Truth via Satisfaction?

NICHOLAS J.J. SMITH[1]

Abstract: One of Tarski's stated aims was to give an explication of the classical conception of truth—truth as 'saying it how it is'. Many subsequent commentators have felt that he achieved this aim. Tarski's core idea of defining truth via satisfaction has now found its way into standard logic textbooks. This paper looks at such textbook definitions of truth in a model for standard first-order languages and argues that they fail from the point of view of explication of the classical notion of truth. The paper furthermore argues that a subtly different definition—also to be found in classic textbooks but much less prevalent than the kind of definition that proceeds via satisfaction—succeeds from this point of view.

Keywords: Truth, Satisfaction, Tarski, Model

1 Introduction

In presenting his now famous definition of truth, one of Tarski's aims was to give an explication of the ordinary notion of truth:

> What will be offered can be treated in principle as a suggestion for a definite way of using the term "true", but the offering will be accompanied by the belief that it is in agreement with the prevailing usage of this term in everyday language. Our understanding of the notion of truth seems to agree essentially with various explanations of this notion that have been given in philosophical literature. What may be the earliest explanation can be found in Aristotle's *Metaphysics*: "To say of what is that it is not, or of what is not that it is, is false, while to

[1]Thanks to John P. Burgess, Max Cresswell, Kit Fine, Rob Goldblatt, Tristan Haze, Ed Mares and Simon Varey for discussions and correspondence. Earlier versions of this paper were presented at the Australasian Association for Logic Annual Conference at the University of Sydney on 3 July 2015, at a Pukeko meeting at Victoria University of Wellington on 10 October 2015 and at Logica 2016 in Hejnice on 24 June 2016; thanks to the audiences for their comments. For the germ of one of the core ideas of this paper see Smith (2012, 503–4, n.23 and n.25); cf. also Smith (2016, 756, n.17).

> say of what is that it is, or of what is not that it is not, is true." ...We shall attempt to obtain here a more precise explanation of the classical conception of truth, one that could supersede the Aristotelian formulation while preserving its basic intentions. (Tarski, 1969, 63–4)[2]

Many have felt that Tarski achieved this aim:

> The two most famous and—in the view of many—most important examples of conceptual analysis in twentieth-century logic were Alfred Tarski's definition of *truth* and Alan Turing's definition of *computability*. In both cases a prior, extensively used, informal or intuitive concept was replaced by one defined in precise mathematical terms. ...in the view of many...Tarski's definition of truth is one of the most important cases of conceptual analysis in twentieth-century logic. (Feferman, 2008, 72, 90)

In this paper I shall argue that Tarski's approach to defining truth does not succeed *from the point of view of explication or conceptual analysis of the ordinary or classical notion of truth.* I thereby go further than Field in his well-known criticism of Tarski's definition. According to Field, the received view is that Tarski reduced truth to non-semantic notions. Field argues, on the contrary, that Tarski reduced truth to other semantic notions:

> My contrary claim will be that Tarski succeeded in reducing the notion of truth *to certain other semantic notions*; but that he did not in any way explicate these other notions, so that his results ought to make the word 'true' acceptable only to someone who already regarded these other semantic notions as acceptable. ...Tarski merely reduced truth to other semantic notions. (Field, 1972, 347, 348)

I shall argue that Tarski does not reduce *truth* to anything: he does not explicate or reduce truth (in the ordinary or classical sense) at all.

Tarski's core idea of defining truth via the notion of satisfaction has now found its way into the logic textbooks. I shall focus on standard textbooks—rather than Tarski's original papers—because they give us a cleaner version

[2]Cf. Tarski (1956, 153).

of Tarski's core idea and arguably have greater contemporary relevance.[3] My central concern is not the historical study of Tarski's contributions but the analysis of truth. My central claims are first that a successful analysis of the ordinary notion of truth cannot be achieved by proceeding via satisfaction in the Tarskian way and second that a subtly different definition of truth—also to be found in classic textbooks but much less prevalent than the kind of definition that proceeds via satisfaction—*does* yield an explication of the classical notion of truth.

2 Two ways of defining truth in a model

We consider a standard first order language with names (individual constants), and predicates of each arity.[4] A model of the language comprises a domain (a nonempty set), and an assignment of a referent (an object in the domain) to each name and an extension (a set of n-tuples of members of the domain) to each n-place predicate.

Let us now set out two ways of defining truth in a model: one that proceeds via satisfaction in the Tarskian way; and one that defines truth directly rather than via first defining satisfaction.[5]

First some preliminary definitions:[6]

- A value assignment $\mathfrak{v}$ on a model $\mathfrak{M}$ is a function from the set V of all variables in the language into the domain of $\mathfrak{M}$.
- $\mathfrak{M}^{\mathfrak{v}}$ is a model $\mathfrak{M}$ together with a value assignment $\mathfrak{v}$ on $\mathfrak{M}$
- A *term* is a name or variable
- Where $\underline{t}$ is a term, $[\underline{t}]_{\mathfrak{M}^{\mathfrak{v}}}$ is:
  - the referent of $\underline{t}$ on $\mathfrak{M}$, in case $\underline{t}$ is a name
  - the value assigned to $\underline{t}$ by $\mathfrak{v}$, in case $\underline{t}$ is a variable

[3]It is irrelevant to my argument whether the authors of these textbooks do or should care about conceptual analysis. Some people certainly care about the analysis of truth—and my point is that they will not find what they seek in the approach taken in these logic texts.

[4]Nothing apart from simplicity of presentation turns on the omission of function symbols.

[5]For the first definition, see e.g. Enderton (2001). For the second definition, see Jeffrey (1967), Boolos and Jeffrey (1989) and Boolos, Burgess, and Jeffrey (2007).

[6]Following Smith (2012, §8.4.2), underlining is used for metavariables; so $\underline{x}$ is any variable, $\underline{a}$ is any name, etc.

- Where $\underline{P}$ is a predicate, $[\underline{P}]_{\mathfrak{M}}$ is the extension of $\underline{P}$ on the model $\mathfrak{M}$

Now the definition of satisfaction:[7]

- $\underline{P^n}\underline{t_1}\ldots\underline{t_n}$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff $\langle[\underline{t_1}]_{\mathfrak{M}^{\mathfrak{v}}},\ldots,[\underline{t_n}]_{\mathfrak{M}^{\mathfrak{v}}}\rangle \in [\underline{P^n}]_{\mathfrak{M}}$
- $\neg\alpha$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α is unsatisfied relative to $\mathfrak{M}^{\mathfrak{v}}$
- $(\alpha \wedge \beta)$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α and β are both satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$
- $(\alpha \vee \beta)$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff one (or both) of α and β is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$
- $(\alpha \rightarrow \beta)$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α is unsatisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ or β is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ (or both)
- $(\alpha \leftrightarrow \beta)$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α and β are both satisfied, or both unsatisfied, relative to $\mathfrak{M}^{\mathfrak{v}}$
- $\forall\underline{x}\alpha$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α is satisfied relative to $\mathfrak{M}^{\mathfrak{v}'}$ for every value assignment $\mathfrak{v}'$ on $\mathfrak{M}$ which differs from $\mathfrak{v}$ at most in what it assigns to $\underline{x}$
- $\exists\underline{x}\alpha$ is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ iff α is satisfied relative to $\mathfrak{M}^{\mathfrak{v}'}$ for some value assignment $\mathfrak{v}'$ on $\mathfrak{M}$ which differs from $\mathfrak{v}$ at most in what it assigns to $\underline{x}$

And now the first definition of truth:[8]

- A wff is true (henceforth 's-true') on a model $\mathfrak{M}$ iff it is satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ for every value assignment $\mathfrak{v}$ on $\mathfrak{M}$.
- A wff is false (henceforth 's-false') on a model $\mathfrak{M}$ iff it is unsatisfied relative to $\mathfrak{M}^{\mathfrak{v}}$ for every value assignment $\mathfrak{v}$ on $\mathfrak{M}$.

Next we set out a second definition of truth—one that defines truth directly rather than via satisfaction. First some preliminary definitions:

- Where $\underline{a}$ is a name, $[\underline{a}]_{\mathfrak{M}}$ is the referent of $\underline{a}$ on the model $\mathfrak{M}$

---

[7] If a wff is *not* satisfied relative to $\mathfrak{M}^{\mathfrak{v}}$, we say that it is *unsatisfied* relative to $\mathfrak{M}^{\mathfrak{v}}$.

[8] We use the terms 's-true' and 's-false' (rather than plain 'true' and 'false') for the properties here defined, in order to avoid ambiguity later when we look at a different definition of truth.

- Where $\alpha(\underline{x})$ is a wff that has no free variables other than $\underline{x}$, $\alpha(\underline{a}/\underline{x})$ is the wff that results from $\alpha(\underline{x})$ by replacing all free occurrences of $\underline{x}$ by the name $\underline{a}$

- Where $\mathfrak{M}$ is a model and $\underline{a}$ is a name that is not assigned a referent in $\mathfrak{M}$, $\mathfrak{M}^{\underline{a}}_{o}$ is the model that is just like $\mathfrak{M}$ except that in it the name $\underline{a}$ is assigned the referent o

Now the definition of truth:[9]

- $\underline{P^n}\underline{a_1}\ldots\underline{a_n}$ is true on $\mathfrak{M}$ iff $\langle[\underline{a_1}]_{\mathfrak{M}},\ldots,[\underline{a_n}]_{\mathfrak{M}}\rangle \in [\underline{P^n}]_{\mathfrak{M}}$

- $\neg\alpha$ is true on $\mathfrak{M}$ iff α is false on $\mathfrak{M}$

- $(\alpha \wedge \beta)$ is true on $\mathfrak{M}$ iff α and β are both true on $\mathfrak{M}$

- $(\alpha \vee \beta)$ is true on $\mathfrak{M}$ iff one (or both) of α and β is true on $\mathfrak{M}$

- $(\alpha \rightarrow \beta)$ is true on $\mathfrak{M}$ iff α is false on $\mathfrak{M}$ or β is true on $\mathfrak{M}$ (or both)

- $(\alpha \leftrightarrow \beta)$ is true on $\mathfrak{M}$ iff α and β are both true on $\mathfrak{M}$ or both false on $\mathfrak{M}$

- $\forall\underline{x}\alpha(\underline{x})$ is true on $\mathfrak{M}$ iff for every object o in the domain of $\mathfrak{M}$, $\alpha(\underline{a}/\underline{x})$ is true on $\mathfrak{M}^{\underline{a}}_{o}$, where $\underline{a}$ is some name that is not assigned a referent in $\mathfrak{M}$

- $\exists\underline{x}\alpha(\underline{x})$ is true on $\mathfrak{M}$ iff there is at least one object o in the domain of $\mathfrak{M}$ such that $\alpha(\underline{a}/\underline{x})$ is true on $\mathfrak{M}^{\underline{a}}_{o}$, where $\underline{a}$ is some name that is not assigned a referent in $\mathfrak{M}$

Note some key points about these two definitions of truth:

- Satisfaction is defined relative to a model and a variable assignment. S-truth and truth are defined relative to a model.[10]

- Satisfaction and s-truth are defined for all wffs. Truth is defined only for closed wffs.[11]

- Truth and s-truth have the same extension amongst closed wffs.

---

[9] If a closed wff is *not* true on $\mathfrak{M}$, we say that it is *false* on $\mathfrak{M}$.

[10] S-truth is satisfaction by the model and *all* variable assignments thereon. It is by quantifying over the variable assignments in this way that we get rid of them—kick away the ladder—and get a notion of s-truth defined relative to a model only.

[11] (i) A closed wff is one that contains no free occurrence of any variable; an open wff is one

3 The classical conception of truth

We are interested in whether the definitions of truth in the previous section can serve as explications of the ordinary or classical notion of truth. It is time to say a little more about this notion.[12]

The core idea is that truth is *saying it how it is*. A claim is true if things are the way it claims them to be; it is false if things are not as it claims them to be.

This idea goes back at least as far as Plato and Aristotle:

> SOCRATES: But how about truth, then? You would acknowledge that there is in words a true and a false?
> HERMOGENES: Certainly.
> SOCRATES: And there are true and false propositions?
> HERMOGENES: To be sure.
> SOCRATES: And a true proposition says that which is, and a false proposition says that which is not?
> HERMOGENES: Yes, what other answer is possible? (Plato, c.360 BC)

> ...we define what the true and the false are. To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true (Aristotle, c.350 BC, Book IV (Γ) §7)

This is just a rough guiding idea about truth. The rough idea has been used to motivate more precise, detailed theories of truth—some of which (such as certain versions of the correspondence theory of truth) are quite contentious. My interest in this paper is not in such detailed theories: it is in the basic guiding idea of truth as saying it how it is.

---

that is not closed. (ii) S-truth is well-defined for all wffs but some open wffs might be neither s-true nor s-false, relative to a given model. (iii) Some proponents of the definition of truth via satisfaction restrict the *term* 'truth' to closed wffs—but this does not change the fact that the property of satisfaction relative to all variable assignments is one that open and closed wffs can possess. We return to this point in §5. (iv) In the definition of truth (i.e. the second definition—not the definition of s-truth), the clause for atomic wffs covers only wffs containing names (not variables) and the clauses for quantified wffs cover only wffs where no variable other than the one in the quantifier occurs free in the remainder of the formula.

[12] I sometimes call the notion of truth under discussion the 'ordinary' notion of truth and sometimes call it the 'classical' notion—but note that our interest in this paper is in the analysis of *this notion* (whatever one wants to call it): it is not in the question whether this notion is the 'folk notion' of truth.

Note the following comment from Tarski:

> we are not interested here in 'formal' languages and sciences in one special sense of the word 'formal', namely sciences to the signs and expressions of which no material sense is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful. (Tarski, 1956, 166)

A formula in a purely formal, uninterpreted language does not make a claim—it does not set up a condition that the world may meet or fail to meet. Hence the question of truth—in the sense in which we are interested—does not arise for such formulas. The question of truth arises only for *contentful* sentences: only when a sentence says something does the question arise whether things are the way it says they are.[13]

4 Explicating truth I

I submit that the second definition of truth in §2 explicates the ordinary notion of truth. In particular, it spells out more precisely what 'saying it how it is' amounts to—and it incorporates the key insights that (i) what it is for a sentence to *say it how it is* varies depending on the *form* of the sentence and (ii) ultimately reduces to the case of *atomic sentences*. The definition contains one case for atomic sentences and then one case per logical operator—hence point (i)—and it is a recursive definition where the truth or falsity of any sentence is ultimately grounded in the truth or falsity of certain atomic sentences—hence point (ii). Let's look at these points in a little more detail. Each clause in the definition can be seen as telling us what it is for a certain kind of sentence to *say it how it is*. The first clause tells us that for Pa to say it how it is is for the thing picked out by a to be in the set of things (that have the property) picked out by P.[14] The second clause tells us that for a negation $\neg\alpha$ to say it how it is is for α to say it how it isn't. The third clause tells us that for a conjunction to say it how it is is for both conjuncts to say it how it is; the fourth clause tells us that for a disjunction to say it how it is

[13]Cf. Hodges (1985–6, 147–8): "The issue is simply this. If a sentence contains symbols without a fixed interpretation, then the sentence is meaningless and doesn't express a determinate thought. But then we can't properly call it true or false." This is Hodges's gloss on Frege (1971, 98): "A proposition that holds only under certain circumstances is not a real proposition."

[14]For Rab to say it how it is is for the pair of things picked out by a and b (in that order) to be in the set of pairs of things (that stand in the relation) picked out by R; and so on.

is for at least one disjunct to say it how it is; the fifth and sixth clauses tell us similar things, mutatis mutandis, for conditionals and biconditionals. The seventh clause tells us that for a universally quantified sentence to say it how it is is for every *particularisation* of it—one for each thing in the domain—to say it how it is. The core idea here is that 'everything is ϕ' says it how it is iff *this* is ϕ, and *this* is ϕ, and so on—through all the things there are. The way the definition spells out this idea is as follows. Consider a name that nothing currently has—say (for the sake of example) 'Rumpelstiltskin'. Then for 'Everyone in the room was born in Tasmania' to say it how it is is for 'Rumpelstiltskin was born in Tasmania' to say it how it is—no matter who in the room we name 'Rumpelstiltskin'. Finally the eighth clause tells us that for an existentially quantified sentence to say it how it is is for at least one particularisation of it to say it how it is.

Note how the truth/falsity of every sentence is explained in terms of the truth/falsity of sentence(s) with fewer logical operators. Hence we eventually get down to atomic sentences—and for them, the truth condition evidently encapsulates the idea of 'saying it how it is'. Thus the definition as a whole provides a genuine explication of the classical notion of truth. The definition is mathematically precise and—in the ways just discussed—it provides genuine insight into the idea of 'saying it how it is'.

Interestingly, point (i) above—which I regard as a key insight of the explication—is the basis of one of Prior's *criticisms* of Tarski:

> When the presupposed definition of 'satisfaction' is examined, however, it will be found that this definition of truth has a further defect. Satisfaction can only be defined in the following roundabout way (I again give the thing roughly): 'x is included in y' is satisfied by the pair of classes a, b if and only if a is included in b; 'not-y' is satisfied by any group of classes which does not satisfy y; 'x or y' is satisfied by any group of classes which either satisfies x or satisfies y; and a function preceded by the universal quantifier is satisfied, etc. Such a piecemeal definition of satisfaction means a similarly piecemeal definition of truth, when it is all spelt out; and the more complex the language considered the more pieces there will be. I know there are plenty of quite un-Tarski-like people who will be entirely happy about this—people who contend that even in 'everyday or colloquial' language the word 'true' has different meanings when applied to sentences of different sorts, so that it can have

> a single meaning only in the sense of a disjunction of these. My own understanding of ordinary language is quite otherwise; there are no doubt dozens of different ways of deciding whether a given sentence is true, but what it means for a sentence to be true is pretty much the same throughout, and pretty much what was suggested at the beginning of this discussion. (Prior, 1957, 408–9)

I take it that Prior is referring at the end of this quotation to his earlier comment "We generally say that a sentence is true if it says that something is so, and it is so" [406]. In response to this criticism, however, I'd say that 'true' *does* have the same meaning throughout the second definition of truth in §2: it means (just as Prior says) *saying it how it is*. However, what exactly 'saying it how it is' amounts to differs depending on the form of the sentence—and one way in which the formal definition is an advance over and explicates the guiding idea is precisely that it reveals this.

5 Explicating truth II

I shall now argue that the first definition of truth in §2 does not explicate the classical notion of truth. The key point to observe is that open wffs can be s-true (relative only to a model—not relative to a model *and* a variable assignment). For example, in a model in which the extension of P is the entire domain, the wff Px is s-true; and in any model whatsoever, the wff $\forall xPx \rightarrow Py$ is s-true. However open wffs do not *say anything* (relative only to a model)—they do not *make any claim*—and so they cannot (on the classical conception of truth) be true or false. (I do not mean that the open wff all by itself does not say anything. I mean that even given a model it does not say anything. For example, even given a domain and an extension for P, Px and $\forall xPx \rightarrow Py$ do not say anything, do not make any claim.) The *reason* they do not say anything is that some symbols in them—the free occurrences of variables—'have no material sense attached to them'. So the problem here turns on the very point that Tarski himself noted in the passage quoted in §3: when no material sense is attached to the expressions in a formula—in this case, the free occurrences of variables (remember that we are talking about open wffs relative to models, not relative to models and variable assignments)—the question of truth "has no relevance, it is not even meaningful".[15]

In short, there are formulas that have the property s-truth but that cannot be regarded as true (in the classical sense). So s-truth does not provide an explication of the classical notion of truth.

Some textbook authors withhold the term 'true' from open wffs: they say only that a closed wff is true (on a model) iff it is satisfied relative to all variable assignments (on that model). But this move does not help with the present problem. Even if we restrict the *term* 'true' to closed wffs, the *property* that we call 'truth' when closed wffs have it (viz., satisfaction relative to all variable assignments) can't yield an explication of truth (in the ordinary sense) because that property is one that open wffs can also have—and they cannot be true in the ordinary sense (because they make no claim). Declining to *call* an open wff 'true' when it has this property does not take away from the fact that one has defined a property that open wffs can possess, and then called a closed wff 'true' when it has *that property*. My point is that because open wffs can also have this property, possessing this property cannot make something (not even if it is a closed wff) *true* in the ordinary sense.

6 Value Assignments vs Sequences of Objects

Where a wff is true or false relative to one thing—a model—a wff is satisfied or unsatisfied relative to *two* things: a model $\mathfrak{M}$ *and* something else ('X'). One option for X is, as we have seen in §2, a value assignment on $\mathfrak{M}$. A second option for X is a denumerable (i.e. countably infinite) sequence of objects from the domain of $\mathfrak{M}$.[16]

Given a countable infinity of variables and an enumeration of them, the two approaches are interchangeable. To get a sequence of objects, given a value assignment, we transfer the enumeration of the variables to objects in the domain, via the value assignment function. To get a value assignment, given a sequence of objects, we assign the nth object in the sequence to the nth variable.

[15] Cf. Hodges (2004, 99). My reminders throughout this paragraph that we are talking about open wffs relative to models—not models and variable assignments—should not be taken to suggest that relative to a model and a variable assignment, any open wff *does* say something. I discuss this issue in §7.

[16] See e.g. Mendelson (1987) and Hunter (1971). Note that the domain need not be infinite: the same object may appear at more than one place in the sequence.

The sequence approach goes back to Tarski himself. Tarski writes at the outset "In this construction I shall not make use of any semantical concept if I am not able previously to reduce it to other concepts" (Tarski, 1956, 152–3). Milne (1999, 151–2) regards an assignment of values to variables as a semantic notion and says that the approach via sequences is a "stroke of genius" that enables this semantic notion to be avoided. I'm not sure (a) that the idea of an assignment of values to variables is a semantic notion—nor (b) that, if it were, the approach via sequences would really avoid semantic notions. On (a): cf. the discussion in §7 below of the point that a model plus a value assignment does *not* make any wff say something or make a claim. On (b): Imagine an old-fashioned school dance, where the headmaster wants to avoid doing anything as vulgar as pairing up boys and girls—so he assigns each boy a number, and each girl a number, and then the students match numbers. If there really were something wrong with pairing up boys and girls, then could this coy approach via numbering seriously be regarded as having avoided wrongdoing?

Of course, to settle issues (a) and (b), we'd need a full discussion of what makes a notion 'semantic'. However we do not need to settle these issues here. The key point for purposes of this paper is that moving from value assignments to sequences does not help with the problem raised in §5. As far as the analysis of truth is concerned, the sequence version just adds a further step of indirection to the variable-assignment version. This certainly doesn't make the definition any better from the point of view of explication of the classical notion of truth; if anything it makes it worse.

7 Notational variants?

I imagine someone might object as follows:

> This is a storm in a teacup. The two approaches to defining truth presented in §2 are simply notational variants of one another. An open wff relative to a model *and* a value assignment can be seen this way: the free occurrences of variables have effectively been made into names and then assigned referents (by the value assignment).[17] So this is exactly like the second approach to defining truth—where we have new names (that

[17]Cf. Field (1972, 349): "The idea is going to be to treat the variables, or at least the free variables, as sort of "temporary names" for the objects assigned to them."

are not assigned a referent on the original model) and extended models that assign them referents.

There are two points to make in response to this objection.

First, even if it were true that satisfaction relative to a model plus value assignment were simply a notational variant on truth as defined in the second way in §2, this would not affect my argument: for even if there were some *other* property involved in the first definition of truth (e.g. the property of satisfaction relative to a model and a variable assignment) that *did* explicate the ordinary notion of truth, that would not affect my point that the property of *s-truth* (satisfaction relative to *all* variable assignments) does *not* explicate the ordinary notion of truth.

Second, we cannot, in general, regard assigning a value to a variable as the same thing as viewing the variable as a name and assigning it a referent. This gloss works fine for free occurrences of variables—but not for bound occurrences. Yet note that a value assignment assigns values to *variables*—not to *occurrences* of variables. It is fine to say that Px says that Bill is P, relative to a value assignment that assigns Bill to x. But what about $\forall xRxy$, relative to a value assignment that assigns Bill to x and Ben to y? It does *not* say 'Everything bears R to Ben'. That would be to ignore the assignment of Bill to x. It actually says something like 'Every Bill is such that Bill bears R to Ben'—which makes no sense at all.[18] The hope (of the imagined objector) was that a model *plus* a value assignment makes *any* wff say something or make a claim—and that satisfaction (relative to a model and a variable assignment) is just like truth as defined in the second way in §2 and hence does explicate truth in the classical sense. However this thought does not pan out. There is a dilemma here (for the imagined objector). Without a variable assignment (i.e. relative only to a model), open wffs have no content—they do not make claims. But *with* a variable assignment (regarded as turning variables into names and assigning them referents), wffs with bound occurrences of variables become meaningless.

[18]Cf. Shoenfield (1967, 13): "Most of our previous remarks about variables are false when applied to the x in $\forall x(x = 0)$. This formula has only one meaning, while $x = 0$ has many meanings. We get a particular meaning of $x = 0$ by substituting 2 for x; but if we substitute 2 for x in $\forall x(x = 0)$, we get the meaningless expression $\forall 2(2 = 0)$."

8 Conclusion

I have considered two definitions of truth: one that proceeds via satisfaction and one that does not. (I have furthermore considered two versions of the first definition: one that appeals to variable assignments and one that appeals to sequences.) I have argued that the first definition fails from the point of view of explicating the classical notion of truth and that the second definition succeeds from this point of view.

Note that these conclusions are quite specific. For example, I have not argued that the second definition is the only possible definition of truth that can serve as an explication of the classical notion. So what about other definitions of truth in the mainstream textbooks—and in the broader literature? A discussion of all these approaches would take far more space than I have available here. It would also raise a host of new issues that have not featured in the discussion in this paper. For example, consider the approach taken by Robbin (1969),[19] who assumes that the set of names in the language is the very same set as the domain of the model—or the approach taken by Robinson (1951, 1963, 1966), who assumes a set of names of arbitrary transfinite cardinal number and a one-one correspondence between the domain of the model and the set of names. Neither of their definitions of truth faces the problem I have raised for the first definition presented in §2 above. However one might still think that they are less than ideal as explications of the ordinary notion of truth, because the ordinary notion applies to languages in which the names used to talk about objects and the objects talked about are distinct, and to languages that do not contain names for every object. However these are arguments that would have to be made: they bring in new considerations and do not follow automatically from the arguments of this paper.

References

Aristotle. (c.350 BC). Metaphysics. (In Barnes (1984), vol.2.)

Barnes, J. (Ed.). (1984). *The Complete Works of Aristotle: The Revised Oxford Translation.* Princeton NJ: Princeton University Press.

Boolos, G. S., Burgess, J. P., & Jeffrey, R. C. (2007). *Computability and Logic* (Fifth ed.). Cambridge: Cambridge University Press.

[19]Thanks to Rob Goldblatt for pointing me to this book.

Boolos, G. S., & Jeffrey, R. C. (1989). *Computability and Logic* (Third ed.). Cambridge: Cambridge University Press.

Enderton, H. B. (2001). *A Mathematical Introduction to Logic* (Second ed.). San Diego: Harcourt/Academic Press.

Feferman, S. (2008). Tarski's conceptual analysis of semantical notions. In D. Patterson (Ed.), *New Essays on Tarski and Philosophy* (pp. 72–93). Oxford: Oxford University Press.

Field, H. (1972). Tarski's theory of truth. *Journal of Philosophy*, *69*(13), 347–375.

Frege, G. (1971). *On the Foundations of Geometry and Formal Theories of Arithmetic*. New Haven: Yale University Press. (Translated by Eike-Henner W. Kluge.)

Hamilton, E., & Cairns, H. (Eds.). (1963). *The Collected Dialogues of Plato*. New York: Pantheon Books.

Hodges, W. (1985–6). Truth in a structure. *Proceedings of the Aristotelian Society*, *86*, 131–51.

Hodges, W. (2004). What languages have Tarski truth definitions? *Annals of Pure and Applied Logic*, *126*, 93–113.

Hunter, G. (1971). *Metalogic: An Introduction to the Metatheory of Standard First Order Logic*. Berkeley: University of California Press.

Jeffrey, R. C. (1967). *Formal Logic: Its Scope and Limits*. New York: McGraw-Hill Book Company.

Mendelson, E. (1987). *Introduction to Mathematical Logic* (Third ed.). Monterey CA: Wadsworth & Brooks/Cole.

Milne, P. (1999). Tarski, truth and model theory. *Proceedings of the Aristotelian Society*, *99*, 141–67.

Plato. (c.360 BC). Cratylus. (In Hamilton and Cairns (1963).)

Prior, A. (1957). Critical notice of *Logic, Semantics, Metamathematics* by Alfred Tarski. *Mind*, *66*(263), 401–10.

Robbin, J. W. (1969). *Mathematical Logic: A First Course*. New York: W.A. Benjamin.

Robinson, A. (1951). *On the Metamathematics of Algebra*. Amsterdam: North-Holland Publishing Company.

Robinson, A. (1963). *Introduction to Model Theory and to the Metamathematics of Algebra*. Amsterdam: North-Holland Publishing Company.

Robinson, A. (1966). *Non-Standard Analysis*. Amsterdam: North-Holland Publishing Company.

Shoenfield, J. R. (1967). *Mathematical Logic*. Reading MA: Addison-Wesley.

Smith, N. J. (2012). *Logic: The Laws of Truth*. Princeton: Princeton University Press.
Smith, N. J. (2016). Truthier than thou: Truth, supertruth and probability of truth. *Noûs*, *50*(4), 740–58.
Tarski, A. (1956). The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (pp. 152–278). Oxford: Clarendon Press. (Translated by J.H. Woodger.)
Tarski, A. (1969). Truth and proof. *Scientific American*, *220*(6), 63–77.

Nicholas J.J. Smith
Department of Philosophy, The University of Sydney
Australia
E-mail: `nicholas.smith@sydney.edu.au`