New Waves in Truth. Edited by Cory Wright and Nicolaj Pedersen. New York: Palgrave Macmillan, 2010.

Falsity

KEVIN SCHARP

The Ohio State University

ABSTRACT: Although there is a massive amount of work on truth, there is very little work on falsity. Most philosophers probably think this is appropriate; after all, once we have a solid understanding of truth, falsity should not prove to be much of a challenge. However, there are several interesting and difficult issues associated with understanding falsity. After considering two prominent definitions of falsity and presenting objections to each one, I propose a definition that avoids their problems.

0. Introduction

Although there is a massive amount of work on truth, there is very little work on falsity.[1] Most philosophers probably think this is appropriate; after all, once we have a solid understanding of truth, falsity should not prove to be much of a challenge. However, there are several interesting and difficult issues associated with understanding falsity. The first is that there are several ways philosophers and logicians have defined falsity, and they are incompatible. Moreover, there are a number of unnoticed problems with even the most popular definitions.

[1] We use 'truth' in English as a count noun (e.g., 'The greatest enemy of any one of our truths may be the rest of our truths.'—William James) and as a mass noun (e.g., 'Truth springs from argument amongst friends'—David Hume). 'Falsity' is the antonym of the mass noun, while 'falsehood' is the antonym of the count noun.

1. Three Definitions

There are at least three obvious ways to define falsity:

(i) $\ulcorner\sigma\urcorner$ is false $=_{df}$ $\ulcorner\sigma\urcorner$ is not true.

(ii) $\ulcorner\sigma\urcorner$ is false $=_{df}$ $\ulcorner\sim\sigma\urcorner$ is true.

(iii) $\ulcorner\sigma\urcorner$ is false $=_{df}$ $\sim\sigma$.[2]

A few comments about these definitions are in order. 'σ' is used as a sentential variable, not as a name; that is, one can substitute a sentence for 'σ' in these definitions and arrive at a well-formed sentence. '⌈' and '⌉' are half brackets—they are used in cooperation with sentential variables as quotation devices; if one substitutes a sentence in for 'σ', then one should substitute the quote-name in for '⌈σ⌉' in the above definitions.

As the '$=_{df}$' expression indicates, these are *definitions*. For example, (ii) implies that the sentence ''snow is black' is false' means that 'snow is not black' is true. One might reject the view that falsity can be defined (many philosophers take this view with respect to truth); if so, then one will reject all three of these definitions.

Sentences are serving as truth bearers in these definitions, which is a matter of convenience; I discuss this choice below. We attribute truth and falsity to many different kinds of things, and it is far from clear how to make sense of this practice.

A negation sign '~' also appears in two of the definitions. It is intended as a regimentation of the natural language term for negation (e.g., 'not'), which would appear in the natural language formulation of these three falsity definitions. Its interpretation becomes a matter of dispute below. As a bit of regimentation, I often use 'not' in sentences being used to call something false and negation symbols (e.g., '~') in sentences being called false.

---

[2] For examples of (i) and (iii) see Horwich (1998: 71-72). For an example of (ii) see Field (2008).

The motivation for each of the three definitions is relevant. (i) is probably the most common definition given (when one is given) outside the tradition of philosophical logic. The idea behind it is simple: truth and falsity are antonyms or opposites—something is true if and only if it is not false, and something is false if and only if it is not true. (ii) is most common in formal work in philosophical logic. It rarely, if ever, receives an explicit justification; one goal of section two is to illustrate its utility. (iii) is deflationary in spirit; it is of a piece with recent work on the nature of truth that denies that truth has a substantive nature or an illuminating analysis. If my characterization of these motivations is accurate, then it is an unfortunate consequence of the fact that the literature on truth cleaves into two relatively independent traditions (one on the nature of truth, the other on the liar paradox) that philosophers unwittingly mean different things by the term 'false'. A goal of this paper is to expose and remedy this problem.

One thing to note about the difference between the three definitions is their implications for falsity attributions. When someone asserts that some sentence p[3] is false, she is: (i) asserting a sentence with the logical form of a negation, and the sentence negated attributes truth to p, (ii) asserting a sentence with atomic logical form that predicates truth of something other than p, namely the negation of p, or (iii) asserting a sentence with the logical form of negation, where the sentence negated is p. Loosely speaking, on (i) one attributes a property (the negation of truth) to p; on (ii) one attributes a property (truth) to something (the negation of p); on (iii) one need not be attributing anything at all, since by calling a sentence false, one merely asserts the negation of that sentence. On (i) ''snow is white' is false' is a metalinguistic statement about the sentence 'snow is white'; on (ii) ''snow is white' is false' is a metalinguistic statement about a different sentence, 'it is not the case that snow is white'; on (iii) ''snow is white' is false' is a

[3] 'p' is an individual constant (e.g., name) here.

linguistic statement about snow, presumably (assuming the negation of a sentence about X is about X). .

For the purposes of this investigation, (iii) will be set aside. The motivation is that I do not think semantic content can pull logical form that far away from sentential form. Consider some examples. 'Waylon loves Monty' is false' has the logical form of '~Lwm', but the sentential form of 'Fs'.[4] If (iii) is correct, then no matter how complicated the target, the falsity attribution has atomic sentential form, but the same logical form as the target. Even worse is the fact that different falsity attributions would have the same sentential form, but different logical forms.

Perhaps one could live with that divergence between sentential form and logical form. However, (iii) also implies that the property of having a logical form diverges from the property of having a sentential form. For example, ''How big is the universe?' is false' seems like it should be grammatical—it has a sentential form. Yet if (iii) is correct and the sentence has a logical form, then it has the logical form of the negation of a question. Since questions do not have negations, there is no such logical form for it to have. ''How big is the universe?' is false' should count as either false or a truth-value gap (i.e., *gappy*) depending on some issues that will arise later. Yet if (iii) is right, then it should have what would be the logical form of a string of letters that is not even a grammatical sentence (e.g., 'It is not the case that how big is the universe?'), if only there were such a logical form. Thus, we have to conclude that if (iii) is correct, then ''How big is the universe?' is false' does not even have a logical form despite the fact that it is clearly a well-formed declarative sentence.

Matters are even worse for sentences like 'the acorn is false'. Here we have a falsity attribution, but there is not even a linguistic entity to negate (assume 'the acorn' successfully

---

[4] Here, 'w', 'm', and 's' are individual constants.

refers to an acorn). At least in the previous example, we could formulate a string of letters (albeit an ungrammatical one), but in this case that is not possible. Presumably, if (iii) were correct, then 'the acorn is false' would have the logical form of a negated acorn. Again, this falsity attribution should turn out to be either false or gappy (again depending on issues that arise below), but (iii) implies that we should treat it as we treat negated acorns, if only there were such things.

Do these claims refute (iii)? No. I can imagine someone defending it in ways that deflationism about truth has been defended. Nevertheless, these are my reasons for setting it aside in the remainder of this paper.

2. Four Cases

We are left with two options. (i) implies that falsity is truth of negation, while (ii) implies that falsity is negation of truth. To investigate (i) and (ii), I propose to consider four situations that arise in formal semantics: (A) falsity attributions from a classical language to items of a classical language, (B) falsity attributions from a classical language to items of a three-valued language, (C) falsity attributions from a three-valued language to items of a classical language, and (D) falsity attributions from a three-valued language to items of a three-valued language.[5]

For the first case (A), let L be a classical first-order interpreted language, and let M be a classical first order interpreted language with a falsity predicate, 'F(x)', and a truth predicate,

[5] Throughout this discussion I avoid unnecessary technical jargon. Since negation is our primary focus, we do not need to be specific about the 3-valued scheme we use for the non-classical languages; however, when it is relevant, I use a strong Kleene scheme. There are many other ways to think about non-classical languages with truth value gaps, but this case provides a good test for the definitions of falsity.

‘T(x)’, both of which are intended to apply to sentences of L. Assume also that there is a some method for constructing names in M of the sentences of L. Let ‘T(x)’ be defined by a traditional Tarski-style truth definition. Because we are using a metalanguage, I assume that we have the resources to avoid liar paradoxes and other things of that sort (I assume the reader is familiar with the details of these assumptions).

Definition (i) implies that ‘F(x)’ is defined as ‘~T(x)’, while definition (ii) implies that it is defined as ‘T(~ x)’. Because L and M are classical, it is easy to show that if q is a sentence of L, then T⌜~ q⌝ if and only if ~ Tq.[6] Indeed, this result echoes a clause in the definition of ‘T(x)’. Therefore, there is no difference between the two definitions of falsity when it comes to case (A). For attributions of falsity from a classical language to a classical language, truth of negation is equivalent to negation of truth.

In cases (B), (C), and (D), we consider languages with truth-value gaps. In languages with truth-value gaps (or three-valued contexts as I sometimes call them), we have to provide more precise interpretations of definitions (i) and (ii) because we have two different negations in play. That is, in three valued contexts, there are at least two ways to define negation: choice negation (-) and exclusion negation (¬). If q is true or false then ⌜- q⌝ and ⌜¬ q⌝ have the same truth value, but if q is gappy, then ⌜- q⌝ is gappy, while ⌜¬ q⌝ is true.

Consider an example. We could treat ‘A(x)’ as a predicate to be interpreted as ‘is alive’, and we could define ‘is dead’ in terms of ‘A(x)’ and choice negation as ‘- A(x)’: something is dead if and only if it is not alive. Thus, if ‘b’ is a constant (e.g., the name of an object), then ‘Ab’ says that b is alive, and ‘- Ab’ says that b is dead. However, many things are neither alive nor dead

---

[6] ‘q’ is being used as an individual constant in this sentence, and ‘⌜’ and ‘⌝’ are corner quotes (e.g., ‘⌜~q⌝’ is the name of the negation of the sentence named by ‘q’).

(e.g., a rock). Thus we might say of a rock that it is not alive and not dead. That is, the rock is not alive and not not alive. Here the 'not's are given different interpretations. When one says that a rock is not alive, one is not saying '- Ar' (using 'r' as a constant interpreted as the name of the rock in question) because that would be to say that the rock is dead. Instead, one is saying '¬ Ar'. When one says that the rock is not dead, one is not saying '- - Ar' because that would be equivalent to 'Ar'.[7] Instead, one is saying '¬ - Ar'. Thus, the claim that the rock is neither alive nor dead would be '¬ Ar ∧ ¬ - Ar'. One need not assume that this is the best way to understand 'alive' and 'dead' in order to find the example illuminating. Either way, the kinds of claims found in this example are commonplace and easily understood, which suggests that we use these two different kinds of negation frequently.[8]

In a classical setting, an interpretation of a language assigns an extension (i.e., a subset of the domain) to each predicate of the language. A predicate's extension is the set of items of the domain of which that predicate is true. Since the interpretation is classical, the predicate's anti-extension (i.e., the set of things of which the predicate is false) is just the complement of the extension. However, in a three valued setting, a predicate's extension and anti-extension need not be jointly exhaustive. Thus, in a three valued setting, the interpretation of the language assigns two sets to each predicate, an extension and an anti-extension.

For the second case (B), assume that M is a classical language just as above and that L is a three-valued first-order interpreted language. Assume as well that L contains two negation signs, one expressing choice negation and the other expressing exclusion negation (since M is classical, it has only one negation sign). Because there is only one negation sign in M, definition (i) needs

[7] I am assuming that certain classical inference rules are valid in this three-valued context.
[8] See Beall (2002) for discussion.

no amending, but since L contains two negation signs, we need to formulate more precise versions of definition (ii):

(iia) ⌈σ⌉ is false $=_{df}$ ⌈- σ⌉ is true.

(iib) ⌈σ⌉ is false $=_{df}$ ⌈¬ σ⌉ is true.

Definition (iia) takes falsity to be truth of choice negation, while (iib) takes falsity to be truth of exclusion negation.

Consider definition (i) first. Since M is classical, we know that for any sentence q of L, either q is true or q is not true (i.e., Tq ∨ ~Tq). Since falsity is negation of truth under definition (i), we derive that every sentence of L is true or false. However, since L is three-valued, there might be sentences of that are not intuitively true or intuitively false. For example, let 'a' be a constant of L and 'F(x)' be a predicate of L. If the member of the domain assigned to 'a' is neither a member of the extension of 'F(x)' nor a member of the anti-extension of 'F(x)', then 'Fa' is not true (because the definition of truth for L never makes it true). However, under this interpretation, 'Fa' should not count as false either, because the item assigned to 'a' is not in the anti-extension of 'F'. Therefore, if falsity is negation of truth, then we have the counterintuitive result that every sentence of L is either true or false, despite the fact that L has truth-value gaps. Moreover, we have no way in M to characterize sentences like 'Fa' where the object to which Fness is predicated is neither in the extension nor the anti-extension of 'F'. Typically these sentences are said to be gappy, but in M we have to call them false.

We could introduce new terms into M in an attempt to rectify this situation. Let 'q is really false' be defined as 'q is false and ⌜- p⌝ is true'. Let 'q is really true' be defined as 'q is true'. Let 'q is really gappy' be defined as 'p is false and ⌜- p⌝ is false'. These introduced terms of M do a much better job of tracking our intuitive judgments about truth, falsity, and truth-value gaps for sentences of L. For example, 'Fa' should turn out to be really gappy; it should not be really true, and it should not be really false. 'Fa' is false and '- Fa' is false. Thus, 'Fa' is really gappy, it is not really true, and it is not really false (according to the definitions of these newly introduced terms). Therefore, if we treat falsity as negation of truth in situations where we are doing semantics in a classical language for a three-valued language (i.e., definition (i) in case (B)), we will be dissatisfied with our definition of falsity. Of course, we can use 'false' as defined by (i) to introduce more intuitive notions (e.g., 'really false', and 'really gappy'), but that just proves that we should not think of 'false' as defined by (i) as a falsity predicate.

Notice that real falsity in this situation turns out to be choice negation of truth. At least, real falsity of q is both falsity of q and truth of choice negation of q. However, it turns out that the first clause is unnecessary. We get the same results if we define 'q is really false' as '⌜- q⌝ is true'. Therefore, our dissatisfaction with definition (i) has led us to definition (iia). This is provides strong support for definition (ii) over definition (i).

We have seen that definition (iia) (i.e., falsity as truth of choice negation) gives us the results we intuitively expect in case (B). How about (iib)? Notice that if p is gappy, then ⌜¬ p⌝ is true. Thus, if we define falsity as exclusion negation of truth, then we get the same unsatisfactory results as above. That is, definition (iib) is equivalent to definition (i). Both treat all untrue sentences of L as false, instead of permitting a three-part distinction between truth, falsity, and gaphood. The moral is that (iia) is the only satisfactory definition of falsity for case (B). So far, it looks like falsity will best be understood as truth of negation.

For the third case (C), assume that L is a classical language just as above and that M is a three-valued first-order interpreted language. Assume as well that M contains two terms for negation, one expressing choice negation and the other expressing exclusion negation (since L is classical, it has only one negation sign). I use 'Cnot' for choice negation and 'Xnot' for exclusion negation in M. Because there is only one negation sign in L, definition (ii) needs no amending, but since M contains two kinds of negation, we need to formulate more precise versions of definition (i):

(ia) $\ulcorner\sigma\urcorner$ is false $=_{df}$ $\ulcorner\sigma\urcorner$ is Cnot true.

(ib) $\ulcorner\sigma\urcorner$ is false $=_{df}$ $\ulcorner\sigma\urcorner$ is Xnot true.

Definition (ia) takes falsity to be choice negation of truth, while (ib) takes falsity to be exclusion negation of truth.

In case (C) we are considering falsity attributions to sentences of a classical language. Thus, our definition of falsity should respect the fact that a sentence of L is either true or false—L has no truth value gaps. Moreover, our metalanguage is three-valued. Thus, a falsity attribution might itself turn out to be gappy.

Consider definition (ia). If we adopt it, then we will be unable to prove that every sentence of L is either true or false. If q is a sentence of L, the sentence 'q is true or q is false' will be 'Tq ∨ - Tq'. It might seem like this sentence says of q that it is either true or false, but that is a mistake. For it is easy to see that the following sentences are compatible: 'Tq ∨ - Tq', '¬Tq', and '¬ - Tq'. Thus, if we use definition (ia), saying that q is either true or false is compatible with the claim that q is not true and the claim that q is not false. When dealing with a three-valued metalanguage, one has to be very careful.

If we want to be able to characterize sentences of L as either true or false, then we need to (ib) instead of (ia). According to (ib), the claim that q is either true or false is 'Tq ∨ ¬ Tq'. Again, it is easy to show that one of these disjuncts must be true. Thus, (ib) is clearly superior to (ia).

How about (ii)? Remember, L is classical, so it only has one negation. Thus, we cannot distinguish (iia) from (iib) in this situation. According to (ii), 'q is false' is synonymous with '⌜~ q⌝ is true'. We can show that this definition is equivalent to definition (ib). That is, '⌜~ q⌝ is true' is equivalent to '¬ q is true'. '⌜~ q⌝ is true' is true if and only if ⌜~ q⌝ is in the extension of 'true'. ⌜~ q⌝ is in the extension of 'true' if and only if q is not in the extension of 'true'. q is not in the extension of 'true' if and only if '¬ q is true' is true. The moral is that (ii) and (ib) are equivalent in case C, and they are satisfactory definitions of falsity.

For the final case (D), assume that L and M are three-valued interpreted languages as above. Assume as well that both L and M contain two negation signs, one expressing choice negation and the other expressing exclusion negation. Because we now have two negation signs in each language, we need to consider all four of our more precise definitions:

(ia) ⌈σ⌉ is false $=_{df}$ ⌈σ⌉ is Cnot true.

(ib) ⌈σ⌉ is false $=_{df}$ ⌈σ⌉ is Xnot true.

(iia) ⌈σ⌉ is false $=_{df}$ ⌈- σ⌉ is true.

(iib) ⌈σ⌉ is false $=_{df}$ ⌈¬ σ⌉ is true.

We know from the discussions above that if we choose (ib), then we will be able to prove that any sentence of L is either true or false; since L has truth-value gaps, we should avoid (ib). The same result holds for (iib). Thus, we should not accept it either.

We can show that in case (D), (ia) and (iia) are equivalent. 'p is Cnot true' is true if and only if ⌜- p⌝ is true. ⌜- p⌝ is true if and only if '⌜- p⌝ is true' is true. Therefore, they are equivalent. And, moreover, both seem to be satisfactory definitions of falsity.

To sum up. In (A), (i) and (ii) are equivalent. In (B), (ii) splits into (iia) and (iib). (iib) is equivalent to (i); both are unacceptable. (iia) is acceptable. In (C), (i) splits into (ia) and (ib). (ii) and (ib) are equivalent and both are acceptable; (ia) is unacceptable. In (D), (ia) and (iia) are equivalent and both are acceptable; (ib) and (iib) are equivalent and both are unacceptable. The following table illustrates these results ('CL' stands for 'Classical Language', 'TVL' stands for 'Three-Valued Langauge', 'Y' indicates acceptability, 'N' indicates unacceptability, '—' indicates imprecision, and 'n/a' indicates non-applicability).

		Cases			
		(A) (CL to CL)	(B) (CL to TVL)	(C) (TVL to CL)	(D) (TVL to TVL)
Definitions of falsity	(i) $\ulcorner \sigma \urcorner$ is not true	Y	N	—	—
	(ia) $\ulcorner \sigma \urcorner$ is Cnot true	n/a	n/a	N	Y
	(ib) $\ulcorner \sigma \urcorner$ is Xnot true	n/a	n/a	Y	N
	(ii) $\ulcorner \sim \sigma \urcorner$ is true	Y	—	Y	—
	(iia) $\ulcorner - \sigma \urcorner$ is true	n/a	Y	n/a	Y
	(iib) $\ulcorner \neg \sigma \urcorner$ is true	n/a	N	n/a	N

Based on these considerations, (ii) is superior to (i). In cases where (ii) is not specific enough (e.g., cases (B) and (D)), one should choose (iia) over (iib).

3. Criticism of (ii)

We seem to have a clear winner based on how (i) and (ii) fare in the four cases. However, there are reasons to worry about (ii).

First, as has already been noted, if (ii) is correct, then sentential form of falsity attributions is rather misleading. Although it seems like an agent who asserts that a sentence p is false is attributing some property, falsity, to p, if (ii) is correct, then the agent is really attributing a different property, truth, to some other object, p's negation. That is, the logical target and the sentential target of falsity attributions pull apart. This result is less distressing than the logical form/sentential form problem discussed in connection with (iii), but it is still a concern. It is rather counterintuitive to say that when an agent says that p is false, she is not even talking about p. Moreover, it is hard to accept that when an agent says that p is false she is calling something true.

If falsity is truth of negation, then in order for something to be false, it needs to have a negation. When I say 'the acorn is false', I am saying 'the negation of the acorn is true'. But acorns do not have negations. So, 'the negation of the acorn' is a non-denoting term. Thus, the status of 'the negation of the acorn is true' will depend on how we think about sentences with non-denoting terms. Call them *semantically defective*. However, this sentence does not seem to be semantically defective in this sense—it does not seem to have a non-denoting term. Moreover, 'the acorn is true' is not semantically defective in this sense, but it seems like either both sentences should be semantically defective or neither should be. So it looks like this definition of falsity will have problems with falsity attributions to things that are not sentences.

Furthermore, if falsity is truth of negation, then any sentence that does not have a negation is not false. If L is a classical language with no negation operator, then none of the sentences of L are false (according to (ii)). For example, consider L^-, which is the negation-free fragment of a first-order classical language, L. Let an interpretation I of L^- assign a proper subset |F| of the domain to a predicate, 'F' of L^-, and let I assign a member of the domain not belonging to |F| to a singular term, 'a' of L^-. We want to be able to say that, under I, the sentence, 'Fa' of L^- is false because, under I, the object assigned to 'a' is not in the extension of 'F'. However, if we adopt definition (ii) of falsity, we cannot say this, because to say that 'Fa' is false would be to say that the negation of 'Fa' is true; of course, since L^- is negation-free, there is no negation of 'Fa'. One way of expressing this problem is to say that definition (ii) severs the connection between 'false' and 'false of'. We want to say that a predicate 'F' is false of an object, a, if and only if 'Fa' is false. However, in this case, 'F' is false of the object assigned to 'a', but 'Fa' is not false (according to definition (ii)).

I can imagine a reader who objects: come on, who cares about languages without negation operators? My reply: actually, there are many of them and they are commonplace in branches of logic and computer science.[9] But this point should not really matter. For, if there are false sentences in such a language and a definition of falsity fails to capture this fact, then the definition is unacceptable, no matter whether these sentences are familiar or not.

A proponent of definition (ii) might be able to overcome this difficulty by taking propositions to be the primary bearers of truth and falsity: a proposition P is false if and only if P's negation is true. However, this solution is not without costs. The most obvious one is that is incompatible with theories of truth that do not take propositions to be the primary bearers of truth and falsity (e.g., disquotationalists).

[9] See Kurtonina and de Rijke (1997) for example.

A more subtle problem is that there are many kinds of negation. For example, let S be the proposition expressed by 'snow is white'. What is the negation of S? It could be: (i) the proposition expressed by 'snow is not white' where 'not' expresses classical negation, (ii) the proposition expressed by 'snow is Cnot white' where 'Cnot' expresses choice negation, (iii) the proposition expressed by 'snow is Xnot white' where 'Xnot' expresses exclusion negation, (iv) the proposition expressed by 'snow is Inot white' where 'Inot' expresses intuitionistic negation, or some other proposition Which of these propositions is being called true when S is called false? It is not obvious how a proponent of definition (ii) should choose or even that a single choice would be the best one in every situation. Again, this is not a knock-down objection, but it does seem like a genuine worry. Addressing the worry is likely to make the definition significantly more complex and bring with it further problems.

Finally, the truth of negation option is not going to work for anyone who wants to endorse an error theory of some discourse. If some sentence p of said discourse is false, then p's negation is true. But presumably p's negation is part of the same discourse. Thus, if falsity is truth of negation, then there is no discourse all of whose sentences are false. Error theories have been very influential in metaethics and other areas of philosophy[10]; even if one rejects error theories, one's definition of falsity should not rule them out as conceptually impossible.

An error theorist might address this problem by saying that only *atomic* sentences of the discourse are false. However, this suggestion faces the following problem. Say someone wants to be an error theorist about discourse involving the terms 'good' and 'bad'. Assume as well that 'bad' is defined as being synonymous with 'not good'. Now the same problem recurs. The error theorist says that 'X is bad' and 'X is good' are both false, but the claim that 'X is good' is false

---

[10] See Mackie (1977) for an example.

just means that ‘X is not good’ is true. But ‘X is not good’ is synonymous with ‘X is bad’, which is false (according to the error theorist).

4. An Alternative Definition

It is one thing to deny that we can define truth, but it is another to say that we cannot define falsity, even if we can define it in terms of truth. We have seen that definition (i)—falsity is negation of truth—is unacceptable and definition (ii)—falsity is truth of negation—has some serious problems. Definition (i) fails in three of the four cases discussed in section two. Definition (ii) faces problems generated by the fact that it has counterintuitive consequences for the targets of falsity attributions.

I suggest that we can define falsity in terms of truth, but we cannot do it with negation. Instead, we need the help of another semantic notion that links predicates or properties to objects or sets of objects. To illustrate, I use the notion of extension. The following is the suggested definition:

(iv) $\ulcorner\sigma\urcorner$ is false $=_{\text{df}}$ $\ulcorner\sigma\urcorner$ is in the anti-extension of truth.

If we think of truth as a property or concept, then its extension is the set of objects that are true. Just as each property or concept has an extension, it also has an anti-extension. Note that the extension and anti-extension need not be jointly exhaustive, although I assume that they are disjoint.

It is easy to check that definition (iv) gives the correct results in each of the four cases considered in section two. Furthermore, if we accept definition (iv), then we avoid the problems discussed in section three that face definition (ii). In particular, definition (iv) treats the sentential target of a falsity attribution as the logical target of that attribution. That is, in the falsity attribution, ‘p is false’, p is the sentential target (the thing being called false based on the form of the sentence), and if (iv) is correct, then p is also the logical target (the thing the falsity attribution is about based on the logical form of the falsity attribution). Thus, definition (iv) brings with it the benefits of definition (ii), but avoids its problems.

I have not shown that (iv) does not encounter other problems—no doubt it does. However, based on the considerations presented in this paper, (iv) might be a better candidate for a definition of falsity than (i) or (ii).[11]

[11] Thanks to Stewart Shaprio, Matti Eklund, and Cory Wright for helpful comments on earlier drafts.

Work Cited

Beall, JC. (2002). “Deflationism and Gaps: Untying ‘not’s in the Debate,” *Analysis* 62: 299–305.

Field, Hartry. (2008). *Saving Truth from Paradox*. Oxford.

Horwich, Paul. (1998). *Truth*, 2nd ed. Oxford.

Kurtonina, Natasha and de Rijke, Maarten. (1997). “Simulating without Negation,” *Journal of Logic and Computation* 7: 501-522.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*, Penguin.