The liar and the new t-schema

El mentiroso y el nuevo esquema T

Stephen Read
University of St Andrews, United Kingdom. slr@st-andrews.ac.uk

Recibido el 30 de septiembre de 2010 y aprobado el 25 de octubre de 2010

Resumen

Desde que Tarski publicó su estudio sobre el concepto de verdad en los años 30, ha sido una práctica ortodoxa el considerar que toda instancia del esquema T es verdadera. Sin embargo, algunas instancias del esquema son falsas. Éstas incluyen las instancias paradójicas ejemplificadas por la oración del mentiroso. Aquí se demuestra que un esquema mejor permite un tratamiento uniforme de la verdad en el que las paradojas semánticas resultan ser simplemente falsas.

Palabras clave

Bradwardine, Carnap, Davidson, esquema T, paradoja del mentiroso, Tarski, teoría de la verdad.

Abstract

Since Tarski published his study of the concept of truth in the 1930s, it has been orthodox practice to suppose that every instance of the T-schema is true. However, some instances of the schema are false. These include the paradoxical instances exemplified by the Liar sentence. It is shown that a better schema allows a uniform treatment of truth in which the semantic paradoxes turn out to be simply false.

Key words

Bradwardine, Carnap, Davidson, T-scheme, liar paradox, Tarski, truth-theory.

1. TARSKI'S T-SCHEME

Carnap, in his *Autobiography*[1] tells an anecdote about a meeting with Tarski in Vienna in the early 1930s. Both Carnap and Tarski had been struggling with the problem of providing a satisfactory definition of truth. Carnap asked Tarski to explain the basic idea: how can one define the truth of an empirical sentence like 'This table is black'. "'This table is black' is true," replied Tarski, "*if and only if this table is black*". Coffa recalls that Carnap later described this as the moment when "*the scales fell from my eyes*"[2].

Over the years since Tarski published his study of the concept of truth[3], it has been an unquestioned orthodoxy that every instance of the T-scheme:

x is true if and only if p (T)

is true, where what replaces 'x' is a name of a sentence whose translation into the metalanguage replaces 'p'. Indeed, in the homophonic case of identity (in sound or in fact) of object language and metalanguage, what replaces 'x' is a quotation-name of the sentence which replaces 'p'.

Nonetheless, not all instances of (T) are true, even where the very same sentence replaces 'x' and 'p', the first in quotation-marks. For example,

'I am tired' is true if and only if I am tired (1)

and

'That book was stolen' is true if and only if that book was stolen (2)

are not true unless the person who utters (1) is the same as the utterer of the sentence cited, and unless the book indicated by the demonstrative pronoun in (2) is the very same as that demonstrated in the sentence whose truth is at stake, as Davidson[4] pointed out. Moreover, even in

[1] CARNAP, R. "Intellectual autobiography in P. Schilpp". In: The Philosophy of Rudolf Carnap. (pp. 3-84). La Salle, IL: Open Court, 1963. p. 60.
[2] COFFA, J. A. The Semantic Tradition from Kant to Carnap: To the Vienna Station. Cambridge: Cambridge University Press, 1991. p. 304.
[3] Cfr. TARSKI, A. (1936). "The establishment of a scientific semantics". In: WOODGER, J. (Ed.). Logic, Semantics, Metamathematics, 2d Ed. Corcoran, J. (Tr.). Indianapolis: Hackett, 1983. p. 152.
[4] Cfr. DAVIDSON, D. "True to the facts". In: Journal of Philosophy, 1969. vol. 66, p. 304. Reprinted in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984. p. 37-54.

the absence of indexicals, the validity of the T-scheme is not assured. Hintikka's example[5] was

'Any man is mortal' is true if and only if any man is mortal (3)

'Any' always takes widest scope, so that whereas 'any' in the quoted sentence acts as a universal quantifier within it, the occurrence of 'any' on the right-hand side functions as a universal quantifier with wide scope over the whole of (3), and so as an existential quantifier with scope restricted to the right-hand side. Thus homophonic translation fails.

Despite such counterexamples, (T) are often taken as a truism. For example, in his paper 'The Establishment of a Scientific Semantics', Tarski wrote:

> Quite generally, we shall accept as valid every instance of the form
>
> *the sentence x is true if and only if p*
>
> where '*p*' is to be replaced by any sentence of the language under investigation and '*x*' by any individual name of that sentence[6].

More recently, Paul Horwich wrote:

> for any declarative sentence 'p' our language generates an equivalent sentence 'The proposition that p is true'[7].

and

> (…) the concept of truth is entirely captured by stipulating the equivalence schema, 'The proposition that *p* is true if and only if *p*' –where *p* can be replaced by any declarative sentence[8].

[5] HINTIKKA, J. "A counterexample to Tarski-type truth-definitions as applied to natural languages". In: Philosophia, 1975. vol. 5, p. 207.
[6] TARSKI. The establishment of a scientific semantics. Op. cit., p. 404.
[7] HORWICH, P. Truth. 2d. Ed. Oxford: Clarendon Press, 1998. p. 4-5.
[8] HORWICH, P. "In the truth domain". In: Times Literary Supplement, 4711, 1993. p. 28.

Lastly, Aladdin Yaqub writes:

> The concept of truth is *completely* and *correctly* defined by the Tarskian scheme, whose instances are *all* the biconditionals obtained from the phrase '*x* is true if and only if *p*' by substituting any sentence for '*p*' and any expression which stands for that sentence for '*x*'[9].

Yaqub invokes Tarski, Ayer, Horwich and Gupta in support of his claim that "*the Tarskian schema is the most fundamental intuition about the concept of truth*"[10].

Tarski did not propose (T) as a definition of truth, though others, e.g., Horwich, have done so since. They all describe (T) as a truism, as something so obviously correct as to need no argument, at least, provided the obvious counterinstances, such as (1), (2) and (3) are excluded. Tarski himself presented (T) as a condition of what he called "material adequacy". It was a necessary condition on any adequate definition of truth that it generates all instances of (T) as theorems:

> A formally correct definition of the symbol '*Tr*' (…) will be called an adequate definition of truth if it has the following consequence:
>
> all sentences which are obtained from the expression '*x* ∈ *Tr* if and only if *p*' by substituting for the symbol '*x*' a structural-descriptive name of any sentence of the language in question and for the symbol '*p*' the *expression which forms the translation of this sentence into the metalanguage*[11].

Tarski's account was for many years described as a version of the correspondence theory of truth –an acceptable version of a theory relieved of its suspect ontology of facts and correspondence. The (T)-scheme presents in a stark and nominalistically acceptable form the correlation of language, on the left-hand side, and world on the right-hand side. What replaces '*x*' refers to a sentence or other truth-bearer, and what replaces '*p*' refers to, or describes, how things are. Thus the (T)-scheme can be summed up as expressing the idea that a sentence is true if things are as it says they are:

[9] YAQUB, A. *The Liar Speaks the Truth*. Oxford: Oxford University Press, 1993. p. 3.
[10] *Ibid*., p. 27.
[11] TARSKI. The establishment of a scientific semantics. Op. cit., p. 187-188.

x is true if and only if things are as *x* says they are (CP)

Crispin Wright describes (CP) as "*the correspondence platitude*"[12], and Julian Dodd agrees: "*a platitude it certainly is (...) it says nothing that would be disagreed with by a correspondence theorist, coherence theorist or deflationist*"[13].

In this paper, I want to show that (CP) is at best equivocal, and (T) is false. This is not just a matter of showing that not all instances of (T) are true. Many proponents of (T) admit that. They nonetheless claim that (T) offers a good account (partial or complete) of the truth-conditions of those sentences for which it is true. However, this is not so. Where it is true, (T) is true by lucky chance. Where true, its instances are also instances of a better scheme, the right scheme of truth. (T) is very far from being a truism, or a good test of the adequacy of a truth-theory.

2. HOW THINGS ARE

How can things be as *x* says they are unless *x* is true? Clearly, 'snow is white' says that snow is white, and 'snow is white' is true just when snow is as 'snow is white' says it is, that is, when snow is white. But 'Every even number greater than 2 is the sum of two primes' not only says that every even number greater than 2 is the sum of two primes, but also in particular that, e.g., 4 is the sum of two primes. So 4's being the sum of two primes is as Goldbach's Conjecture says things are. But Goldbach's Conjecture is not true just because 4 is the sum of two primes. 4's being the sum of two primes is necessary for the truth of Goldbach's Conjecture, but it is not sufficient. So (CP) needs more careful expression. What was meant was

x is true if and only if things are wholly as *x* says they are, (S)

or

x is true if and only if however *x* says things are, they are (S')

A sentence says many things, and however it says they are, they must be, for the sentence to be true. Anything of the form '*p* and *q*' says *inter alia* that *p*, but the truth of *p* does not suffice for that of '*p* and *q*'.

[12] WRIGHT, C. J. G. Truth and Objectivity. Cambridge, MA: Harvard University Press, 1992. p. 27.
[13] DODD, J. "Is truth supervenient on being?" In: Proceedings of the Aristotelian Society, 2001-2002. vol. 102, p. 75.

Indeed, the phrase 'how x says they are' or even 'however x says they are' is imprecise in another way. 'All men are mortal' says that if Socrates is a man then Socrates is mortal. But Socrates was a man. So if 'All men are mortal' is true, Socrates was mortal. Whether the universal sentence actually says that Socrates was mortal is, perhaps, unclear. If Socrates was not mortal, then 'All men are mortal' is false –unless Socrates was not a man after all. If x is true, things must be however x implies they are. Conversely, if things are however x implies they are, x will be true. "Safety first" would support the identification of 'however x says they are' with 'what x implies'. This may make 'how x says they are' more precise than is natural, but in the context of (CP) it is welcome precision. We find such a conception of 'saying that' as closed under implication in Wittgenstein's *Tractatus*: "*A proposition affirms every proposition that follows from it (…) If p follows from q, the sense of 'p' is contained in the sense of 'q'*"[14]. By 'follows from', Wittgenstein here clearly intended the logical truth of the conditional (as John Corcoran pointed out to me). Elementary propositions are logically independent and p follows from q only if the truth-conditions or truth-grounds (Sect. 5.101), of q contain those of p. For Wittgenstein, the conditional itself was a material one. I will follow him in this, for the purposes of this paper. Thus logical implication is presently strict implication, that is, the necessitation (or logical truth) of the material conditional. Certain move in the argument to follow work only for such an implication.

Let us represent 'x says that p' by '$x{:}p$', where 'x' designates a sentence, strict implication by '$\Rightarrow$' and the material conditional by '$\rightarrow$'.
Then we can represent the closure of 'saying that' under logical implication by

$$(\forall p, q)((p \Rightarrow q) \rightarrow (x{:}p \Rightarrow x{:}q)) \text{ (K)}$$

The context $x{:}p$ is an intensional context for p, allowing substitution only of logical equivalents (and then, only if we accept the closure of $x{:}p$ (saying that) under '$\Rightarrow$')[15]. The truth-condition given by (S) or (S′) is then represented by

[14] WITTGENSTEIN, L. Tractatus Logico-Philosophicus. Pears, D. & McGuinness, B. (Trs.). London: Routledge and Kegan Paul, 1961. Sects. 5.122 and 5.124.
[15] Indeed, if '$\Rightarrow$' were '$\rightarrow$' and x said anything, say p, then since $(p \rightarrow q) \vee (p \rightarrow \neg q)$ is a tautology, we would have for all q, $x{:}q \vee x{:}\neg q$, which would trivialize the whole project and render $x{:}p$ extensional. What one would like for '$\Rightarrow$' is some enthymematic conditional, which includes the case of Socrates above without trivializing the whole notion. But identifying such a conditional is, presently, I believe, an unrealized project.

$$\mathrm{T}x \Leftrightarrow \forall p(x{:}p \rightarrow p) \text{ (A)}$$

Here, '⇔' represents strict equivalence. For Tx and $\forall p(x{:}p \rightarrow p)$ do not just happen to be equivalent. The right-hand side of (S) gives a philosophical analysis, or explication, or truth. (A) contrasts with (T) in two important ways. First, (A) is an analysis of truth; (T) is merely, for Tarski, a condition of material adequacy, a test of the accuracy of his and others' theories. In particular, (A) is not the final analysis of truth: that awaits an analysis of 'saying that'. Thus (A) need not conflict with Tarski's requirement that all semantic notions be defined and eliminated, any more than does his own account of truth:

$$\mathrm{T}x \Leftrightarrow \forall s(s \text{ is a sequence} \rightarrow s \text{ satisfies } x).$$

Tarski defined truth in terms of satisfaction, but then proceeded to define satisfaction in non-semantic terms.

Secondly, (A) is a logical equivalence, whereas (T) was for Tarski merely a material equivalence. For Horwich, the partial definitions whose form is (T) are given as material equivalences, but, he says, their a priori role allows us to derive the corresponding strict equivalences[16]. That is fitting for his use of (T) in defining truth, a use which Tarski did not make. Accordingly, if (A) is a logical equivalence in providing the correct analysis of truth, we must assume that what a sentence says, it says of necessity too. Of course, any particular form of words expresses what it does only contingently. When I speak of the truth of sentences, and what they say, in what follows, I will be speaking of a sentence taken as meaning what it does –of a proposition, in Horwich's terminology, or of a sentence in a context of speaker, place and time.

If we take the more cautious line of rejecting (K), we can still capture (S) and (S′) by

$$\mathrm{T}x \Leftrightarrow \forall p(\exists q(x{:}q \wedge (q \Rightarrow p)) \rightarrow p)^{17} \text{ (A}'\text{)}$$

Whereas (A) expresses the thought that a sentence is true if things are however it says they are, (A′) spells out that it is true if things are

[16] Cfr. HORWICH, P. Truth. 2d. Ed. Oxford: Clarendon Press, 1998. Chap. 2, Sect. 3.

[17] Note that (K), (A) and (A') should be thought of as expressions in type theory, where x has type ι (of individuals) and p has type o (of propositions). See: e.g., CHURCH, A. "A formulation of the simple theory of types". In: Journal of Symbolic Logic, 1940. vol. 5, p. 56-68.

however what it says (logically) implies them to be. They are equivalent just if we identify what a sentence says with what it implies. Of course, taking '→' to be material implication, (A) and (A′) will make true all sentences which say nothing. One solution is to conjoin $(\exists p)\ x{:}p$ to the right-hand side of (A) and (A′). But for simplicity, I will assume in what follows that each sentence to which we apply (A) and (A′) does say something.

Another possible objection to (A) –and to (S)– concerns ambiguity. A different way for a sentence to say more than one thing is for it to be ambiguous. But if x has two meanings, e.g., 'Visiting relatives can be boring', we do not require that both senses be the case for x to be true –either will suffice.

In reply, remember that the point of formalization is to avoid ambiguity. Accordingly, we must remove ambiguity from x and require that all sentences governed by (A) be unambiguous. For example, we might deem 'Visiting relatives can be boring' to be a pair of homophones, as we do with 'I have a pair/pear'. Nonetheless, an unambiguous sentence, like 'All men are mortal', can say many things –that Socrates is mortal, that Plato is mortal and so on. To be true, all the things it says must hold– things must be however it says they are.

The crucial advantage of (A), or (A′), over (CP) is that it makes clear that everything which a sentence says or implies must be the case for the sentence to be true. (CP) is equivocal: x may say things are in many ways. They are all severally necessary, but only jointly sufficient. Things must be not only how x says they are, but however x says they are, in order that x be true.

Tarski's later discussion helps us see where he went wrong. Denoting 'snow is white' by 'S', he writes:

> What do we mean by saying that S is true…? The answer to this question is simple: (…) by saying that S is true we mean simply that snow is white (…) By eliminating the symbol 'S' we arrive at the following formulation:
>
> (1) 'snow is white' is true if and only if snow is white[18].

[18] TARSKI. The establishment of a scientific semantics. Op. cit., p. 64.

But he generalizes this later on the same page:

> Partial definitions of truth analogous to (1) (…) can be constructed for other sentences as well. Each of these definitions has the form:
>
> (3) ‘*p*’ is true if and only if *p*,
>
> where ‘*p*’ is to be replaced on both sides of (3) by the sentence for which the definition is constructed[19].

That snow is white is, arguably, all that ‘snow is white’ says. So the right-hand side of (1) gives a sufficient, as well as necessary condition for its truth. But not all cases are as straightforward as this. In general we need to do more than simply repeat ‘*p*’ on the right-hand side. (T) is a special case of (A) where, as for ‘snow is white’; a sentence reveals explicitly all that it says. The crucial importance of using (A), and the disastrous consequences of using (T) in its place, will be demonstrated in Sects. 3 and 4.

3. THE “LIAR PARADOX”

Tarski and others exclude the Liar paradox from the scope of the T-scheme. Tarski did it by forbidding semantic closure, and so making the semantic paradoxes inexpressible[20]. Horwich[21] simply excludes the paradoxical instances of (T) from the infinite conjunction of its instances which he takes to define truth. The consequence is that neither author can offer an account of the semantics of the Liar paradox and its ilk. For the one, the material test of the correctness of his theory of truth is that the right instances of (T) be derivable in the theory; for the other, the theory of truth is constituted by the conjunction of the correct instances. Each implicitly concedes, therefore, that not all instances of (T), even the homophonic ones where what replaces ‘*x*’ is a quotation-name of the sentence which replaces ‘*p*’, are true. Nonetheless, they persevere in keeping (T) as the cornerstone of their theory of truth, rather than looking for a better scheme. A better scheme would be one all of whose instances are true, one which did not need an ad hoc exclusion clause, excluding any counterexamples by *fiat*.

---

[19] *Ibidem*.
[20] See TARSKI. Sect. 1 and p. 402.
[21] See HORWICH. Sect. 10.

There are many ways to formulate the Liar paradox, but the general idea is to find a sentence-name L such that L and 'L is not true' are interchangeable in light of some fact about L (e.g., where L is written, or through the correlations of some system of Gödel-numbering, and so on). Recall that we are assuming that a sentence (or proposition) necessarily says what it says, even, for example, in the case of contingent paradox[22]. Then since L implies itself, it implies that L is not true (by substitution), and indeed, if 'says that' is closed under implication (as proposed in Sect. 2) or if the connection is close enough, L says that L is not true. Thus we can substitute 'L is not true' for 'p' and 'L' for 'x' in (T), or (CP), to obtain the paradoxical instance:

L is true if and only if L is not true (P)

Without serious and damaging revisions to logical principles, (P) is false, for it entails:

L is true and L is not true(P')

Those who reject Ex Falso Quodlibet (the principle that a contradiction implies every sentence) may attempt to maintain that (P') is true without collapsing into triviality. But the Curry paradox, finding for an arbitrary sentence q, a sentence Cq which is interchangeable with 'If Cq is true then q', produces the paradox of triviality without recourse to contradiction. Substituting 'Cq' for 'x' in (T), we obtain

Cq is true iff if Cq is true then q,

and by familiar steps of Modus Ponens and Contraction, we get a proof of q. Since q was arbitrary, the truth-theory based on (T), and (CP), is shown to be unacceptable. (T) is simply false, for it abounds in false instances.

4. TRUTH-THEORY WITHOUT PARADOX

What we have recalled in Sect. 3 is that if we look to (T) for an account of the semantics of L and Cq, the result will be trivial and paradoxical –in a word, wrong. The mistake encapsulated in (T) is to suppose that the expression replacing 'p' on the right-hand side fully expresses

[22] See KRIPKE, S. "Outline of the theory of truth". In: Journal of Philosophy, 1975. vol. 72, p. 691.

its own truth-condition. This may seem a plausible assumption. But at least in the case of the semantic paradoxes, as we will see, it is a mistaken assumption. For sentences imply other things which they do not themselves explicitly say. By the lights of (A), they say them nonetheless. By the lights of (A) or (A′), if those other things are not the case, then what the original sentence says cannot hold either, and so the original sentence is false. Reading (CP) as (T) misses that implication. (T) Identifies the truth of a sentence with things being how it says they are; but not with however it says they are. Reading (CP) as (S), and so as (A) or (A′), includes in the truth-conditions of a sentence all its consequences.

Consider L, for example. L says that L is not true, that is, L: ¬TL. It may say more: call it q, that is, supposes L: (¬TL ∧ q). (If ¬TL is all that L says, we can simply let q be ¬TL, or a tautology.) By (A),

$$\mathrm{T}L \Leftrightarrow \forall p(L\colon p \rightarrow p)$$

i.e.,

$$\mathrm{T}L \Leftrightarrow (\neg \mathrm{T}L \wedge q)$$

since $\neg \mathrm{T}L \wedge q$ is all L says, whence

$$\neg \mathrm{T}L \Rightarrow \neg\ (\neg \mathrm{T}L \wedge q)$$

i.e.,

$$\neg \mathrm{T}L \Rightarrow (\mathrm{T}L \vee \neg q).$$

If saying that is closed under '⇒', as (K) says, then, since L: $\neg \mathrm{T}L$,

$$L\colon (\mathrm{T}L \vee \neg q)$$

and so (since L:q)

$$L\colon ((\mathrm{T}L \vee \neg q) \wedge q)$$

whence

$$L\colon \mathrm{T}L$$

since $((\mathrm{T}L \vee \neg q) \wedge q) \Rightarrow \mathrm{T}L$. Thus, if L says $\neg \mathrm{T}L$, it says $\mathrm{T}L$ as well. That is, $L\colon(\neg \mathrm{T}L$ c $\mathrm{T}L)$. So by (A),

$$TL \Leftrightarrow (\neg TL \wedge TL \wedge \ldots)$$

whence

$$\neg TL.$$

The argument can be repeated using (A′) in place of (A), and dropping the assumption that saying that is closed under '$\Rightarrow$'. Assume again that L: $(\neg TL \wedge q)$. By (A′), since $\neg TL \wedge q$ is all that L says,

$$TL \Leftrightarrow \forall p(((\neg TL \wedge q) \Rightarrow p) \to p),$$

so

$$\neg TL \Rightarrow \exists p(((\neg TL \wedge q) \Rightarrow p) \wedge \neg p),$$

i.e.,

$$\neg TL \Rightarrow \exists p(\neg p \wedge (\neg p \Rightarrow (TL \vee \neg q))),$$

so

$$\neg TL \Rightarrow \exists p(\neg p \wedge (TL \vee \neg q)),$$

whence

$$\neg TL \Rightarrow (\neg TL \to \neg q),$$

which by Contraction yields,

$$\neg TL \Rightarrow \neg q.$$

Contraposing, we have

$$q \Rightarrow TL.$$

Using (A′) once again,

$$TL \Leftrightarrow (\neg TL \wedge TL \wedge \ldots)$$

whence

$$\neg TL.$$

It may not be immediately obvious what a breakthrough this is. Of course L is not true. By (A)

$$\mathrm{T}L \Leftrightarrow (L\colon \neg\mathrm{T}L \rightarrow \neg\mathrm{T}L) \wedge \ldots$$

whence

$$\mathrm{T}L \Rightarrow (L\colon \neg\mathrm{T}L \rightarrow \neg\mathrm{T}L).$$

Since L:$\neg$TL,

$$\mathrm{T}L \Rightarrow \neg\mathrm{T}L$$

and so

$$\neg\mathrm{T}L.$$

But the longer argument concerning L shows more than that L is not true. In Sect. 3, using the (T)-scheme, we concluded not only that L was not true but also that L was true, and so landed in paradox. This reasoning fails when the truth-conditions of L (and all sentences) are given by (A), or (A′). To use (T) to infer that L is true, all we need to show is that L is not true, as in Sect. 3. To use (A) or (A′) to show that L is true we need to show both that L is not true and, in addition, that L is true. But that is impossible. L cannot be true, for to be true, it would have to be both true and not true. Nothing can be both true and not true. So L cannot be true. L is not only not true, it cannot (also) be true, as (A) reveals.

Note that I have not shown that ¬TL ⇒ TL. The closure principle (K) implies that L says whatever is implied by what L says, and, by the argument above, it follows from the fact that L: ¬TL that L: TL. But ¬TL ⇒TL is in fact false. It entails ¬TL → TL, which is equivalent to TL. But TL is false, as we have seen, and consequently, so is ¬TL ⇒ TL.

The problem with (T), as a condition of truth, is that it implies the impossible, for it declares (unless, of course, suitably restricted) that L is both true and not true. Thus (T) is false. (T) has the false instance:

$$\mathrm{T}L \Leftrightarrow \neg\mathrm{T}L.(\dagger)$$

The mistake made by Tarski, Horwich and others is to think that a suitably restricted version of (T) can serve as a determiner of the right theory of truth. It cannot. Either for no good reason other than to preserve the (universal) truth of (T), the counterinstances are deemed ill-formed, a thoroughly ad hoc manoeuvre; or the counterinstances are simply excluded from the theory of truth by *fiat*, again ad hoc and involving the further consequence that there are apparently well-formed sentences (such as *L*) for which there is no theory of truth. How, unless we can apply our theory of truth to *L* can we discover that it is paradoxical?

The solution is ready to hand. Abandon (T) and realise that the correct theory of truth is given by (A) –or (A′), depending on one's preferred theory of meaning– governing all well-formed sentences in a semantically closed language. As applied to *L*, we obtain the correct truth-condition:

$$\mathrm{T}L \Leftrightarrow (\neg \mathrm{T}L \wedge \mathrm{T}L).(\dagger\dagger)$$

I said that (CP) equivocates, and perhaps (T) does so too, or at least has been misunderstood. Tarski requires that any adequate theory of truth for a language must generate all instances of (T) where what replaces '*p*' is a translation of the sentence whose name replaces '*x*'. But a proper translation must spell out all that the sentence says. So arguably, (†) is not a correct instance of (T), for the right-hand side does not say all that *L* says. Rather, the correct instance of (T) is (††), whose right-hand side says all that *L* says.

5. PARADOX WITHOUT TRUTH

Tarski[23] observed that one can repeat the Liar reasoning in the absence of a truth-predicate, simply with a quotation-functor. Even after replacing (T) by (A), there is a danger that we might similarly find paradox has returned, through use of the 'says that' predicate. For consider

$$\forall p(C{:}p \rightarrow \neg p)(\mathrm{C})$$

Clearly C: $\forall p(C{:}p \rightarrow \neg p)$. Suppose that $\forall p(C{:}p \rightarrow \neg p)$. By Universal Instantiation,

[23] TARSKI, The establishment of a scientific semantics. Op. cit., p. 161.

$$C\colon \forall p(C{:}p \to \neg p) \to \neg \forall p(C{:}p \to \neg p)$$

so by Modus Ponens,

$$\neg \forall p(C{:}p \to \neg p)$$

and so by *reductio*,

$$\neg \forall p(C{:}p \to \neg p),$$

whence

$$\exists p(C{:}p \wedge p).(*)$$

We might be tempted to infer

$$\forall p(C{:}p \to \neg p)$$

from (*), since that is what C seems to say. But we have already shown $\neg \forall p(C{:}p \to \neg p)$, and so contradiction would have returned, without the involvement of (T) or (A) and the truth-predicate.

However, although $C\colon \forall p(C{:}p \to \neg p)$, this may not be all that C says. Suppose C also says that q, that is, $C\colon (\forall p(C{:}p \to \neg p) \wedge q)$. Then by (A),

$$\mathrm{T}C \Leftrightarrow (\forall p(C{:}p \to \neg p) \wedge q)$$

whence

$$(\neg \forall p(C{:}p \to \neg p) \vee \neg q) \Rightarrow \neg \mathrm{T}C$$

and so

$$\neg \forall p(C{:}p \to \neg p) \Rightarrow \neg \mathrm{T}C,$$

whence, since $\forall p(C{:}p \to \neg p) \Rightarrow \neg \forall p(C{:}p \to \neg p)$,

$$\forall p(C{:}p \to \neg p) \Rightarrow \neg \mathrm{T}C$$

and so

$$C{:}\neg \mathrm{T}C.$$

Thus it is true that $\exists p(C{:}p \land p)$, for $C{:}\neg TC \land \neg TC$. So from (*) we cannot legitimately infer that whatever C says is the case. Some of what C says is the case, namely, $\neg TC$. But (*) is no ground for inferring that whatever C says is the case. In particular, $\forall p(C{:}p \rightarrow \neg p)$ is not the case, since C is false.

Although this frees C from the threat of immediate paradox, it raises a further worry. For it is crucial to the present solution to the paradoxes that not every proposition say of itself that it is true. The reason that L, C and Cq turn out not to be paradoxical is that they say of themselves that they are true, like the Truth-teller; but since they also directly or indirectly say that they are not true, things cannot be as they say they are, and so by (A) they are not true but false.

But if we could show that every sentence said (albeit indirectly) of itself that it was true, the truth-condition (A) would be crippled. For (A) says that a sentence is true only if things are as it says they are, and if part of what it said was that it was true, (A) would, correctly but unhelpfully, endorse a sentence as true if it was true, but would not reveal whether it was true or not. Every sentence (bar the paradoxical ones) would become a Truth-teller, true if true and false if not, but of undetermined truth-value.

What induces this anxiety that every sentence is a truth-teller? The thought is this: implicit in endorsing (A) as the right truth-condition (in place of (T)) was that 'saying that' is closed under implication –every sentence says (implicitly, or indirectly) whatever is implied by what it says. But by (A), things being as it says they are implies that a sentence is true. It seems to follow that every sentence says that it is true. That it is true is implied by whatever it says, so if it says whatever it implies, it says that it is true.

Fortunately, this is a specious argument, which we can see from our formalization of it. (A) says

$$\mathrm{T}x \Leftrightarrow \forall p(x{:}p \rightarrow p)$$

and the closure of 'saying that' under implication, (K), is

$$(\forall p, q)((p \Rightarrow q) \rightarrow (x{:}p \Rightarrow x{:}q)).$$

These do not entail x:Tx. For a counterexample, read x:p as $p \in X$, where X is a theory, that is, a collection closed under logical implication. Then the closure assumption is true, i.e.,

$$(\forall p, q)((p \Rightarrow q) \rightarrow (p \in X \Rightarrow q \in X)).$$

We want x:Tx to be false. (A) identifies Tx with $\forall p(p \in X \rightarrow p)$. So we need to show that it is possible that $\forall p(p \in X \rightarrow p) \notin X$. What $\forall p(p \in X \rightarrow p)$ says is that X is correct –that X contains only truths. But theories can be mistaken, and it would be an arrogant theory which included the statement $\forall p(p \in X \rightarrow p)$, which says that everything X says is true. Thus x:Tx does not follow from (A) together with the closure principle.

Thus the argument that every sentence says of itself that it is true is sophistical. A sentence says whatever it says, but it does not say that whatever it says is the case –it does not say that things are as it says they are. The Truth-teller says that. But that is what is wrong with the Truth-teller. To say that things are as it says they are is to say nothing. Things must be as a sentence says they are for it to be true, but it does not (in general) say that things are as it says they are.

It follows that paradox, with or without truth, is prevented by the adoption of the correct account of truth, namely, the truth-condition (A).

6. SOME HISTORY

The argument in Sect. 4 to the effect that L says that L is true since it says that L is not true is due to Thomas Bradwardine, briefly Archbishop of Canterbury in 1349 before succumbing to the Black Death. His treatment of the semantic paradoxes dates from some 25 years earlier, when he was a young teacher at the University of Oxford. His solution was taken up and adapted by John Buridan and Albert of Saxony at the University of Paris, and in Buridan's version was discussed in the 1950s and sixties by Ernest Moody (1953) and Arthur Prior (1962). However, both John and Albert omit the proof from Sect. 4., John (2001, 969) simply asserting that every sentence implies its own truth, Albert (1974, f. 43*rb*) inferring that every sentence says that it itself is true from the assumptions that every (affirmative) sentence says that its subject and predicate are identical, and that every such sentence whose subject and predicate are identical is true (and likewise for negatives). It was this reasoning of Albert's which inspired the sophistical reasoning at the close of Sect. 5.

But if every sentence says that it itself is true, or implies its own truth, then no sentence is true, or at least no sentence can be shown, using (A) or (A′), to be true. (A) simply says that any such sentence is true if it is true (and whatever else it says or implies is the case). John and Albert are shown to have no adequate theory of truth at all[24].

The same problem does not affect Bradwardine's treatment. He is not committed to saying that every sentence says of itself that it is true, only that paradoxical sentences such as *L* (and *C* and *Cq*, if *q* is not true) do so. His theory shows that they are simply not true, and so false, since he accepts Bivalence.

References

ALBERT OF SAXONY. (1974). Perutilis Logica. Venice 1522, repr. Olms.
BURIDAN, J. (2001). Summulae de Dialectica. Klima, G. (Tr.). New Haven, CT: Yale University Press.
CARNAP, R. (1963). "Intellectual autobiography in P. Schilpp". In: The Philosophy of Rudolf Carnap. (pp. 3-84). La Salle, IL: Open Court.
COFFA, J. A. (1991). The Semantic Tradition from Kant to Carnap: To the Vienna Station. Cambridge: Cambridge University Press.
CHURCH, A. (1940). "A formulation of the simple theory of types". In: Journal of Symbolic Logic, vol. 5, p. 56-68.
DAVIDSON, D. (1969). "True to the facts". In: Journal of Philosophy, vol. 66, p. 304-323. Reprinted in his *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984. p. 37-54.
DODD, J. (2001-2002). "Is truth supervenient on being?" In: Proceedings of the Aristotelian Society, vol. 102, p. 69-86.
HINTIKKA, J. (1975). "A counterexample to Tarski-type truth-definitions as applied to natural languages". In: Philosophia, vol. 5, p. 207-212.
HORWICH, P. (1993). "In the truth domain". In: Times Literary Supplement 4711, p. 28.
________. (1998). Truth. 2d. Ed. Oxford: Clarendon Press.
INSOLUBILIA. (2009). "Inncyclopedia Britannica". In: Roure, M. L. La problematique des propositions insolubles au XIII*e* siecle et au debut

[24] A fuller treatment of John's, Albert's and Thomas' attempts at a solution was given in (READ, S. "The liar raradox from John Buridan back to Thomas Bradwardine". In: Vivarium, 2002. vol. 40, p. 189). I am grateful to John Corcoran and Ignacio Jané, among others at a workshop at the University of Santiago, Spain, for helpful comments on an earlier version of this paper and o Thomas Schmidt and others at the University of Göttingen, Germany, for further helpful comments.

du XIV*e*, suivie de l'edition des traites de W. Shyreswood, W. Burleigh et Th. Bradwardine. In: Archives d'Histoire Doctrinale et Littéraire du Moyen Age, vol. 37, p. 205-326.
KRIPKE, S. (1975). "Outline of the theory of truth". In: Journal of Philosophy, vol. 72, p. 690-716.
MOODY, E. A. (1953). Truth and Consequence in Medieval Logic. New York: Greenwood Press.
PRIOR, A. (1962). "Some problems of self-reference in John Buridan". In: Proceedings of the British Academy, vol. 48, p. 281-296.
READ, S. (2002). "The liar raradox from John Buridan back to Thomas Bradwardine". In: Vivarium, vol. 40, p. 189-218.
TARSKI, A. (1936). "The establishment of a scientific semantics" (Tr.). In: WOODGER, J. (Ed.). Logic, Semantics, Metamathematics. 2d Ed. Corcoran, J. (Tr.). Indianapolis: Hackett, 1983. p. 401-408.
________. (1983). Logic, Semantics, Metamathematics. 2d Ed. Woodger, J. (Ed.). Corcoran, J. (Tr.). Indianapolis: Hackett.
WITTGENSTEIN, L. (1961). Tractatus Logico-Philosophicus. Pears, D. & McGuinness, B. (Trs.). London: Routledge and Kegan Paul.
WRIGHT, C. J. G. (1992). Truth and Objectivity. Cambridge, MA: Harvard University Press.
YAQUB, A. (1993). The Liar Speaks the Truth. Oxford: Oxford University Press.