Do we need to reform the old T-scheme?

¿Necesitamos reformar el viejo esquema T?

Jan Woleński
University of Management and Information, Rzeszów, Finlandia. wol@if.edu.pl

Recibido el 30 de abril de 2012 y aprobado el 7 de mayo de 2012

Resumen

Este artículo comenta de manera crítica el análisis de Stephen Read del esquema T ofrecido en su artículo: "The Liar and the new T-scheme". Sostengo que Read malinterpreta los planteamientos de Tarski en algunos puntos y que el esquema A que introduce Read en su artículo, es o bien reducible al viejo esquema T (tarskiano) o bien plantea una serie de dudas esenciales.

Palabras clave

Falsedad, inconsistencia, intencionalidad, probabilidad, Read, Tarski, tautología, el mentiroso, verdad, oraciones T.

Abstract

This paper critically comments Stephen Read's analysis of the T-scheme offered in his paper: "The Liar and the new T-scheme". I argue that Read misinterprets Tarski in some points and that the A-scheme introduced in Read's paper is either reducible to the old (Tarskian) T-scheme or raises essential doubts.

Key words

Falsehood, inconsistency, intensionality, provability, Read, Tarski, tautology, the Liar, truth, T-sentences.

Read 2010 almost verbatim reproduces Read 2008. I was invited to make comments about the later paper (Woleński). In what follows I partly repeat my earlier critical comments about Read's interpretation of the T-scheme, but I also add new remarks and improve old ones; some of additions are inspired by other contributors participating in Rahman, Tulenheimo, Genot (2008), including Stehpen Read's replies in this volume (2008) as well as the previous discussion in *Discusiones Filosóficas* (Miller, Read, Sandu).

At first, let me remind Leśniewski's-Tarski's diagnosis of the Liar paradox. They pointed out (Tarski 1944) that the derivation of the paradox uses: I) self-referential sentences asserting semantic properties; II) the T-scheme, and III) classical, that is, bivalent logic. Hence, we can conceive three strategies in order to solve the Liar: i) to exclude self-referentiality; ii) to reject or modify the T-scheme; iii) to change logic. It would be mistaken to maintain that there is a solution free of costs or some artificiality. It concerns Tarski who choose i), Kripke (and many other logicians) who opted for ii) (Sandu makes several remarks about this way out), and Read who tries to reform the T-scheme. More precisely, Read argues that the old (Tarskian) T-scheme is inaccurate and proposes its modification. I would like to show that Read's reading of Tarski is incorrect and that his (Read's) leads to some difficulties.

Read says that, according to Tarski, every instance of the T-scheme, that is the formula

(**T**) x is true if and only if p,

is true; Read even says that his understanding of (**T**) is "an unquestioned orthodoxy". However, Tarski focused on the provability of T-sentences from his truth-definition, but not on their truth. In Read's reply to my criticism (Read, *The Truth-Schema* 218), he agrees with my standpoint, but he adds "But he [Tarski] clearly imposed that requirement because he thought those instances were true". I should note that Read uses the proper formulation in another place of his main paper 2010 (*The Liar and* 122). He probably thinks that both formulations are equivalent. However, they are not, unless we assume that the metatheory of truth-theory is ω-complete. Anyway, Read's statement about "an unquestioned orthodoxy" is certainly incorrect. For example, Miller and Sandu in their contributions published in *Discusiones Filosóficas* formulate (**T**) via provability, not truth.

I see no place in Tarski's writings which could justify the view that he "clearly imposed, etc." The observation that provable instances of (**T**) are true is trivial and has no relevance for Tarski's proposal of how to solve the Liar paradox. On the other hand, the provability of T-sentences matters very much and the role of this fact has the best illustration in the problem of the definability of the T-predicate. Combining the fixed-point theorem and the Tarski undefinability theorem, leads to the unprovability of some instances of (**T**). This fact concerns the Liar sentence, independently of its formulation as 'this sentence is false' or 'this sentence is not true', both recorded via arithmetization.

The situation is clear in formal arithmetic of natural numbers, but the results that hold for arithmetic cannot be directly applied to natural language. However, *pace* Tarski, excluding the Liar sentence from the stock of permitted formulas of natural language plays a quite similar role to showing that some formalized instances of (**T**) are not provable. In fact, this is an analogical move as the exclusion of division by 0 in arithmetic.

Although the formula $^{m}/_{0}$ = n is grammatically correct and perfectly understandable, it must be rejected as producing inconsistency. As far as I know Tarski never said that the liar sentence is nonsensical, meaningless, etc. He only recommended that so-called closed languages, that is, languages containing semantic concept self-referentially used, should be avoid. This restriction is by no means that counterexamples to (**T**) are excluded by *fiat* as Read suggests (127). On the contrary, they are avoided by a subtle and elegant reasoning.

That T-sentences are assumed to be provable is important for their status. Read says (*The Liar and* 125) that according to Tarski (**T**) is "a merely a material equivalence". Consequently, the instances of (**T**) have the same status. Although Tarski himself was not quite explicit about this issue, it is rather obvious that provable theorems of the form *A B* are something more than material equivalences. In particular, one cannot replace *A* or *B* by their material equivalences. For instance, the equivalence: 'the sentence 'Stephen Read wrote a paper on T-scheme' is true if and only if Stephen Read wrote a paper about T-scheme' cannot be replaced by the equivalence: 'Stephen Read wrote a paper about T-scheme if and only if Jan Woleński commented Read's paper', although both equivalences are true, and, thereby, materially equivalent, and, moreover, both consist of true constituents.

I do not suggest that Read considers such a replacement as possible or justified. Yet I claim that the difference between merely material equivalences and provable material equivalences is important, particularly for a proper interpretation of Tarski's truth-theory.

Read entirely neglects some properties of languages for which the semantic concept of truth is defined. Firstly, such a language $\mathcal{L}$ is formalized or at least it has specified structure; this latter concept was introduced in Tarski 1944. Without entering into details, $\mathcal{L}$ must be well-described as a syntactic object. In particular, we should know what belongs to the vocabulary of $\mathcal{L}$ and how the class of its sentences is defined. Secondly and more importantly, $\mathcal{L}$ is an interpreted language. Let me quote the following words of Tarski (1933):

> It remains perhaps to add that 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 meaning is attached. For such sciences the problem here discussed has no relevance, it is not even meaningful. We shall always ascribe quite concrete and, for us, intelligible meanings to the signs which occur in the language we shall to consider. The expressions which we call sentences still remain sentences after the sign which occur in them have been translated into colloquial language. The sentences which are distinguished as axioms seem to us to be materially true, and in choosing rules of inferences we are always guided by the principle that when such rules are applied to true sentences the sentences obtained by their use should also be true. (166-67)

Tarski did not explain what meaning was for him. However, it is clear that meanings of words, ordinary or artificial, for instance, the convention that 'black' expresses the property of being white, dictate semantic interpretations. Speaking more precisely and using the model-theoretic terminology, meanings generate so-called interpretation functions which ascribe denotations of individual terms and predicates in given models.

Although Tarski was skeptical as far as the matter concerns consistent semantic constructions for natural language, the difference between it and interpreted formalized (or with a specified structure) $\mathcal{L}$ is less radical than it is frequently assumed by interpreters and critics of Tarski.

Taking facts about $\mathcal{L}$ into account one can easily demonstrate that all of the counterexamples given by Read fail. This concerns the sentences:

(1) 'I am tired' is true if and only if I am tired;
(2) 'That book was stolen' is true if and only if that book was stolen;
(3) 'Any man is mortal' is true if and only if any man is mortal.

The argument is that the words 'I', 'that' and 'any' can have different meanings on the left and right sides of the respective equivalences. However, this cannot happen by definition in the case of properly interpreted languages. In particular, such languages do not require a special principle of uniformity (Dutilh Novaes) blocking ambiguities and other defects of expressions, because they are automatically excluded by technology of the semantic interpretation.

The valuation is a function and 'I', 'that' and 'any' have to have the same interpretations in all their occurrences in (1)-(3). This is particularly clear for indexicals, like 'I' and 'that'. An ostensive specification of references (for example, supplementing the act of using 'I' or 'that' by a gesture) fixes the ascribed objects.

On the other hand, if the references of such words are not made precise, they function as variables and the problem of truth (falsehood) of (1)-(3) does not arise at all. I do not argue that ambiguities do not occur in ordinary language, but only point out that the semantic definition of truth assumes that $\mathcal{L}$ for which it works is formalized (or has a specified structure) and its interpretation is fixed, even when $\mathcal{L}$ is selected as a suitable (not closed!) part of colloquial parlance.

Read proposes to replace (**T**) by

$$(\mathbf{A})\ \mathbf{T}x \Leftrightarrow \forall p(x : p \to p).$$

This scheme codes an intuition expressed by

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

Since what a sentence says is covered by all its implications, (**A**) formalizes (**S**), provided that the symbol $\Leftrightarrow$ denotes strict equivalence.

Read says (*The Liar and* 125) that "(**A**) is a logical equivalence" (expressed by the symbol $\Leftrightarrow$) contrary to (**T**) as a material equivalence.

However, Read's qualification of (**A**) as a logical equivalence is vague. If we say that A is a logical sentence, we can mean various things. Firstly, (a) A is logical if and only if it is coded by signs belonging to the language of logic, but, secondly, (b) A is logical if and only if it is a logical theorem.

Now, (**T**) and (**A**), are logical equivalences in the same sense, if the first understanding is assumed, but they both are not logical theorems.

Thus, neither (a) nor (b) qualify (**A**) as a logical equivalence, contrary to (**T**). Perhaps Read intends to say that (**A**) is a logical equivalence, because strict connectives are logically stronger than material ones. However, due to my previous remarks about the status of the instances of (**T**), provable equivalences are "something more" than material equivalences. In fact, the difference between the formulas $X \vdash A \leftrightarrow B$ and $A \Leftrightarrow B$ seems secondary, although, at least in my view, the former is much clearer than the latter.

Read points out that (**A**) is intensional for 'says that', but he does not see any problem with it. In particular, he seems to think that everything is solved by the closure of $x : p$ by "allowing substitution only of logical equivalents" (Read, *The Liar and* 124).

Unfortunately, the matter is not so simple, because 'says that' is strongly intensional. Consider two equivalent formulations of a mathematical axiom, let say, the parallel postulate. Denote them by F and F', respectively. They are logically equivalent. Assume that we have a person O who does not know that F and F' are provably equivalent. Thus, the formulas $x : F$ and $x : F'$ do not say the same for O, although Read claims that they do.

Otherwise speaking, 'says that' forms contexts which are, contrary to Read, not closed under replacement by logical equivalents. Defining 'says that' as satisfying such a kind of closure is, in my opinion, at odds with the ordinary meaning of this operator.

There are other implausible consequences of (**A**). Consider (I do not assume that Peano arithmetic is first-order)

(4) $\mathbf{T}$('Peano axioms') $\Leftrightarrow \forall p$ ('Peano axioms': $p \rightarrow p$).

Peano axioms are true in standard and non-standard model of arithmetic. Let **N** be the standard model and **N′** a nonstandard one. Take the sentence (*) 'all natural numbers have finitely many predecessors', which is true in **N**, but false in **N′**. Intuitively speaking, Peano axioms say (*) in **N**, but its negation in **N′**.

Thus, 'says that' requires a relativisation to a model, but (**T**) without it has false instances. This also shows that defining the context $x : p$ by the consequences of p may be insufficient in some cases, because a reference to models is required.

There is also a problem with (**A**) as applied to falsehoods (this is also pointed out by Miller and Sandu). The definition of **F** ('is false') corresponding to **(A′)** can be recorded as

$$(\mathbf{A''})\ \mathbf{F}x \Leftrightarrow \exists p\ ((x : p) \wedge \neg p).$$

As an example we have

(5) $\mathbf{F}$ ('Warsaw is the capital of France') $\Leftrightarrow p$ ('Warsaw is the capital of France': $p) \wedge \neg p$)).

However, (5) has its instantiation in

(6) $\mathbf{F}$ ('Warsaw is the capital of France') $\Leftrightarrow$ (('Warsaw is the capital of France'): the greatest Polish city is the capital of France) $\wedge$ the greatest Polish city not is the capital of France.

Now assume that someone knows that a person *O* knows that the sentence 'the greatest Polish city is the capital of France' is false and that the right side of (6) is a correct (true) instantiation of the right side of (5).

Hence, he knows that the sentence 'Warsaw is the capital of France' is also false; we use here the principle 'if an instantiation of *A* is false, *A* is false too. However, to justify that, one must assume that 'Warsaw' and 'the greatest Polish city' are co-denotative (I neglect that the latter is a description). This consideration shows that we do not need to worry whether the sentences 'Warsaw is the capital of France' and 'the greatest Polish city is the capital of France' say the same or, eventually, in which circumstances they cover the same content, because it is sufficient to known the values of nominal expressions.

Thus, even if we agree that that logically equivalent sentences say the same thing, this observation does not close the issue, because it can happen, as in the case of (5) and (6) that sentences are equivalent in theories or some language systems modulo denotative conventions, although they are not equivalent on purely logical grounds. Thus the interpretation of a language has a crucial importance for establishing what sentences say and when they are true or false.

Further, the negation of $(x : p)$ can be interpreted either as $\neg(x : p)$ or as $(x : \neg p)$ (the latter is stronger than the former). It matters in the case of negative sentences because

(7) **T** ('Warsaw is not the capital of France') $\Leftrightarrow \forall p\, (\neg p \rightarrow$ 'Warsaw is not the capital of Poland': $\neg p)$ looks more plausible than

(8) **T** ('Warsaw is not the capital of Poland') $\Leftrightarrow \forall p\, (\neg p \rightarrow$ (it is not the case that 'Warsaw is not the capital of Poland': $\neg p$)).

Perhaps the most important critical observation concerns metalogical properties of **T**. The law of the excluded middle can be stated as

(9) $$\forall p\, (x : p \rightarrow p) \vee \exists p((x : p) \wedge \neg p),$$

which is an instance of

(10) $$\mathbf{T}A \vee \neg\mathbf{T}A.$$

However, the relation of $\mathbf{T}A$ and $\mathbf{T}\neg A$ is not clear under Read's definition. In general we have

(11) $$\mathbf{T}\neg A \rightarrow \neg\mathbf{T}A,$$

and in the classical bivalent (two-valued) case the equivalence

(12) $$\mathbf{T}\neg A \leftrightarrow \neg\mathbf{T}A \leftrightarrow \mathbf{F}A$$

holds. Now $\mathbf{T}\neg A$ has its interpretation (according to Read's definition) in

(13) $$\forall p\ (x : \neg p \rightarrow \neg p).$$

However, (13) does not implies

(14) $$\exists p\ ((x : p) \wedge \neg p),$$

although we have (or should have)

(15) $$(x : \neg p) \rightarrow \neg(\ x : p).$$

I do not claim that (12) is indispensable, but only note that important relations are unclear under Read's proposals. In particular, the functor of negation commutes with truth (in its semantic understanding), but not with 'says that'.

Returning to 'x says that p', a full analysis of this phrase seems to be much more complicated than Reads maintains. The general problem is that sentences can say quite different things for different persons or even for the same person depending of various pragmatic factors. For instance, we should distinguish direct (explicit) and indirect (implicit saying that. Every sentence entails infinitely many logical consequences. Even if we assume that the direct linguistic content of a sentence A is definable (it is a very optimistic presumption), the implicit content is much vaguer and the amount of its grasping, always partial, cannot be accounted in advanced. Reads seems to consider the intensionality of (**A**) as its advantage, but it is a very dubious view.

I am inclined to think that the intensionality of (**A**) prevents a satisfactory definition of saying wholly as things are. This is very important, because the real virtue of the scheme introduced by Read essentially depends on such a definition.

(**T**) essentially differ from (**A**), because the former it is purely extensional. In spite of this difference, I will argue that both schemes record almost the same intuitions. It is interesting that Tarski's starting point was similar to that proposed by Read. The initial intuition of the semantic definition of truth was presented by the following formulat of Tarski (1933) (155); Tarski followed Tadeusz Kotarbiński:

(*) A true sentence is one which says that the states of affairs are so and so, and the state of affairs indeed is so and so.

The original Polish formulation is much closer to (**A**), because it runs (in English translation): "a true sentence is a sentence which expresses that things are so and so and things indeed are so and so". Tarski explicitly formalized (*) by **T**, presumably also for eliminating 'expresses' as an intensional factor. That $\mathcal{L}$ is interpreted and that its expressions have "intelligible meanings" illuminate the issue at stake.

(**T**) can be supplemented by assertions of the type 'x says that p', 'x means that p', 'x asserts that p', etc. All are external with respect to instance of (**T**) and serve to fix, define, explain, etc. the valuation function connecting $\mathcal{L}$ -expressions with their denotations. However, 'x **:** p', although internally embedded into (**A**) plays exactly the same role, provided that semantics a la Tarski is associated with this scheme.

However, Read explicitly says (*The Liar and* 125) that this is just the case. In my opinion, (**T**) has similar intuitive advantages as (**A**), including the correspondence platitude, that is, the ability to express the basic content of *adequatio rei et intellectus.*

On the other hand, the Tarski scheme avoids all problems caused by the internal intensionality of (**A**), in particular, the treatment of 'is false' (remember Russell's requirement that any satisfactory theory of truth should be also the theory of falsehood). In order to explain this point let me refer to Miller 2010 (223). He attributes to Tarski the following scheme (I use the notation of the present paper):

$$(\mathbf{T'})\ \mathbf{T}x \leftrightarrow (x \in \mathcal{L}) \wedge p.$$

The first conjunct in the right side of (**T'**) indicates that the sentence named by x belongs to $\mathcal{L}$.

Miller define 'is false' by

$$(\mathbf{F'})\ \mathbf{F}x \leftrightarrow (x \in \mathcal{L}) \wedge \neg p.$$

Since **T**x and **F**x are mutual negations in classical logic, the same relation must hold between the right sides of (**T'**) and (**F'**).

However, $(x \in \mathscr{L}) \wedge p$ and $(x \in \mathscr{L}) \wedge \neg p$ do not negate each other. The negation of the former formula yields the sentence $\neg(x \in \mathscr{L}) \vee \neg p$. Assume that this disjunction is true in favor of $\neg(x \in \mathscr{L})$. Now, since x does not belong to $\mathscr{L}$, it is not a sentence.

Consequently, it is not true and not false, because only sentences of $\mathscr{L}$ can be true or false. Clearly, $\mathbf{F}x$ and $\neg\mathbf{T}x$ are not equivalent under (**T'**) and **(F')**. This example additionally shows difficulties when truth is defined by conjunctions of conditions. It much better to treat clauses, like $x \in \mathscr{L}$ as external with respect to (**T**).

Read claims that (**A**) is better, because it correctly solves the Liar paradox. According to Read, (**T**) is false, because it has a false instance (the letter l refers to the Liar sentence):

(#) l is true iff and only if l is not true.

On the other hand, (**A**) produces a correct truth-condition of l in the form:

(##) $\mathbf{T}l \Leftrightarrow \neg\mathbf{T}l \wedge \mathbf{T}l$.

However, this success is entirely apparent. The formula (#) does not give a complete analysis of the Liar. First of all, (#), as inconsistent, is not provable in the semantic theory of truth. Since (##) is true for Read, the formula $\mathbf{T}l$ is either false or inconsistent. However, both cases must be excluded. A simple transformation of (##) gives:

(####) $\neg\mathbf{T}l \Leftrightarrow \mathbf{T}l \vee \neg\mathbf{T}l$,

which is not satisfactory, because it says that a tautology is not true, that is, inconsistent. In fact, adding $\neg\mathbf{T}l$ (this is an unprovable formula!) to the stock of theorems of propositional logic immediately abolishes the Post (absolute) consistency of this logic. The actual situation of l is displayed by the formula

(####) $\mathbf{T}l \leftrightarrow \mathbf{F}l \leftrightarrow \neg\mathbf{T}l \wedge \mathbf{T}l$.

Roughly speaking, every semantic assumption about l, that is, concerning its truth-value modulo the bivalence, entails a contradiction. It clearly shows, firstly, why some instances of (**T**) are excluded as unprovable and, secondly, that (##) does not formulate any truth-condition. This concurs with Sandu's remark that (Sandu 289), it is not clear what the Liar sentence says, although I am inclined to a more radical conclusion, namely that *l* says nothing.

To sum up, if (**A**) is reduced to (**T**), the former solves the Liar in the same way as the latter does, but if is not reduced, the issue is still open, because the proponent of (**A**) must decide whether this schema has true, but unprovable instances.

Finally, I would like to note that the problem of a philosophical significance of the **T**-scheme is still open. Does it code the correspondence intuition or not? Should we modify (**T**) in order to express the correspondence platitude? I do not address to these and similar questions in this paper. I hope to give an account of the philosophical significance, if any, of Tarski's theory of truth in Woleński (in preparation).

Bibliographical references

Dutillh Novaes, Catarina. "Tarski's hidden theory of meaning: Sentences say exactly". *Unity, truth and the Liar*. *The modern relevance of medieval solutions to the Liar Paradox*. Shahid Rahman, Tero Tulenheimo, Emmanuel Genot. Dordrecht: Springer, 2008. 41-64. Print.

Miller, David. "Verdades, antiguas y modernas". *Discusiones Filosóficas*. Ene.-Jun. 2011: 207-227. Impreso.

Rahman, Shahid., Tero Tulenheimo and Emmanuel Genot. Eds. *Unity, truth and the Liar*. *The modern relevance of medieval solutions to the Liar Paradox*. Dordrecht: Springer, 2008. Print.

Read, Stephen. "The Truth-Schema and the Liar". *Unity, truth and the Liar*. *The modern relevance of medieval solutions to the Liar Paradox*. Shahid Rahman, Tero Tulenheimo, Emmanuel Genot. Dordrecht: Springer, 2008. 3-18. Print.

---. "Further thoughts on Tarski's T-scheme and the Liar". *Unity, truth and the Liar*. *The modern relevance of medieval solutions to the Liar Paradox*. Shahid Rahman, Tero Tulenheimo, Emmanuel Genot. Dordrecht: Springer, 2008. 205-228. Print.

---. "The Liar and the new truth-Schema". *Discusiones Filosóficas.* Jul.-Dic. 2010: 119-137. Print.

---. “Miller, Bradwardine and truth”. *Discusiones Filosóficas.* Ene.-Jul. 2011: 229-235. Print.

Sandu, Gabriel. “Read on the Liar”. *Discusiones Filosóficas.* Jul.-Dic. 2011: 277-290. Print.

Tarski, Alfred. “The concept of truth in formalized languages”. *Logic, semantics, Metamathematics*. Oxford: Oxford University Press, 1933. Print.

---. “The semantic concept of truth and the foundations of semantics”. *Philosophy and Phenomenological Research.* 1944: 341-375. Print.

Woleński, Jan. “Read about T-scheme”. *Unity, truth and the Liar*. *The modern relevance of medieval solutions to the Liar Paradox*. Shahid Rahman, Tero Tulenheimo, Emmanuel Genot. Dordrecht: Springer, 2008. 199-203. Print.

---. *Semantics and truth.* Forthcoming.