuaged by G. If L_2 is, as G affirms, neither true nor false, then in particular it is not true. But the reasoning just advanced purports to show that one can refute this claim: if L_2 is not true, then it is true (since *not true* is just what it says itself to be). If G is supported by ideas about grounding, the problem is simply that an ungrounded sentence is

not true: that was the whole idea. So if L_2 is ungrounded, L_2 is *not true*; we are committed to the problematic L_2 itself.

One may be tempted to try to modify G, for example to something entailing

G' : L_2 is neither true nor not true.

First, this appears to be a contradiction. Standard reasoning would enable us to infer from G that L_2 is both true and not true (see question 6.6). Second, like G, G' implies directly that L_2 is not true: it entails the paradoxical sentence itself.

6.6

The reasoning depends on the equivalence between
neither P nor Q
and
both not- P and not- Q .

On what other principle does the reasoning depend?

L_2 and the attendant reasoning is sometimes known as the “Strengthened Liar.” A standard view is that it shows the inadequacy of theories like the one based on the notion of grounding as resolutions of Liar paradoxes. More generally, it might be taken to show that any approach which tries to resolve the paradox by finding some semantic defect in L_2 is doomed, since what is semantically defective is not true.

Leaving this claim in abeyance for the moment, I now turn to an approach to resolving these paradoxes, derived from Tarski, for which the Strengthened Liar poses no special problem, no problem not already posed by the ordinary Liar. This approach also finds something semantically defective in sentences like L_1 (“ L_1 is false”) and L_2 (“ L_2 is not true”), but of a quite different kind.

6.5 Levels

In deriving apparently unacceptable conclusions from L_1 and L_2 , we relied upon two principles:

if a sentence is true, then things are as it says they are;
if things are as a sentence says they are, then the sentence is true.

Tarski stressed the feature of truth these principles capture. He expressed it somewhat more formally. Let us use “ σ ” to stand in for a name of any

sentence, and “ p ” to stand in for a sentence. Then, Tarski claimed, for any acceptable language we must accept every instance of

T : σ is true iff p

provided that the sentence named by σ means the same as the sentence that replaces p . In the limiting case, these can be the same sentence; so an instance of T (putting “Snow is white” for “ σ ” and “snow is white” for “ p ”) is

“Snow is white” is true iff snow is white.

T may seem utterly platitudinous, but the Strengthened Liar shows that it has contradictory instances. Putting “ L_2 ” for “ σ ” and “ L_2 is not true” for “ p ,” we get:

(*) L_2 is true iff L_2 is not true.

Since L_2 is “ L_2 is not true,” (*) presumably meets the requirement that the sentence named by “ L_2 ” (namely “ L_2 is not true”) means the same as the sentence which replaces p (namely “ L_2 is not true”).

One aspect of the problem posed by the Liar is that the apparently platitudinous T leads by apparently correct reasoning to the contradictory (*). Tarski’s response is that the ordinary concept of truth, the one we use every day, is incoherent and must be rejected. According to Tarski, it needs to be replaced by a series of concepts of truth, hierarchically arranged, and each expressed in a language different from any natural language (i.e. from any language that has evolved naturally).

Suppose some language λ_0 contains a predicate “ Tr_1 ” that applies to all and only the true sentences of λ_0 . Suppose also that λ_0 contains a sentence σ that says of itself that it is not Tr_1 . Then, granting T , we have a version of the Liar: if Tr_1 applies to σ , then σ says truly that Tr_1 does not apply to it; but if Tr_1 does not apply to it, then, since this is what it says, it is true, and so Tr_1 does apply to it. Tarski took the contradiction to refute the supposition that σ belongs to λ_0 . The natural explanation of how this could be is that Tr_1 is not an expression of λ_0 . Hence, no sentence belongs to λ_0 if it contains Tr_1 . This blocks the paradox in the following sense: the proposed language, since it does not contain a predicate true just of its true sentences, is one in which the paradoxical sentence cannot be formulated. One can write down the words, but they are claimed to have no significance: they are semantically wholly defective.

We can enlarge a language by adding new expressions. In particular, we could enlarge λ_0 , taken to contain no occurrence of “ Tr_1 ,” by adding “ Tr_1 .” We could call the newly formed language λ_1 : it contains all the sentences of λ_0 together with all sentences which can be formed from these

by using “ Tr_1 ”; so it contains σ . Paradox is still avoided: σ does not belong to λ_0 , and since Tr_1 is defined only for λ_0 sentences, there is no question of Tr_1 applying to σ . The expression σ (= “ σ is not Tr_1 ”) does not belong to λ_0 , and so it is not one of which “ Tr_1 ” can be significantly affirmed or denied.

It is not that there is no predicate true of just the sentences of λ_1 . There is: call it “ Tr_2 ” (see question 6.7). However, for familiar reasons, it cannot belong to λ_1 (See question 6.8). In general, a predicate Tr_n cannot belong to a language λ_{n-1} but only to a language of level at least n .

6.7

Is σ Tr_2 ?

6.8

How does the supposition that Tr_2 belongs to λ_1 lead to paradox?

No paradoxical Liar sentence can be formulated in any of the languages in Tarski’s hierarchy. How is this supposed to provide a “solution” to the paradox? The paradox arises in our language, so a proper defusing of it must say something about our language, and not merely offer a replacement.

What Tarski says about our language is that the Liar shows it to be incoherent. We must replace our actual, but incoherent, concept of truth by a family of new concepts, each fixed to a level in the hierarchy, in the way just described. Many people have sought something less radical, a response that preserves more of our ordinary thought and talk.

One such less radical response draws on a Tarskian notion of hierarchy, but claims that this is already implicit in our actual use of “true.” Unlike Tarski’s account, which claimed that ordinary language is irremediably defective, this alternative claims that the defects are mere appearance: the underlying reality is that we already use a Tarski-like hierarchy of concepts of truth.

A major difficulty with this suggestion is that there would appear to be nothing in our usage reflecting the appropriate sensitivity to Tarski-style, fixed-in-advance levels. For example, suppose I say:

What you said just now is not true.

On the face of it, anyone, including myself, could quite well know what I have said without knowing what you have said. (Imagine a game on the lines of paper, stone, and scissors, in which two players have to make a simultaneous

declaration. The task of one is to say whether the other has declared something true. Normally things work well: you declare "Snow is white" and I declare "Not true!" and you win. But what happens when I declare "Not true" and you declare "You win"? My declaration is intelligible in advance of knowing the content of yours.) On a hierarchical view in which levels are fixed in advance, something in my use of this sentence determines an association between "true" and some level. Presumably the normal (default) level would be 1. If you have said "Snow is white," there is no problem. But suppose you have said "What Mark will say is true." On the present theory, the intelligibility of my utterance requires my "true" to be on a higher level than yours; but if my utterance can be understood without knowing what you have said, its level of truth must get fixed independently of the content of what you have said. This suggests that it will be difficult to apply this kind of hierarchy response to natural language. (However, compare Burge 1979.)

So far we have considered two main ways of making good the claim that Liar-paradoxical sentences are semantically defective. One used the notion of grounding, in terms of which there seemed to be some hope of defending the view that L_1 is neither true nor false; though the hope that this would lead to a resolution of all versions of the Liar was apparently dashed by the Strengthened Liar (L_2 is not true). The other was Tarski's claim that any non-hierarchical notion of truth is incoherent. The Strengthened Liar creates no special problem for this view (see question 6.9). However, it has difficulties. To jettison our ordinary concept of truth seems too radical; yet it seems incorrect to suppose that our concept already contains, implicitly, the required segregation into levels. Where else might one look for an account of the semantic defectiveness of Liar sentences?

6.9

Why not? You might like to answer by criticizing one or both versions of the following reasoning:

Version 1:

Even when levels of truth are made explicit, as Tarski requires, we can formulate a Liar sentence, e.g.:

L_N : L_N is not true_n.

If this is defective, through infringing levels requirements, then it is not true_n; but since this is what it says it is, it must be true_n after all.

Version 2:

A sentence which violates levels is semantically defective and so not true; so one can always construct a Strengthened Liar sentence to refute a levels approach to the paradoxes. (Cf. the argument mentioned at the end of section 6.4 above.)

6.6 Self-reference

It is natural to think that something about the self-referential character of Liar paradoxical sentences is the main source of their paradoxical nature. There may be something in this thought, but as it stands it is both incorrect and inadequate.

It is incorrect because a sentence can refer to itself, as for example this very sentence does, without leading to any kind of semantic defect or paradox. So sentential self-reference cannot be the sole source of Liar paradoxes.

It is inadequate because one can construct Liar paradoxes without using any sentence which refers to itself. One example of this phenomenon involves liar cycles like the following.

(A) (said by α on Monday): Everything β will say on Tuesday is true.

(B) (said by β on Tuesday): Nothing α said on Monday is true.

If α and β said nothing other than, respectively, (A) and (B) on, respectively, Monday and Tuesday, we have a paradox of essentially the Liar type. Suppose (B) is true; then (A) is not true, and β will say something not true on Tuesday. Since β only says (B), (B) is not true. So if (B) is true, then it is not true. Suppose (B) is not true; then α said something true on Monday. Since α only said (A), (A) is true, that is, everything β will say on Tuesday is true. This includes (B), so (B) is true. Thus if (B) is not true, it is true.

Neither of the sentences in the story literally refers to itself. Rather, there is a kind of circle, so perhaps we should talk of "circular reference" rather than self-reference. However, as the circularity does not strictly involve reference at all, but rather quantification, it might be safer still just to speak of circularity.

We could expand the story of (A) and (B) by imagining a third utterance:

(C) (said by γ on Tuesday): Nothing α said on Monday is true.

The fact that β and γ use the very same sentence, yet only one of them is circular in the relevant way, shows that circularity is not a property of sentences as such. Being meaningful or meaningless is a property of sentences. Since there is nothing paradoxical about (C), there is no reason to say it is other than meaningful, and since (B) is the same sentence, it follows that the property which circularity prevents is not that of being meaningful. We need a more refined notion, one sensitive to the use to which a sentence is put on a specific occasion. Such a notion emerges naturally from a consideration of indexicality; and the remaining two responses to the Liar which I shall consider both claim to discover, in reasoning related to the Strengthened Liar, some element of

indexically. One response locates the indexicality in the specific kind of self-reference involved in the Liar; the other, briefly mentioned in the last paragraph of section 6.8, locates it in the predicate "true."

To summarize: if we are to finger self-reference as the villain of the piece, the relevant kind of self-reference must contain an indexical element. However, once indexicality is allowed, we also open the way to a Tarski-like hierarchy of levels, triggered by indexical features of "true."

6.7 Indexicality

For reasons independent of Liar paradoxes, it is necessary to distinguish between sentences, regarded as things which can be uttered by different people and on different occasions, and the things which people can say or express by using sentences. The reason is the "indexicality" of language: the fact that the same words may, without exploiting ambiguity, be used on different occasions to say different things. Indexicality in pronouns provides a familiar example: if you use the sentence "I am hungry" affirmatively, you say one thing, and if I use it affirmatively I say another. The things said are different because it could be that what you say is true whereas what I say is false.

I will use *statement* for what a sentence is used on a specific occasion to say or express. Indexicality shows that it is only statements and not sentences that can properly be said to be true or false. I shall assume that bivalence holds for statements, so that every statement is either true or false. We have already in effect seen that a sentence can be meaningful, yet on a specific occasion be used in such a way as to fail to make a statement ("That elephant is about to charge"). Although sentences can be self-referential, or more generally can have the kind of circularity associated with paradox, it may be that statements cannot. Thus, reverting to the example in section 6.6, we might be able to justify the claim that whereas both β and γ use the same sentence, only γ thereby succeeds in making a statement. The notion of a statement thus seems to have the features we were looking for: it is a function not only of the meaning of a sentence, but also of the use to which it is put on a specific occasion.

The Strengthened Liar needs to be adapted to the distinction between sentence and statement. One way to do so is as follows:

L_2^* : L_2^* does not express a true statement.

Reconsideration of the reasoning involved in the Strengthened Liar supports the view that some kind of indexicality is at work. We contemplate L_2^* and regard it as defective. When we come to express this, we may do so by words which are, or entail, L_2^* itself, for example: " L_2^* is semantically defective, so (*a fortiori*) it does not express a true statement; only the

use of a non-defective sentence could do that." Intuitively this at first glance seems perfectly sensible (until we realize that we have ourselves re-used the very words we wish to say are defective). This intuition could be vindicated if we could show that the same words, even referring to the same thing, and applying the same predicate to it, may not say the same thing on two occasions of use. We want to say that the first use of L_2^* is defective, but the second use of those very words is not, since they are then used to express a truth.

The general feasibility of such an approach is suggested by considerations like the following. Suppose that the displayed sentence is the only sentence written on the board in room 101:

The sentence written on the board in room 101 does not express a true statement.

It appears perfectly consistent for me to write on this page that, because of some semantic defect, the sentence written on the board in room 101 does not express a true statement. I use the words which, as written on the board in room 101, are defective, in circumstances in which there is nothing defective about their use. This suggests that the same words, used to refer to the same thing, and applying the same predicate to it, do not necessarily make the same statement. The sentence written on the board in room 101 makes no statement, in its use in room 101; whereas I use those words, on this page, to make a true statement. I did not have to use the same words. Under suitable circumstances I could simply have said "That sentence does not express a true statement." That there are special circumstances under which I can reuse the same words is an accident of our use of language, and does not affect the truth of the statement I wish to make.

We are still some way from our goal, for three tasks remain: (i) to provide a more detailed account of what the problematic circularity is; (ii) to give some independent justification for saying that a statement cannot possess it; and (iii) to return to the problems posed by the Strengthened Liar.