Rysiek Sliwinski (ed.), *Philosophical Crumbs. Essays dedicated to Ann-Mari Henchen-Dahlquist on the occasion of her seventy-fifth* birthday, Uppsala: Department of Philosophy, Uppsala University, 1999 (Uppsala Philosophical studies 49), pp. 103-121.

HORWICH'S MINIMALIST CONCEPTION OF TRUTH: SOME LOGICAL DIFFICULTIES*

Sten Lindström

Aristotle's words in the *Metaphysics*: "*to say of what is that it is, or of what is not that it is not, is true*" are often understood as indicating a *correspondence view* of truth: a statement is true if it corresponds to something in the world that makes it true. Aristotle's words can also be interpreted in a *deflationary*, i.e., metaphysically less loaded, way. According to the latter view, the concept of truth is contained in platitudes like: 'It is true that snow is white iff snow is white', 'It is true that neutrinos have mass iff neutrinos have mass', etc. Our understanding of the concept of truth is exhausted by these and similar equivalences. This is all there is to truth.

In his book *Truth* (Second edition 1998), Paul Horwich develops *minimalism*, a special variant of the deflationary view. According to Horwich's minimalism, truth is an indefinable property of propositions characterized by what he calls the *minimal theory*, i.e., all (nonparadoxical) propositions of the form: It is true that p if and only if p.

Although the idea of minimalism is simple and straightforward, the proper formulation of Horwich's theory is no simple matter. In this paper, I shall discuss some of the difficulties of a logical nature that arise. First, I discuss problems that arise when we try to give a rigorous characterization of the theory without presupposing a prior understanding of the notion of truth. Next I turn to Horwich's treatment of the Liar paradox and a paradox about the totality of all propositions that was first formulated by Russell (1903). My conclusion is that Horwich's minimal theory cannot deal with these difficulties in an adequate way, and that it has to be revised in fundamental ways in order to do so. Once such revisions have been carried out the theory may, however, have lost some of its appealing simplicity.

1. Deflationism

Various views about truth go under the label of *deflationism*. The general idea is that truth is a 'thin' or 'logical' concept, the essential content of which is explained along the following lines:

* An earlier version of this paper was presented at the Friday Seminar of the Department of Philosophy and Linguistics, Umeå University. I am grateful to the participants, especially Gunnar Andersson, Anders Berglund, Ingvar Johansson and Peter Melander, for their helpful comments. Special thanks are due to Wlodek Rabinowicz for his insightful written remarks and to Bertil Strömberg for valuable discussions.

(1) for any bearer of truth x, the claim that x is true is equivalent to the claim (expressed by) x itself.

To understand the notion of truth it is, according to *deflationism*, sufficient to recognize the truth of (1). Accordingly, truth is not a philosophically contentious concept in need of an analysis in metaphysical or epistemological terms but rather a notion whose central use is logical or semantical. Deflationists thus agree in rejecting 'thick' or *inflationary* conceptions of truth like those expressed by correspondence theories, verificationist theories, and coherence theories. In particular, deflationists agree in rejecting any kind of *truth-maker analysis*:[1]

(2) x is true iff there exists a truth-maker for x,

of which correspondence theories are the most obvious examples. Compare Wittgenstein's *picture theory* in the Tractatus stating that an elementary proposition is true just in case it depicts an existing state-of-affairs. Verificationist theories are of this kind too: a sentence is true if, and only if, there exists a verification (or a proof) of it. Even some versions of the *coherence theory* of truth are covered by (2): x is true iff there exists a *coherent* set of beliefs to which x belongs. Deflationists reject truth-maker analyses regardless of whether their aim is to analyze the concept of truth or to uncover the nature of truth.

Deflationary theories differ concerning what kind of objects they take as (the primary) *bearers of truth*. Here the principal alternatives are: (i) the *sentential view* that truth is primarily a property of meaningful (eternal) *sentences* of a natural or constructed language; (ii) the *propositional view* that takes *propositions* as the primary bearers of truth and ascribes truth to sentences only insofar as they express true propositions. Propositions are here taken to be the kind of abstract entities that serve as informational contents of sentences and are objects of propositional attitudes like belief and desire.

Formal theories of truth, like those proposed by Tarski and Kripke, are often interpreted in a deflationary spirit, but they do not prima facie preclude 'thick', 'inflationary' interpretations of the concept of truth.[2]

Often deflationism is presented by means of a schema, sometimes called the *equivalence schema*:

(E) S is true iff p,

---

[1] My terminology here is inspired by Sundholm (1994)

[2] Cf. Tarski (1983), (1943), (1969) and Kripke (1975).

where ‘p’ is to be replaced by an (eternal) sentence and ‘S’ by a standard name for this sentence or the proposition that it expresses. Instances of the schema take one of the forms:

(i) the sentence ‘The Earth is round’ is true iff the Earth is round,
(ii) the proposition that the Earth is round is true iff the Earth is round.

depending on whether sentences or propositions are taken as the primary bearers of truth.

Deflationists differ with respect to the strength of the equivalence they read into (E). Presumably, most will agree that each instance of (E), at least in its propositional version, is both necessarily and a priori true.[3] But some go further and assert *identity* between the proposition that x is true and the proposition (expressed by) x itself. That is, they accept every (nonparadoxical) instance of the *identity schema*, in one of the formulations:

(a) the proposition that p is true = the proposition that p
(b) the proposition that the sentence ‘p’ is true = the proposition that p.

---

[3] That sentences like (ii) are necessarily and a priori true seems noncontroversial, at least for philosophers that accept talk of propositions. Sentences of type (i) are more problematic. Given that both occurrences of ‘The Earth is round’ in (i) are interpreted as actually having the same meaning, (i) is a priori true. One might argue, though, that (i) is not necessarily true: If the sentence ‘The Earth is round’ had meant, for instance, that the Earth is flat, wouldn’t (i) have been false? Consider the sentence:

(iii) Necessarily, the sentence ‘The Earth is round’ is true iff the Earth is round.

Is (iii) true or not? Isn’t it obvious that (iii) is false? However, this is too fast. The sentence (iii) is ambiguous. We can interpret it either as:

(iii’) It is necessarily the case that the sentence ‘The Earth is round’, with the meaning that it *actually* has, is true iff the Earth is round.
(iii’’) It is necessarily the case that the sentence ‘The Earth is round’, with whatever meaning it turns out to have, is true iff the Earth is round.

According to interpretation (iii’), the sentence ‘The Earth is round’, with the meaning it has in the *actual world*, is evaluated for truth in any possible world w. Since, presumably, ‘The Earth is round’ means in the actual world that the Earth is round, under this interpretation, the sentence ‘The Earth is round’ is true in any world w iff the Earth is round in w. Thus, on interpretation (iii’), (iii) comes out as true. However, according to interpretation (iii’’), (i) may fail to be true in some possible worlds. For example, consider a world w, very much like ours except for the English word ‘round’ meaning in w *flat* and the English word ‘flat’ meaning in w *round*. Then, in w ‘The Earth is round’ means (in English) that the Earth is flat. When we evaluate (iii) according to this latter interpretation, it comes out as false. Of course, the words on the right hand side of (i) or (iii) are *used* not *mentioned*, and should therefore be interpreted according to our actual usage. Interpretation (iii’) seems, to me at least, to be the more natural one. So there is a natural reading of (i) according to which it is both a priori and necessarily true.

Version (b) of the identity schema is quite implausible. Consider someone who believes that the Earth is round but does not speak any English. He may not have any opinion, one way or the other, whether the sentence 'The Earth is round' is true. Thus, he believes the proposition that the Earth is round without believing the proposition that the sentence 'The Earth is round' is true. This indicates that the schema (b) must be false.

Frege in his article 'Thoughts' subscribes to the identity schema in the form (a):

> ...we cannot recognize a property of a thing without at the same time finding the thought *this thing has this property* to be true. So with every property of a thing is tied up a property of a thought, namely truth. It also worth noticing that the sentence ' I smell the scent of violets' has just the same content as the sentence 'It is true that I smell the scent of violets.' So it seems, then, that nothing is added to the thought by my ascribing to it the property of truth. And yet is it not a great result when the scientist after much hesitation and laborious research can finally say 'My conjecture is true'? The meaning of the word 'true' seems to be altogether *sui generis.* May we not be dealing here with something which cannot be called a property in the ordinary sense at all? In spite of this doubt I shall begin by expressing myself in accordance with ordinary usage, as if truth were a property, until some more appropriate way of speaking is found.[4]

In spite of what Frege here says, the schema (a) is not self-evident. Consider the proposition expressed by the sentence 'The Earth is round'. This proposition, let us call it P, concerns the Earth and ascribes a property to it. Consider, on the other hand, the proposition expressed by 'The proposition that the Earth is round is true'. The latter proposition is about the proposition P and describes it as being true. Since these two propositions have different subject matters, it is reasonable to conclude that they are distinct.

Deflationism, as I have described it, should be distinguished from various versions of *nihilism* about truth.[5] According to deflationists, there is a property of truth – albeit a "nonsubstantial" one – being applicable to truth bearers of one kind or another. Their disagreement with correspondence theorists, coherence theorists and other "inflationists" concerns the character of this property. Nihilists about truth, on the other hand, deny that there is a property of truth. The predicate *true* on the nihilist view does not express a property; and to say that a truth bearer x is true is not to ascribe a property to x at all. Nihilists maintain that 'It is true that the Earth is round' expresses the same proposition as 'The

---

[4] Frege (1984), pp. 354-355.

[5] Tarski (1969) speaks about "the nihilistic approach to the theory of truth" and attributes the term to Kotarbinski. I have adopted the term 'nihilism about truth' from Soames (1999). Nihilism about truth also goes under the name the *redundancy theory.*

Earth is round'. Nihilists may also claim that 'The sentence 'the Earth is round' is true' is a more long-winded way of saying 'The Earth is round'. Thus, nihilists accept the identity schema, at least in the propositional form (a).

In the quote above, Frege seems to be wavering between two different views concerning the concept of truth. First, there is the view that truth is an indefinable property of propositions. On this view, it is quite plausible to maintain that 'The Earth is round' and 'The proposition that the Earth is round' express two different, although necessarily equivalent, propositions. The two sentences do not have the same cognitive significance, and therefore do not express the same proposition. From the quote, one can also infer that Frege was drawn to the nihilistic conception that truth is not a property at all. The fact that he was tempted by the latter view, in conjunction with his idea that the concept of truth is implicated in every proposition, explains why he accepted the identity schema. It seems, though, that Frege could not consistently hold on to the nihilistic view. For one thing, truth as a predicate is definable within Frege's semantics:

> a proposition p is true iff p designates the object (the truth-value) *Truth*.

It is hard to see why this predicate should not express a property of propositions. Why shouldn't *designating Truth* be a perfectly respectable property of propositions?

One version of the nihilistic approach to truth, often referred to as the *performative theory* of truth, goes back to Strawson (1950). According to this theory, when we say that it is true that the Earth is round, we do several things at once:

(i) we assert the proposition that the Earth is round,
(ii) we perform a speech act of endorsing or conceding the proposition that the Earth is round.

Although, 'The Earth is round' and 'It is true that the Earth is round' express the same proposition, the two sentences have different uses. It is this difference in use that explains the utility of the truth-predicate. Like other nihilistic approaches, Strawson's theory has problems with accounting for the use of 'true' in sentences that are not of the simple type: 'It is true that p' or 'the sentence 'p' is true'. How could one, on a nihilistic approach, account for sentences like: 'Most of the Pope's assertions are true', or 'I wish that what she said is true'?

2. Horwich's minimalist conception of truth

One particular brand of deflationism is Paul Horwich's *minimalist conception of truth* (or *minimalism* as it is also called).[6] Horwich's minimalism has three ingredients:

(i) *An account of the concept truth:* Horwich claims that the word 'true' picks out an indefinable property of propositions whose content is exhausted by (or is "implicitly defined by") a certain theory which he calls *the minimal theory of truth*, or MT for short. Roughly speaking, the axioms of MT are all propositions that are expressed by (nonparadoxical) instances of the schema:

(E) The proposition that p is true iff p.

Horwich accepts an argument, due to Patrick Grim, for the claim that the collection of axioms of MT is too large to form a set. If so, there is no question of expressing MT in any single language.

(ii) *An account of the utility of the truth predicate:* If the truth predicate only occurred in what we may call *primary contexts*:

(a) *The proposition that snow is white is true* (or its sentential counterparts: *'Snow is white' is true*),

then, it could be eliminated by means of the schema (E) and would thus be *redundant* (at least in extensional contexts). However, we also want to use the truth predicate to say things like:

(b) The continuum hypothesis is true.
(c) There are true propositions that are not supported by the available evidence.
(d) Every sentence is such that either it or its negation is true.
(e) Most statements that Clinton made in his deposition were true.

In the latter sentences, however, the truth predicate cannot be eliminated by means of the (E) schema. According to Horwich, *the sole purpose of having a truth predicate at all* is to is to be able to express claims of this latter kind.

(iii) *An account of the nature of truth:* The property truth does not have any underlying nature and the explanatory basic facts about truth are instances of the (E) schema.

---

[6] Cf. Horwich (1998).

3. On the proper formulation of the minimal theory

The formulation of Horwich's minimal theory MT meets with a number of difficulties. First, we have the problem of characterizing in a precise way the axioms of MT for those propositions that are not expressible in English. As a first approximation, we may say that a proposition is an axiom of the theory MT if it is expressed by an English sentence of the form:

(E) The proposition that p is true iff p.

Horwich analyses the expression 'that' (or 'the proposition that ...') as an term-forming operator which applied to a sentence S yields a singular term standing for the proposition expressed by S Thus, expressions like 'that snow is white' or 'the proposition that snow is white' are interpreted as singular terms designating propositions. Introducing '<p>' as a shorthand for 'the proposition that p', Horwich writes the schema (E) as:

(E) <p> is true iff p.

But so far, we have only specified the axioms of MT for those propositions that are expressible by (eternal) sentences of English. We know that if a proposition P is expressed by a certain (eternal) sentence '...' of English, then the sentence:

<...> iff ...,

is an axiom of MT. But the theory MT is supposed to contain for any proposition P, a *truth-axiom* for P, intuitively specifying under what conditions P is true.

If we were content with specifying truth axioms only for those propositions that are expressible in English, we could instead of (E) write:

The proposition expressed in English by 'p' is true iff p,

or shorter:

(T) 'p' is true in English iff p,

thereby avoiding the intensional (i.e., referentially opaque) construction 'The proposition that ...'. The minimal theory, though, is not concerned only with the truth of propositions expressible in English, but with the predicate 'true' as it is applied to any proposition regardless of whether it is expressible in English or not. If the theory were applicable only to propositions that are expressible in English, or any other natural language, it would be *incomplete* as a theory of truth.

So the problem remains of specifying truth axioms for those propositions that are not expressible in English. Now one might think that this problem should

have a simple solution. Suppose, namely, that we can quantify over the collection of all propositions. Then, it appears that we can express the minimal theory by *one single axiom*:

(□) □P(P is true iff P),

and our theory of truth will have the requisite generality. Couldn't we then just infer all the instances of the schema (E) from (A) by predicate logic? We could, it seems, infer from (A):

(1) The proposition that snow is white is true iff snow is white.

This, however, is a dubious proposal. First of all, it is not clear that the axiom A makes sense, unless we interpret the right-hand side of:

(2) P is true iff P,

simply as a shorthand for 'P is true'. Remember that 'P' is supposed to be a referential variable taking propositions as values. So, on the face of it, (2) is ungrammatical. But on the suggested reading, (A) just says that:

□P(P is true iff P is true),

which is correct but, presumably, not what we wanted to express.

It is also dubious, to say the least, whether (1) really follows from (A). This kind of inference would be correct only if we interpreted ordinary sentences of English, in a Russellian way, as designating propositions. But then 'Snow is white' and 'That snow is white is true' would presumably designate the same proposition. The minimal theory would then be completely tautologous and it wouldn't tell us anything about the concept of truth. We would have some kind of redundancy theory, rather than Horwich's minimalist one.

The view that sentences of English designate propositions is not, however, Horwich's view. He speaks of sentences of English as expressing rather than designating propositions and of that-clauses as singular terms referring to propositions. On such a view, it does not make sense to replace the two occurrences of 'snow is white' in:

The proposition that snow is white is true iff snow is white,

by a referential, bindable variable x. To write:

The proposition that x is true iff x,

or,

for all x, the proposition that x is true iff x,

is not meaningful. There simply is no appropriate kind of entities available for the variable 'x' to range over.

On the other hand, the *schema* (E) makes good sense and we can apply *substitutional quantification* to it, thus obtaining:

Πp(the proposition that p is true iff p),

where Π is a universal substitutional quantifier whose substitution class is a set of English sentences. But substitutional quantification does not help us solve the problem of specifying the axioms of MT that correspond to propositions that are not expressible in English.

Horwich suggests various ways of specifying the axioms of MT. One idea is that the schema (E) supplies us with the *propositional structure*:

(E*) <<p> is true iff p>,

and that the axioms of MT are all propositions that have this structure. The suggestion is that the schema (E) has a meaning which when applied to, say, the proposition:

<snow is white>,

yields the proposition:

<<snow is white> is true iff snow is white>.

The meaning expressed by (E) is denoted by the schema (E*).

According to this analysis, the schema (E) expresses a *propositional function*, which takes any proposition P into the result of applying the schema (E) to P. Or to quote Horwich (1998), p 19-20:

> Indeed, when applied to any proposition, y, this structure (or function) yields a corresponding axiom of the minimal theory, MT.
> In other words the axioms of MT are given by the principle
> (5) For any object x: x is an axiom of the minimal theory if and only if, for some y, when the function E* is applied to, its value is x.
> Or in logical notation,
> (5*) $(x)(x \text{ is an axiom of MT} \leftrightarrow (\exists y)(x = E^*(y)))$.

Thus, the schema (E) is interpreted as expressing a propositional function E*, which can be applied to propositions in general, regardless of whether they are expressible in English or any other language. This interpretation seems at odds with (E)'s character as a linguistic schema. It does not seem right to associate the schema with a uniform sense or meaning in this way. Perhaps, Horwich is looking upon (E) as the expression of a *rule* rather than as a *schema*. But to me, at least, this interpretation of (E) seems far-fetched. I can understand projecting (E) to possible extensions of English or to other languages. But I do not under-

stand what it means to apply it directly to a proposition in a way unmediated by language.

We understand how to apply (E) to a proposition by replacing the letter 'p' in E by a *sentence* that expresses the proposition in question. If there is no sentence available to substitute into (E), we don't understand what it means to apply the schema (E) to a proposition. It is not even clear that the result of applying (E) to a proposition is independent of the choice of linguistic representation of the proposition in question.[7]

I am not implying however, that, given a sufficiently powerful conceptual apparatus, we cannot define the theory MT. What is doubtful, however, is whether we can do so in a way that does not already presupposes the concept of truth. In order to define the theory MT, we may need to use strong logical notions that themselves presuppose the notion of truth. If so, it is hard to understand what is meant by saying that MT exhausts our understanding of truth. In order to make my point clearer, I shall now describe an alternative method of characterizing the theory MT.

---

[7] In the footnote on p. 19, Horwich presents an alternative method of specifying the axioms of MT:

> ...we can characterize the 'equivalence axioms' for unformulatable propositions by considering what would result if we *could* formulate them and could instantiate those formulations in our equivalence schema. Thus we may specify the axioms of the theory of truth as what are expressed when the schema
>
> (E) '<p> is true iff p'
>
> is instantiated by sentences in any possible extension of English.

Here, Horwich assumes that every proposition is at least expressible in some possible extension of English. The question remains whether the collection of all possible extensions of English is a well-defined totality. But perhaps, the idea of looking at (E) as specifying a rule gives us the best clue to the interpretation of Horwich's theory. The rule looks somewhat like this:

If s is a sentence of some possible extension of English, then the result of replacing 'p' in the schema (E) by the sentence s is the expression of an axiom of MT.

Taking the theory in this way gives it an open-ended character. Looking at the axioms of MT that are expressible in a given extension of English never exhausts the content of the theory. There is always the possibility of extending the language with new sentences and thereby also with new instances of the schema (E). Perhaps, we should say that MT, interpreted in this way, is not a single theory but rather a recipe for constructing theories of truth. For any single language L, the part of the theory that concerns L is expressed by those axioms that are expressible in L. But it is important for our understanding of the truth predicate and for the notion of truth to realize that the predicate can always be applied to new sentences that are not expressible in L. I am, however, not sure whether Horwich would accept this interpretation. To me this version of the minimal theory seems to be the most sensible one.

Let **P** be the collection of absolutely all propositions. We suppose that there is a subcollection **T** of **P** containing exactly those propositions that are true. We also assume that there are Boolean operations +, □, □, □ on propositions corresponding to disjunction, conjunction, negation and material implication. For all propositions P, Q,

(1) P + Q □ **T** iff P □ **T** or Q □ **T**
(2) P □ Q □ **T** iff P □ **T** and Q □ **T**
(3) □P □ **T** iff P □ **T**
(4) P □ Q □ **T** iff either P □ **T** or Q □ **T**

In addition, we postulate that, for every P □ **P**, there is the proposition **Tr**(P) □ **P**, intutively, saying that P is true. Thus, **Tr** is an operation from propositions to propositions. For all propositions P,

(5) **Tr**(P) □ **T** iff P □ **T**.

In terms of the Boolean operations already given, we can define the Boolean operator □ (material equivalence) in the usual manner. Thus, for all propositions P, Q,

(6) P □ Q □ **T** iff (P □ **T** iff Q □ **T**).

From (5) and (6) we get,

(7) **Tr**(P) □ P □ **T**,

for all propositions P. For any proposition P, we can define the *T-axiom* for P as the proposition **Tr**(P) □ P. We can then define the set of axioms of MT as follows,

> A proposition P is an *axiom of MT* iff there is a proposition Q such that P = (**Tr**(Q) □ Q), i.e., P is the T-axiom for Q.

We may use the notation Axiom_{MT} for the collection of all axioms of MT.[8]

Consider now any sentence 'p' of English. If 'p' expresses the proposition P, then, apparently, 'It is true that p iff p` expresses the proposition (**Tr**(P) □ P). Thus, all English instances of the schema

(E): It is true that p iff p,

express axioms of MT. The same would be true also for instances of (E) in possible extensions of English.

---

[8] Notice that the specification of the axioms of MT is dependent on the notion of propositional identity that seems to presuppose a *principle of individuation* for propositions. Until such a principle is provided our specification of the axioms is incomplete.

If we were given a determinate theory of propositions of the indicated type, we could define the axioms of MT. But the question would still remain of determining what the *theorems* of MT are. This is a question that Horwich pays very little attention to. In order to answer it, we would need a notion of *logical consequence* for propositions in general. That is, we would need a consequence operation Cn that was defined for every collection X of propositions. Given such an operation, we could define the theorems of MT as the propositions that belong to $Cn(Axiom_{MT})$.

The difficulties involved in developing a general theory of propositions together with an appropriate notion of logical consequence applicable to absolutely all propositions seem formidable. As we shall see later, the notion of a collection of absolutely all propositions is hardly a coherent one. And even if we could make sense of it, we have hardly any idea of how to define a notion of logical consequence applicable to all propositions. The option of identifying propositions with classes of possible worlds and logical consequence with class-inclusion hardly seems open to Horwich. This construction presupposes the notion of a proposition being *true* at a world. And it is part of Horwich's project to be able to speak of propositions and logical consequence without presupposing the notion of truth.[9] But then he cannot avail himself of the notion of *truth in a world* either. If he had the latter notion, he could presumably define truth simpliciter as truth in the actual world.

The idea of defining the axioms of MT from the schema (E) is appealing, but it does not seem to work. On the other hand, if we have a sufficiently strong theory of propositions and logical consequence for propositions, then we can surely define the theory MT in the way indicated above. However, we seem to have no idea of how we could get a firm grasp of the requisite notions of proposition and logical consequence without a prior understanding of truth.[10]

---

[9] Cf. Horwich (1998), p. 4: "...the concept of truth plays no substantial role in the justification of logic".

[10] An additional objection to Horwich's minimalism is that truth does not seem to be the only concept satisfying the axioms of MT. Suppose, for instance, that for every proposition P, there is a proposition **A**(P) meaning that P is *actually true*. Thus, **A**(P) is true in a possible world w iff P is true (in the actual world). Then, for every P, **A**(P) ↔ P is a true proposition, i.e., all the axioms of MT hold for actual truth. However, **A**(P) ↔ P differs from **Tr**(P) ↔ P in not being necessarily true. So perhaps MT needs to be strenghtened in order to exclude unintended interpretations of the truth predicate.

4. Minimalism and paradox

Philosophical theories of truth are faced with logical puzzles and paradoxes. Horwich's minimal theory is no exception. Horwich does not pay much attention to logical paradoxes, but it is clear that they affect what he has to say about truth. Here I am going to focus on two problems of a logical nature that confronts the minimal theory. First, there is an apparent paradox concerning the class of all the proposition that are axioms of MT; a fact that was pointed out to Horwich by Patrick Grim.[11] As a matter of fact, the paradox in question is a variant of one that was first formulated by Bertrand Russell (1903) concerning the class of absolutely all propositions. Secondly, there is the Liar paradox and the way it is handled by Horwich.

Let me begin by discussing the paradox concerning the class of all propositions due to Russell. In *The Principles of Mathematics* Russell writes:[12]

> If m be a class of propositions, the proposition "every m is true" may or may not be itself an m. But there is a one-one relation of this proposition to m: if n be different from m, "every n is true" is not the same proposition as "every m is true." Consider now the whole class of propositions of the form "every m is true," and having the property of not being members of their respective m's. Let this class be w, and let p be the proposition "every w is true". If p is a w, it must posses the defining property of w; but this property demands that p should not be a w. On the other hand, if p be not a w, then p does possess the defining property of w, and therefore is a w. Thus, the contradiction appears unavoidable.
>
> ...The totality of all logical objects, or of all propositions, involves, it would seem, a fundamental logical difficulty. What the complete solution of the difficulty may be, I have not succeeded in discovering; but as it affects the very foundations of reasoning, I earnestly commend the study of it to the attention of all students of logic.

In his later work, within the framework of the ramified theory, Russell resolves the paradox by denying the existence of a totality of all propositions that one can quantify over:

> Whatever we suppose to be the totality of propositions, statements about this totality generate new propositions which, on pain of contradiction, must lie outside this totality. It is useless to enlarge the totality, for that equally enlarges the scope of statements about the totality. Hence, there must be no totality of propositions, and "all propositions" must be a meaningless phrase.[13]

---

[11] Cf. Horwich (1998), p. 20, note 4.
[12] Russell (1903), Appendix B, p. 527-528.
[13] Russell (1908), (p. 154 in van Heijenoort (1967)).

The version of this argument that is due to Grim, concerns instead the class of all propositions, the class of propositions that are axioms of the minimal theory MT. Horwich writes:[14]

> ...the minimal theory cannot be regarded as *the set* of propositions of the form <<p> is true iff p>; for there is no such set. The argument for this conclusion is that if there were such a set, then there would be distinct propositions regarding *each* of its subsets, and there would have to be distinct axioms of the theory corresponding to these propositions. Therefore there would be a 1-1 function correlating the subsets of MT with some of its members. But Cantor's diagonal argument shows that there can be no such function. Therefore MT is not a set.

Horwich's proposed solution appeals to the technical notion of a set: the idea is that Grim's paradox about the totality of propositions can be resolved in the same way as the set-theoretical paradoxes (Cantor's paradox, Russell's paradox, etc.), i.e., by denying that the troublesome class forms a *set*. The collection of all axioms of MT is, according to Horwich, not a set but a *proper class*. The same holds, a fortiori, for the collection of all propositions.

It is not clear why Horwich thinks that Grim's paradox concerning the axioms of MT can be resolved by denying that they can be collected into a *set*. The paradoxes of Russell and Grim seem to be directed against the assumption that there is a *well-defined totality* of absolutely all propositions that it makes sense to quantify over, not against the assumption that this totality forms a set, for example, in the sense of belonging to Zermelo's cumulative hierarchy of sets.

The quantifiers of standard set theory range, according to the intended interpretation, over the totality of all sets – a proper class. Quantification over classes is not disallowed in standard set theory. Neither does Horwich hesitate to quantify over the collection of absolutely all propositions. For instance, he defines falsity as the absence of truth, by quantifying over all propositions:[15]

(1) $\neg\exists x$(x is false $\leftrightarrow$ x is a proposition $\wedge$ x is not true).

and one of the points of the theory is that it can accommodate general statements like

(2) $\forall x \forall y$(if x implies y and x is true, then y is true).

It is the assumption that there is a *well-defined totality* of all propositions that one can quantify over that leads to the paradoxes of Russell and Grim, rather than the assumption that there is a *set* of all propositions.

---

[14] Horwich (1998), p. 20, note 4.
[15] Horwich (1998), p. 71.

To see this, let us carefully restate Russell's paradox about propositions. We make the following assumptions:

(i) There is a class **P** of *all* propositions.

(ii) For any class X of propositions, there is a proposition λ_X such that λ_X is true iff all the propositions in X are true. Intuitively, λ_X is the proposition $\langle \lambda x(x \in X \rightarrow x \text{ is true}) \rangle$. Let us call λ_X the *characteristic proposition* of the class X.

(iii) For all classes X, Y of propositions, if $X \neq Y$, then $\lambda_X \neq \lambda_Y$. The, intuitive motivation for this is that if X and Y are distinct, then it is possible to have a certain attitude (for instance, belief) towards λ_X without having the same attitude towards λ_Y.

Notice that there is no assumption here that classes are like sets, objects that can themselves be members of classes.

Assumptions (i) and (iii) are accepted by Horwich in his discussion of Grim's argument. And (ii) should be acceptable to him in the light of his acceptance of quantification over proper classes of propositions.

We now say that a proposition x is *irreflexive* if, and only if, for some class X of propositions, x is the characteristic proposition λ_X of X and x does not belong to X. That is, for every proposition x in **P**,

x is irreflexive iff $(\exists X)(x = \lambda_X \wedge x \notin X)$.

We now make an additional assumption:

(iv) The predicate "irreflexive" defines a class I of propositions. That is, there is a class I such that:
$I = \{x \in \mathbf{P}: (\exists X)(x = \lambda_X \wedge x \notin X)\}$.

If there is a class of absolutely all propositions, it seems hard to deny that there is a class of all irreflexive propositions. Informally, we can argue for (iv) in the following way: Given (i), both the class **P** of all propositions and the collection $\wp(\mathbf{P})$ of all classes of propositions are *well-defined totalities*. This means, in particular, that we can define properties of propositions by quantifying over these totalities. Hence, there is a well-defined property of a proposition being irreflexive. Given this property, we can define the corresponding class I of propositions.

Consider now the characteristic proposition λ_I of the class I of all irreflexive propositions. This proposition exists by assumptions (ii) and (iv). The proposition λ_I belongs to **P** by assumption (i). Does the proposition λ_I belong to I or not?

By the assumption (i), λ_I belongs to **P**. Hence,

$$\lambda_I \in I \text{ iff } (\forall X)(\lambda_I = \lambda_X \rightarrow \lambda_I \in X).$$

But by assumption (iii),

$$\lambda_I = \lambda_X \text{ iff } I = X.$$

Thus,

$$\lambda_I \in I \text{ iff } \lambda_I \notin I.$$

From this we can infer, both $\lambda_I \in I$ and $\lambda_I \notin I$, so we have a paradox.

Of the assumptions (i) - (iv), the one that seems easiest to reject seems to be (i), the assumption that there is a *well-defined totality* of absolutely all propositions. But if we reject that assumption, we must reject the assumption, built into Horwich's theory, that it is meaningful to quantify over the collection of all propositions. Then, we cannot make the claim that the minimal theory MT contains an axiom for every proposition. There is no meaningful notion of an arbitrary proposition and there is no minimal theory of truth that is applicable to arbitrary propositions.

After having rejected the assumption that there is a class of absolutely *all* propositions, Russell's argument does not lead to a paradox any more. The conclusion, then, is only that the proposition λ_I cannot belong to the given class of propositions. That is, we can look upon the argument as a proof by *diagonalization* that for any well-defined totality **P** of propositions there are propositions that do not belong to the totality **P**.

It is now time to turn to the Liar paradox. Let P be a property of propositions and consider the proposition that all propositions having the property P are false. We call this proposition λ_P, i.e.,

(1) $\lambda_P = \langle \forall x(P(x) \rightarrow x \text{ is not true}) \rangle$.

Suppose also that λ_P itself has the property P and that λ_P is the only proposition having the property P, that is:

(2) $\forall x(P(x) \rightarrow x = \lambda_P)$.

We could, for example, let P be the property of being one of the propositions that are formulated in the box (below) and let λ_P be the proposition expressed by the sentence in the box:

ALL PROPOSITIONS FORMULATED IN THIS BOX ARE FALSE

Assuming that the sentence in the box really expresses a (unique) proposition, we can verify that this proposition is the one and only proposition having the property P. Hence, it seems that both conditions (1) and (2) are satisfied.

We now continue to reason as follows:

(3)	$\Box_P$ is true	Assumption
(4)	$P(\Box_P)$	from (2) by predicate logic
(5)	$\langle\Box x(P(x) \leftrightarrow$ x is not true$)\rangle$ is true	(1), (3) by the logic of identity
(6)	$\Box x(P(x) \leftrightarrow$ x is not true)	5, by the (E) schema
(7)	$\Box_P$ is not true	4, 6 by predicate logic
(8)	$\Box_P$ is not true	3-7, RAA
(9)	$\Box x(P(x) \leftrightarrow$ x is not true)	2, 8 by predicate logic
(10)	$\langle\Box x(P(x) \leftrightarrow$ x is not true$)\rangle$ is true	9, by the (E) schema
(11)	$\Box_P$ is true	1, 10 by the logic of identity
(12)	$\Box_P$ is true $\wedge$ $\Box_P$ is not true.	contradiction.

Horwich considers the following options for avoiding the paradox:

(1) to deny classical logic; (2) to deny that the concept of truth can be coherently applied to propositions like $\Box_P$ that themselves involve the concept of truth; (3) to deny that the sentence in the box expresses a proposition; and (4) to reject certain instances of the (E) schema. Of these solutions he rejects all except (4).

Thus, Horwich's way out is to exclude the problematic proposition:

($\Box$) $\langle\langle\Box x(P(x) \leftrightarrow$ x is not true$)\rangle$ is true iff $\Box x(P(x) \leftrightarrow$ x is not true$)\rangle$

from the axioms of MT. Then, the paradoxical conclusion cannot be derived in MT. However, the proposition ($\Box$) must be either true or not. Suppose it is true. Then, the theory MT is incomplete, since it does not contain every true instance of the schema (E). And although the contradiction cannot be derived in MT, it can still be derived from true premises by means of truth-preserving rules of inference. This conclusion is unacceptable, since it means that the paradoxical conclusion is true even though it cannot be derived in MT.

So the proposition ($\Box$) must be false (= not true). That is,

$\langle\Box x(P(x) \leftrightarrow$ x is not true$)\rangle$ is true iff $\Box x(P(x) \leftrightarrow$ x is not true)

is false. But this contradicts our strong feeling that every instance of the schema:

(E) The proposition that p is true iff p,

is analytically true. To claim that an instance of this schema is actually false is counterintuitive, to say the least. The conclusion is that Horwich does not have an acceptable solution to the Liar paradox.

There is, however, one way out that Horwich has not considered. We are free to look upon the Liar paradox as still another *diagonal argument* for the conclusion that there is no well-defined totality of absolutely all propositions. Assuming that the quantifier in the Liar sentence ranges over a certain totality **P** of propositions, then the Liar argument can be viewed as a proof that the proposition expressed by this sentence cannot itself, on pain of contradiction, be a member of **P**. We can deny that the Liar sentence expresses a proposition in the given totality **P**, without denying that it expresses a proposition at all. This is, in effect, Russell's way out in (1908).

5. Conclusion

Horwich's minimalism is built on the idea that there is a well-defined totality of absolutely all propositions and that there exists a single property of truth that is applicable to all the propositions in that totality. In section 3, I argued that it is questionable whether Horwich's minimal theory of truth can be formulated for such a comprehensive collection of propositions in a way that does not itself presuppose the notion of truth. The (E) schema, that Horwich uses to specify his axioms, defines truth only for propositions that are expressed by sentences of a language. For propositions that are not so expressible, the schema does not specify any truth-axioms. And even if we could give an abstract algebraic description of the collection of all axioms of the theory MT, it is still questionable whether we could define the *theorems* of the theory without presupposing the notion of truth. Such an enterprise would involve defining a notion of logical consequence applicable to absolutely all propositions. It is doubtful whether this can be done – at least in a way that does not presuppose the concept of truth.

In section 4, we saw reasons to question the coherence of the concept of a totality of absolutely all proposition. In view of these observations, we might have to give up the dream of a 'global' theory of truth, and instead devise 'local' theories that are relativized to given languages or classes of propositions. After we have subjected minimalism to these drastic changes, what will then be left of its appealing simplicity? Will it still deserve to be called 'minimalist'? To be a minimalist with respect to truth, is perhaps more difficult than it first appeared.

References

Frege, G., 1984, 'Thoughts' in G. Frege, *Collected Papers on Mathematics, Logic, and Philosophy*, ed. B. McGuinness, trans. P. Geach and R. H. Stoothoff, Oxford: Basil Blackwell, 1984, pp. 351-72.

Horwich, P., (ed.), 1994, Theories *of Truth*, The International Research Library of Philosophy 8, Darthmouth Publishing Company, Brookfield, Vermont.

Horwich, P., 1998, *Truth*, second edition, Oxford University Press, Oxford. First edition published by Basil Blackwell, Oxford in 1990.

Kripke, S., 1975, 'Outline of a theory of truth', *Journal of Philosophy* 72, pp. 690-716.

Russell, B., 1903, *The Principles of Mathematics,* Cambridge University Press; second ed. with new introduction, George Allen and Unwin, London, 1937.

Russell, B., 1908, 'Mathematical Logic as Based on the Theory of Types', *American Journal of Mathematics* 30, 22-262. Reprinted in van Heijenoort, (1967).

Soames, S. 1999, *Understanding Truth*, Oxford University Press, New York.

Strawson, P. F., 1950, 'Truth', *Proceedings of the Aristotelian Society,* Supplement 24, pp. 129-56. Reprinted i Horwich (1994).

Sundholm, G., 1994, 'Proof-theoretical semantics and Fregean identity criteria for propositions', *The Monist*, vol. 77, no. 3, pp. 294-314.

Tarski, A., 1983, 'The Concept of Truth in Formalized Languages', in *Logic, Semantics and Metamathematics*, Papers from 1923 to 1938 by Alfred Tarski, Translated by J. H. Woodger, Second edition edited and introduced by John Corcoran, Hackett Publishing Company, Indianapolis, Indiana (First edition published in 1956 by Oxford University Press).

Tarski, A., 1943, 'The Semantic Conception of Truth', *Philosophy and Phenomenological Research*, 4, pp. 341-75. Reprinted in Horwich 1994.

Tarski, A., 1969, 'Truth and Proof', *Scientific American*, June 1969, pp. 63-77.

van Heijenoort, J., 1967, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*, Harvard University Press, Cambridge University Press, Cambridge, Massachusetts.