Truth IV: The Liar

Consider the sentence:

(1) Sentence (1) is false

That gives us a puzzle. But Tarski (and many of his contemporaries) thought it was much worse than that: they were concerned that a language containing a truth predicate was actually *inconsistent*. How can a *language* be inconsistent? Consider Prior's connective 'tonk' which has the introduction rule of disjunction, and the elimination rule of conjunction. So from *p*, one can go to *p tonk* ~*p*, and from there to ~*p*. Adding 'tonk' to the language makes the language inconsistent, in that one can immediately derive a contradiction: from *p*, ~*p* follows. Tarski thought that the truth predicate did much the same thing, though of course we need to add some rules to see this. Rather than doing this with introduction and elimination rules, Tarski did it with a biconditional, or more accurately a schema for a biconditional. He thought that an understanding of the truth predicate in ordinary English commits one to accepting each instance of:

The T-schema: 'p' is true iff p

where 'p' is replaced with a sentence of English in each occurrence, so in each case the first instance is quoted, and the second is used. (Actually Tarski did this with structural descriptions rather than quotation, but this won't matter for our purposes.) Then we have the following argument:

P1: Sentence (1) is 'Sentence (1) is false'. (from (1))
P2: 'Sentence (1) is false' is true iff sentence (1) is false. (instance of the *T-schema*)
C1: Sentence (1) is true iff sentence (1) is false. (instantiating identity from P1 into left side of P2)
C2: Sentence (1) is true and sentence (1) is false. (Assuming that (1) is either true or false, and deriving consequences from C1.)

Which assumptions did we need to get that? Not many: just the universal applicability of the T-schema, which is exactly what a grasp of the notion of truth seemed to require, elementary application of a principle of substitution, and bivalence: the claim that every sentence is either true or false.

Perhaps we could reject bivalence and say that (1) is neither true nor false, though that is probably a bad way to put it for reasons we've seen. Perhaps it has some third status. Let's just say it's *bad*. But then we get the Strengthened Liar:

(2) Sentence (2) is either false or bad.

We can now run though the same reasoning, provided that, in place of the principle of bivalence, we have the principle that every sentence is either true, false, or bad.

P1: Sentence (2) is 'Sentence (2) is either false or bad'. (from (2))
P2: 'Sentence (2) is either false or bad' is true iff sentence (2) is false or bad. (instance of the *T-schema*)
C1: Sentence (2) is true iff sentence (2) is either false or bad (instantiating identity from P1 into left side of P2)
C2: Sentence (2) is true and sentence (2) is either false or bad (Assuming that (2) is true or either false or bad, and deriving consequences from C1.)

TARSKI'S APPROACH: LANGUAGES AND METALANGUAGES

Tarski thought that ordinary languages like English are hopeless. The only hope is to construct new formal languages which don't risk paradox. There his central device for avoiding paradox was to make a distinction between the object language—the language containing the sentences *of which* truth is predicated—and the metalanguage—the language *in which* the truth predication is made. The truth predicate is in the metalanguage, not in the object language. Object language/metalanguage suggests a simple binary division. But it is really a relative notion. So we can think in terms of a hierarchy of languages, each of which contains the truth predicates of the languages below it, but not its own.

Traditionally liar sentences involve self reference. e.g.

(A) This sentence is not true

But truth must be relative to a language. Suppose that (A) is in the language L_1. So, trying to capture the intuitive sense of (A) we might reformulate it as:

(A*) This sentence is not true in L_1

But the truth predicate for L_1 is not part of L_1. So sentence (A*) is simply ungrammatical. Alternatively, the truth predicate might be for the language below it:

(A**) This sentence is not true in L_0

But now (A**) is true, i.e. true in L_1, since it is not in L_0, and so *a fortiori* is not one of the true sentences of L_0.

PARADOX WITHOUT SELF-REFERENCE

Paradoxical sentences need not involve immediate self-reference; but Tarski's approach can certainly sometimes help here. Consider:

(B) Sentence (C) is true
(C) Sentence (B) is not true.

But it can't be that each of these sentences is in the metalanguage for the other. So at least one of them must be ungrammatical. For a harder case, consider Yablo's paradox, the infinite sequence of sentences:

S_0 All later sentences in this sequence are not true
S_1 All later sentences in this sequence are not true
S_2 All later sentences in this sequence are not true ...

This gives rise to paradox if all of the sentences contain the same truth predicate. (Suppose the first sentence is true; then the second must be false, but since it is entailed by the first, it can't be. Suppose then that the first sentence is false; then at least one of the sentences below it must be true; but now reapply the earlier reasoning to that sentence.) (See Yablo, 'Circularity and Paradox')

What happens if we use a hierarchy of truth predicates in Tarski's fashion. Suppose that, going down the list, each sentence is in a language one level lower than the one before, and each truth predicate is correspondingly for a language one level lower than the one it is formulated in. Can they now get a consistent assignment of truth values? Apparently not for the same reasons as before. Tarski doesn't just need a hierarchy of truth predicates; he also needs there to be a lowest level.

Kripke's Criticisms

Tarski insisted on a family of languages, each containing a truth predicate for those below it. So if one wants to know whether a given sentence containing a truth predicate is true or false (and in many cases, if one wants to know if it is even grammatical) one will need to know which language it is in.

This is highly unnatural. Further, as Kripke points out, it isn't really workable. The problem comes from the fact that paradox is not generated just by self-reference but also by reference to other sentences. Consider again the two sentences

(1) Sentence (2) is true
(2) Sentence (1) is false.

How does Tarski solve this? By insisting that the truth predicates cannot both be in a language higher than the other, so that at least one of the sentences must be paradoxical.

But suppose that we are in an argument, and I shout at you:

Hardly anything that you have said is true

and you shout back

Hardly anything that you have said is true

We will both want our truth predicates to be higher than the others, since whoever gets the higher will render their opponent's claim ungrammatical. So we might each try to make ours higher (by muttering a level under our breath?). But that's ridiculous. And moreover, intuitively it seems that both of our assertions could be true or false; perhaps we had both been lying almost all of the time, in which case both of our assertions should be true.

Or consider Kripke's own example.

> Dean says: Most of Nixon's Watergate sentences are false
>
> Nixon says: Most of Dean's Watergate sentences are true

Tarski would say that one of these must be ungrammatical, but in most circumstances they would give no risk of paradox; indeed they might easily be true. They would only give rise to paradox in exceptional circumstances: suppose up till now exactly 50% of each of their Watergate sentences were true, and 50% false. Then they would be like (1) and (2).

KRIPKE'S ACCOUNT

Kripke's main idea is that we can stick with a single truth predicate, if we let each of the utterances find its own level by an inductive procedure. Rather than having a hierarchy of languages, we can have a hierarchy of levels, at each of which we can add instances of the truth predicate. But it is the same truth predicate at each level. Moreover, if we use an axiomatic approach, rather than defining truth explicitly, we do not need a distinction between the object language and the metalanguage. (The need for an axiomatic approach here comes from the fact, shown by Tarski, that any moderately powerful language that tries to define its own truth predicate will be inconsistent. Kripke himself did give a definition of truth, and so worked with a separate object language and metalanguage, but there are axiomatic developments of his approach that do not; for discussion see Volker Halbach's piece on axiomatic theories of truth in the *Stanford Encyclopedia*). To keep things simple, first consider just the atomic sentences; we'll come to the compound sentences later.

From the class of all the atomic declarative sentences that don't contain a truth predicate, take those that you would be prepared to assert (there may be infinitely many. Construct a new set containing just them. Then extend that set by adding to it each sentence formed by predicating truth of each of the sentences it already contains; extend it again by predicating truth of those sentences; and so on. Intuitively that set contains only true sentences. At each step add in from the original class all those sentences that do contain truth predicates that are made true by the growing set of true sentences. So if you have added 'Grass is green' and 'Hilary said that grass is green' to the set of true sentences, you can now add 'Hilary said something true'. And so on.

Now do a parallel thing to construct the class of false sentences. Go back to the set of atomic sentences that don't contain a truth predicate. This time, take those that you would be prepared to deny and make a set of them. Predicate truth of each them and add the resulting sentences to the set, and so on. Intuitively that set contains only false sentences (Kripke

includes in this class the things that are not sentences, since they are not true; but this adds some complexity, so we'll just ignore the non-sentences).

Then we can make moves between the two classes: predicate falsity of a sentence in one class and then add that to the other class (alternatively predicate not-truth: we make no distinction between being false and being not true). Once a sentence has been added to either of the classes in this way, it will remain there no matter how many other sentences are added: the process has the property of monotonicity. And provided that an initial class of atomic sentences was consistent (as we might hope is the case with the true ones) it will remain consistent. Continue this process so that every sentence of the language that is made true (or false) by the sentences already in the true (or false) class to that class; for instance, that may be the case with the Nixon sentences. (We'll come back to how you do this for the compound sentences shortly.) Obviously this process will never stop: you can always going on predicating truth of sentences that you have predicated truth of. But Kripke proves that there is a consistent and complete way of doing this to construct classes with infinitely many members. Kripke calls such a pair of classes, one for the extension of the truth predicate (the true sentences) and one for the anti-extension of the truth predicate (the false senteces), a 'fixed point'.(See the Gauker piece for a clear outline of how this works.)

At the *least* fixed point (i.e. the fixed point which is contained by all the other fixed points — Kripke also proves that there is such a thing) the paradoxical sentences like (1) and (2), and also paradoxical sentences like (3):

(3) Sentence (3) is not true

will not be in the either of the classes, that is, they will not be in the extension of the truth predicate, nor in its antiextension. The same will hold for 'truth teller' sentences like

(4) Sentence (4) is true

All such sentences will be, in Kripke's term *ungrounded*. Whether or not the Nixon/Dean sentences will be there will depend on the empirical facts of the case.

We won't get into trouble if we add (4) to the class of true sentences of our minimal fixed point, and then predicate truth of it: that will give rise to a new fixed point, though the addition may seem arbitrary, since we could equally add it to the class of false sentences. In contrast, if we try to add (3), we won't arrive at another fixed point.

To recap: the effect of this will be to effectively define an extension for the truth predicate (all of the sentences that occur in the minimal fixed point with truth predicated of them), and an anti-extension for the truth predicate (all of those sentences that have falsity or not-truth predicated of them). But not every grammatically well-formed sentence occurs in one or other of the sets. There is a third set: the set of sentences that occur neither in the truth class, nor in the falsity class. It is tempting to say that these sentences are neither true nor false, (i. e. that they are not true, and that they are not false) but for familiar reasons we can't do that, at least not in the object language; the sentences that are not true are those that are in the anti-extension of the truth predicate, but these sentences aren't there. Kripke remarks that we

might be able to say they are neither true not false if we move up a language: say it in the meta-language, not the object language; to that extent, he thinks that the 'ghost of the Tarski hierarchy' remains. But alternatively, we might want to simply resist saying anything about the truth status of the third class.

(If we think that there are independent reasons for thinking that some sentences are not truth-apt—as a result of presuppositional failure etc., as discussed earlier—we might put them into the third class right from the start; however, Kripke doesn't discuss this.)

The truth predicate thus ends up as a 'gappy' predicate. Following Soames, we might compare it to an artificial predicate like 'magnaped' whose extension and anti-extension are fixed by the following stipulative definition:

Someone who has a shoe size of 11 or greater is a magnaped;
Someone who has a shoe size of less than 10 is not a magnaped.

What can we say about someone who has a shoe size of 10.5? Clearly they aren't in either the extension or the anti-extension. But can we say that they are neither a magnaped nor not a magnaped? Apparently not without fear of contradiction. The same applies to truth.

How do we now handle the non-atomic sentences, i.e. those constructed from the atomic ones using the standard connectives?. Once we think of the truth predicate as a partial predicate, we have two possibilities. We might use a three-valued logic, or we might use a supervaluational approach. Kripke focuses on the first, using strong Kleene tables, but briefly discusses the second.

The Strong Kleene three-valued truth-table for disjunction

v	T	–	F
T	T	T	T
–	T	–	–
F	T	–	F

for negation:

~	
T	F
–	–
F	T

On this approach, a true sentence that is disjoined with the liar will be true.If we take the Kleene approach we will clearly have to deny that (p or not-p) is a logical truth. If we instead use supervaluations we might hope to maintain that as a logical truth, while not accepting that every sentence is either true or false; but that will very much depend on the details.