The liar paradox

It seems that some sentences refer to themselves. For example, “This sentence is the one that I am uttering now'” seems to make sense, and be true.

But other sentences which refer to themselves are not so harmless. Consider the following sentence:

> L1. L1 is false.

This sentence says of itself that it is false. But this seems to lead to the following line of argument:

> Suppose that L1 is true. Then what it says must be the case; so it must be the case that L1 is false.
> So, if L1 is true, then it is false.
>
> Suppose that L1 is false. This is what L1 says is the case; since if what a sentence says is the case is in fact the case the sentence is true, if L1 is false, then L1 must be true.

So L1 has the peculiar property that it is true if false and false if true. Since nothing can be **both** true and false, it follows that L1 is **neither** true nor false.

This is not by itself paradoxical. But it may seem somewhat surprising --- it can seem as though every declarative sentence must either describe the world accurately or not describe it accurately; it is true in the former case, and false in the latter. So what we might want is a theory about how “true” works which **explains** how sentences like L1 can turn out to be neither true nor false.

One attempt to do this is via the notion of a sentence being **grounded**. Suppose we want to figure out how the word “true” works --- you can imagine that you need to explain it to someone to understands English, but just doesn't understand the word “true.” You might give them the following general rule: just in case they're willing to endorse a sentence “S”, then they should also be willing to endorse the sentence “S is true.” They might use this rule to decide what sentences involving the word “true” to accept. For example, if they are willing to endorse, or accept

> Grass is green.

they should also be willing to endorse

> “Grass is green” is true.

L1. L1 is false.

So L1 has the peculiar property that it is true if false and false if true. Since nothing can be **both** true and false, it follows that L1 is **neither** true nor false.

This is not by itself paradoxical. But it may seem somewhat surprising --- it can seem as though every declarative sentence must either describe the world accurately or not describe it accurately; it is true in the former case, and false in the latter. So what we might want is a theory about how "true" works which **explains** how sentences like L1 can turn out to be neither true nor false.

One attempt to do this is via the notion of a sentence being **grounded**. Suppose we want to figure out how the word "true" works --- you can imagine that you need to explain it to someone to understands English, but just doesn't understand the word "true." You might give them the following general rule: just in case they're willing to endorse a sentence "S", then they should also be willing to endorse the sentence "S is true." They might use this rule to decide what sentences involving the word "true" to accept. For example, if they are willing to endorse, or accept

Grass is green.

they should also be willing to endorse

"Grass is green" is true.

Conversely, since they reject

Grass is red.

they will not be willing to endorse, or accept

"Grass is red" is true.

In fact, they will know to reject it. The might also then master the rule which tells them that if they are willing to reject a sentence "S", they should be willing to endorse the sentence "S is false."

This will give them a lot of information about how to use the words "true" and "false." But it will not tell them which to apply to every sentence of the language. Consider, for example, the "truthteller" sentence:

L1. L1 is false.

So L1 has the peculiar property that it is true if false and false if true. Since nothing can be **both** true and false, it follows that L1 is **neither** true nor false.

This is not by itself paradoxical. But it may seem somewhat surprising --- it can seem as though every declarative sentence must either describe the world accurately or not describe it accurately; it is true in the former case, and false in the latter. So what we might want is a theory about how “true” works which **explains** how sentences like L1 can turn out to be neither true nor false.

One attempt to do this is via the notion of a sentence being **grounded**. Suppose we want to figure out how the word “true” works --- you can imagine that you need to explain it to someone to understands English, but just doesn't understand the word “true.” You might give them the following general rule: just in case they're willing to endorse a sentence “S”, then they should also be willing to endorse the sentence “S is true.” They might use this rule to decide what sentences involving the word “true” to accept.

This will give them a lot of information about how to use the words “true” and “false.” But it will not tell them which to apply to every sentence of the language. Consider, for example, the “truthteller” sentence:

T1. T1 is true.

If our imaginary language learner looks at T1 to decide whether to accept or reject it, he first notices that the word “true” occurs in the sentence; so, following his rule, he looks at the sentence to which “truth” is applied and asks whether he accepts **that** sentence. But this sentence itself involves the notion of truth ... and so we never get an answer to the question of whether we should accept or reject T1. When this is the case, we say that a sentence is **ungrounded**. Ungrounded sentences, on this view, are neither true nor false -- since the rules which govern “true” and “false” simply do not give a result when applied to them.

It comes as no surprise that L1 is also ungrounded. This may explain why it is neither true nor false -- which is exactly the result we need to block the derivation of a contradiction from L1.

This looks great as far as it goes; but we are not out of the woods yet. Consider the following sentence, which is called the “strengthened liar':

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.

This looks great as far as it goes; but we are not out of the woods yet. Consider the following sentence, which is called the “strengthened liar':

SL1. SL1 is not true.

SL1 is ungrounded, so on the above view it is neither true nor false; in other words, it is not true and not false. So, in particular, SL1 is not true. But this is just what SL1 says is the case, so it must be true. So it looks like the attempt to solve the Strengthened Liar via the view that ungrounded sentences are neither true nor false is a failure.

Maybe we should change the view in order to provide an account of SL1 as follows: perhaps we should say that ungrounded sentences are not just neither true nor false, but also neither true nor not true.

But this modified solution itself faces two apparently decisive objections:

➡ To say that SL1 is neither true nor not true is to say that it is both not true and not not true. But that is a contradiction.

➡ To say that SL1 is neither true nor not true is to say that it is both not true and not true. But that means that it involves saying that SL1 is not true; which is just what SL1 says, which means that SL1 is true.

These problems might suggest that we need a completely different sort of solution to the Liar paradox. An alternative to the grounding approach is the view of truth defended by the midcentury Polish logician, Alfred Tarski.

Tarski thought that three features of natural languages like English give rise to the Liar paradox:

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.

These problems might suggest that we need a completely different sort of solution to the Liar paradox. An alternative to the grounding approach is the view of truth defended by the midcentury Polish logician, Alfred Tarski.

Tarski thought that three features of natural languages like English give rise to the Liar paradox:

1. L contains the resources for stating facts about the truth or falsity of its own sentences. Tarski calls this L being "semantically closed."

2. L contains the capacity to refer to its own expressions.

3. For every meaningful declarative sentence S, the "T-sentence" formed using S and a name of S as follows

   "S" is true if and only if S.

is true.

Since these three assumptions about a language lead to a contradiction, we must reject one of them. The problem is that each of the three seem very plausible. In particular, it is very plausible that English meets all three conditions.

Tarski's conclusion is that we must reject the first of the three assumptions: "Accordingly, we decide *not to use any language which is semantically closed* in the sense given." According to Tarski, no language can contain a word "true" which can apply to its own sentences.

So how should we understand claims involving the word "true'? Strictly speaking, we should think of them as belonging to a different language. Let's call the set of sentences which can be formed by English words other than "true" and "false" the language "L1". Then we can introduce a word, "true1", which applies to the true sentences of L1. However, "true1" is not itself a word of L1, so it cannot be a part of sentences of L1. Sentences which involve words of L1, plus true1, will therefore belong to a new language - L2. Since "true1" is a word of this language, it of course cannot be applied to sentences of L2 - for that, we need another predicate, "true2." And so on.

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.

Tarski's conclusion is that we must reject the first of the three assumptions: “Accordingly, we decide *not to use any language which is semantically closed* in the sense given.” According to Tarski, no language can contain a word “true” which can apply to its own sentences.

So how should we understand claims involving the word “true'? Strictly speaking, we should think of them as belonging to a different language. Let's call the set of sentences which can be formed by English words other than “true” and “false” the language “L1”. Then we can introduce a word, “true1”, which applies to the true sentences of L1. However, “true1” is not itself a word of L1, so it cannot be a part of sentences of L1. Sentences which involve words of L1, plus true1, will therefore belong to a new language - L2. Since “true1” is a word of this language, it of course cannot be applied to sentences of L2 - for that, we need another predicate, “true2.” And so on.

This might be called the hierarchical solution to the Liar paradox, since the idea is that we avoid the paradox by adopting a language which contains a hierarchy of truth predicates, each of which applies only to sentences of the language immediately below it in the hierarchy.

How would this help with the Liar paradox? Consider the word “true” in SL1. Which truth predicate would this be? It seems that it cannot be any truth predicate in the hierarchy. Suppose that it is “true8”. Then SL1 must be a sentence of L8, since “true8” can only be applied to sentences of this language. But then “true8” can’t be a part of SL1 after all, since **it** is not a part of L8 - it is a word in L9. Hence, in our new language, sentences like SL1 simply cannot be formulated.

But what about our old language - the one we have been speaking all along? We were not using subscripts on “true” when we were speaking that language --- so how do we avoid the conclusion that, when in that language we formulated SL1, it was both true and not true?

Tarski’s view was that natural languages like English are inconsistent. But it’s not obvious what this would mean, for two reasons: (i) It’s hard to see how a language, as opposed to a theory formulated in a language, could be inconsistent. (ii) If English is inconsistent, this would suggest that there are sentences of English to which the English words “true” and “not true” both apply. But this is hard to believe.

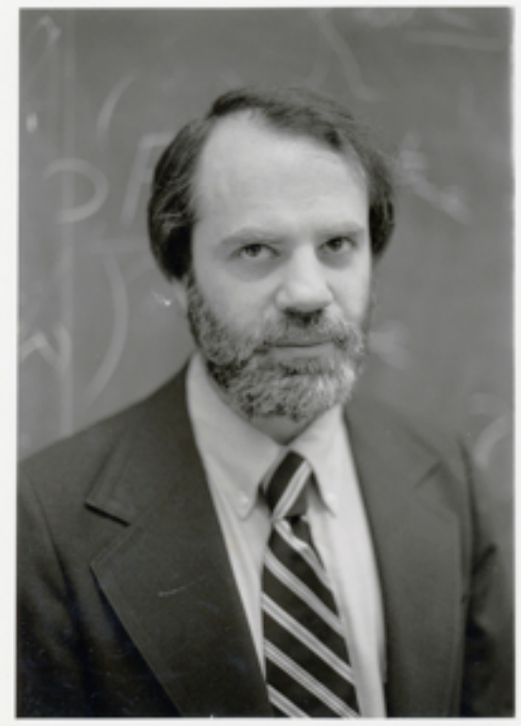
A natural way to adapt Tarski’s views to English *without* taking English to be inconsistent is that although we did not actually write down the relevant subscripts, still we were using a single word - “true” - to express many different concepts - those corresponding to the various languages in the hierarchy. Sometimes we used “true” to mean “true8” - just those times in which we were applying the word to a sentence of L7. On this view, the “true” of ordinary language “automatically” assumes the right level.

But there are two problems with this extension of Tarski’s hierarchical resolution of the Liar, both of which were emphasized by Saul Kripke in his 1975 paper, “Outline of a theory of truth.”

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.



But there are two problems with this extension of Tarski's hierarchical resolution of the Liar, both of which were emphasized by Saul Kripke in his 1975 paper, "Outline of a theory of truth."

First, Kripke pointed out, there's something unrealistic about the application of this sort of hierarchical view to natural language, since often speakers will have no idea which truth predicate they are supposed to be using. He considers the example of a speaker, Jones, saying

(1) Most of Nixon's assertions about Watergate are false.

and says, of the hierarchical theory,

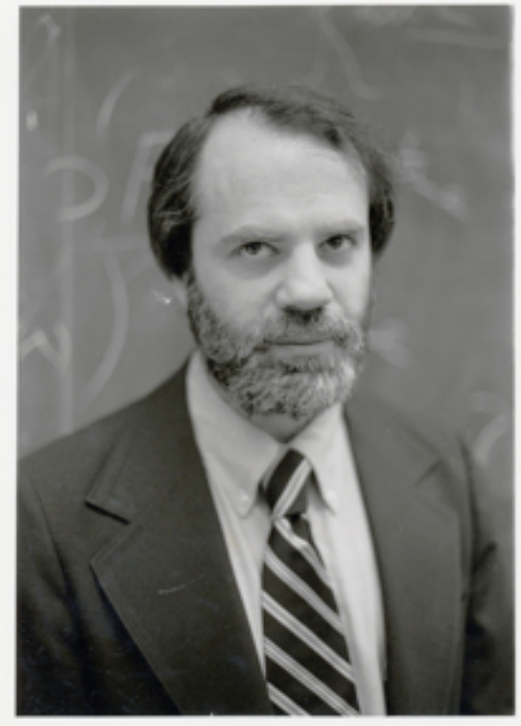
> Unfortunately this picture seems unfaithful to the facts. If someone makes such an utterance as (1), he does *not* attach a subscript, explicit or implicit, to his utterance of 'false', which determines the "level of language" on which he speaks. An implicit subscript would cause no trouble if we were sure of the "level" of *Nixon's* utterances; we could then cover them all, in the utterance of (1) or even of the stronger
>
> (4) All of Nixon's utterances about Watergate are false.
>
> simply by choosing a subscript higher than the levels of any involved in Nixon's Watergate-related utterances. Ordinarily, however, a speaker *has no way of knowing the "levels" of Nixon's relevant utterances*. Thus Nixon may have said, "Dean is a liar," or "Haldeman told the truth when he said that Dean lied," etc., and the **"levels" of these may yet depend on the levels of Dean's utterances, and so on. If the speaker is forced to assign a "level" to (4) in advance [or to the word 'false' in (4)], he may be unsure how high a level to choose; if, in ignorance of the "level" of Nixon's utterances, he chooses too low, his utterance (4) will fail of its purpose.**

Second, and more importantly, sometimes there is no level of truth predicate which can be consistently assigned to a use of the word "true":

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.



But there are two problems with this extension of Tarski's hierarchical resolution of the Liar, both of which were emphasized by Saul Kripke in his 1975 paper, "Outline of a theory of truth."

Second, and more importantly, sometimes there is no level of truth predicate which can be consistently assigned to a use of the word "true". To show this, Kripke again uses the example:

(4) All of Nixon's assertions about Watergate are false.

> Another situation is even harder to accommodate within the confines of the orthodox approach. Suppose Dean asserts (4), while Nixon in turn asserts
>
> (5) Everything Dean says about Watergate is false.
>
> Dean, in asserting the sweeping (4), wishes to include Nixon's assertion (5) within its scope (as one of the Nixonian assertions about Watergate which is said to be false); and Nixon, in asserting (5), wishes to do the same with Dean's (4). Now on any theory that assigns intrinsic "levels" to such statements, so that a statement of a given level can speak only of the truth or falsity of statements of lower levels, it is plainly impossible for both to succeed: if the two statements are on the same level, neither can talk about the truth or falsity of the other, while otherwise the higher can talk about the lower, but not conversely. Yet intuitively, we can often assign unambiguous truth values to (4) and (5). Suppose Dean has made at least one true statement about Watergate [other than (4)]. Then, independently of any assessment of (4), we can decide that Nixon's (5) is false. If all Nixon's other assertions about Watergate are false as well, Dean's (4) is true; if one of them is true, (4) is false.

It thus seems that the hierarchical theory has trouble making sense of our ordinary use of the word "true."

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.

It thus seems that the hierarchical theory has trouble making sense of our ordinary use of the word “true.”

This seems to leave us without anything plausible to say about the strengthened liar, SL1. Let’s return to the idea that the solution to this paradox has something or other to do with the idea that SL1 is ungrounded.

The initial application of this idea is that, if S is ungrounded, that we should deny both that S is true and that S is false. This, as we observed, is no help at all with SL1 (even if it does help with L1). So we imagined extending this idea to say that if S is ungrounded, that **we must deny both that S is true and that S is not true**.

Above we considered two arguments against this sort of view:

➡ To say that SL1 is neither true nor not true is so say that it is both not true and not true. But that is a contradiction.

➡ To say that SL1 is neither true nor not true is so say that it is both not true and not true. But that means that it involves saying that SL1 is not true; which is just what SL1 says, which means that SL1 is true.

Both of these arguments rest on the following assumption: **if one denies S, one must affirm that S is not true**. But perhaps this assumption could be rejected; perhaps we can reject, or deny, a sentence without affirming that it is not true. In this case we would reject each of the following:

SL1 is true.
SL1 is not true.

Without affirming either of:

SL1 is not true.
SL1 is not not true.

Could this be the key to the problems posed by SL1?

SL1. SL1 is not true.

L1. L1 is false.

T1. T1 is true.

It thus seems that the hierarchical theory has trouble making sense of our ordinary use of the word “true.”

This seems to leave us without anything plausible to say about the strengthened liar, SL1. Let’s return to the idea that the solution to this paradox has something or other to do with the idea that SL1 is ungrounded.

The initial application of this idea is that, if S is ungrounded, that we should deny both that S is true and that S is false. This, as we observed, is no help at all with SL1 (even if it does help with L1). So we imagined extending this idea to say that if S is ungrounded, that **we must deny both that S is true and that S is not true**.

Could this be the key to the problems posed by SL1?

There are at least two objections to this idea. First, it is not clear what it means to reject, or deny a sentence if this does not involve the claim that the sentence in question is not true.

Second, it is very tempting to think of the present view as including the claim that SL1 is ungrounded. This seems to indicate that the predicate “is ungrounded” is an intelligible expression of our language. But now consider the following revised strengthened liar:

SL2. SL2 is not true or is ungrounded.

Since SL2, like SL1, is ungrounded, it appears that SL2 is true. But this can’t be, since the whole idea behind our use of the notion of “groundedness” is that **no sentence which is ungrounded can be true**.

This is known as the problem of **revenge**, or the **revenge liar**. Intuitively, the idea is this. We try to solve the liar paradox by saying that the problematic sentences all have some characteristic in common - in our case, this characteristic was ungroundedness. But we can then formulate a sentence which says of itself that it is either not true or has that special characteristic. Since, by hypothesis, this sentence will have that characteristic, it will be true - but this undercuts the core idea that nothing with that characteristic could be true.

In a way, this pushes us back to a view which is somewhat like Tarski’s: we must say that the “special characteristic”, whatever it is, is **inexpressible in our language**. Appearances to the contrary, “is ungrounded” cannot be an intelligible expression of English. But doesn’t it seem to be an intelligible expression?