

## TRUTHS ANCIENT AND MODERN<sup>1</sup>

DAVID MILLER

*Department of Philosophy  
University of Warwick  
Coventry CV4 7AL  
UK*

*dwmiller57@yahoo.com*

**Abstract:** The paper presents a comparison of the theories of truth, and the solutions of the liar paradox, proposed by Thomas Bradwardine (c. 1290–1349), Jean Buridan (c. 1295–1358), and Alfred Tarski (1901–1983).

**Keywords:** Self-reference. Bradwardine. Buridan. Tarski. Liar paradox. Method of diagonalization. Theory of signification. Theories of truth.

### 1. INTRODUCTION

The subject of this paper is truth, and its opposite, falsity, especially when they live together. That is to say, I wish to turn my attention once again to the ancient paradox of the liar, one of the most venerable problems of logic. My pretext for returning to a subject so familiar and so deeply explored by others is the recent revival by my

---

<sup>1</sup>This is an augmented and corrected version of an invited lecture given at the meeting SCIENCE, TRUTH AND CONSISTENCY, held at the State University of Campinas (UNICAMP) from 25–28 August 2009 in celebration of the 80th birthday of Newton da Costa. Versions of the lecture have been given also at Kyoto University in July 2009, at the XVII FORO NACIONAL DE FILOSOFÍA de Colombia, held at the University of Caldas (sede Palogrande) from 15–18 September 2009, at the University of Valparaíso in October 2009, and at the University of Edinburgh in March 2010. I am indebted to Stephen Read for several valuable comments and corrections. The Spanish version of the lecture is expected to appear in due course in the journal *Discusiones Filosóficas* (Manizales), together with a brief response from Read.

compatriot, the philosopher Stephen Read, of a resolution of this paradox, and others, that was proposed in the 14th century by Thomas Bradwardine, a little known Englishman today, but a man of much intellectual distinction. Read (2008a, 2008b, 2010) not only presents Bradwardine's theory of truth sympathetically in modern dress. He is also an enthusiastic proponent of it, maintaining that it is in various respects superior to the predominant theory today; that is, the theory due to Alfred Tarski. I am by no means in agreement with this judgement. I think, nevertheless, that Bradwardine's contribution deserves our attention, and that we can learn much from an investigation, even a brief one, into this new-old solution.

## 2. THE LIAR PARADOX

In speaking of the liar paradox, I may take for granted the general idea, that is, of a sentence, to all appearances well constructed and intelligible, that is true if & only if it is false (or not true). The paradox has assumed many different forms in the course of its history, from the most everyday to the most mathematical. There are simple statements such as 'I am lying', in which there is an implicit reference to the time at which the statement is made. There are additionally the sentences 'This sentence states something false' and 'This is a false sentence', which use demonstrative adjectives or pronouns to indicate what they are talking about. I shall follow here the common custom of calling such sentences *self-referential*, by which I mean that it is the grammatical subject of the sentence, though not the sentence itself, that refers to the whole sentence.

An important version of the liar paradox uses only indirect self-reference, and shows that direct self-reference does not constitute the heart of the problem. If, for example, Daphnis says that Chloe is speaking falsely, and Chloe says that Daphnis is speaking truly, then

there is no consistent attribution of truth values to these utterances. If Daphnis' utterance is true, then Chloe's is false, in which case Daphnis' utterance is also false; and hence Chloe's utterance is true, and so Daphnis' utterance is true after all. We have travelled in a complete circle without any stable attribution of truth or falsity to either of the two utterances.

There are versions of the paradox that depend for their effectiveness on an empirical premise that does not appear explicitly. For example, there is the paradox of Epimenides the Cretan, who said that all Cretans are liars; that is, that all Cretans persistently and unfailingly utter falsehoods. This utterance of Epimenides cannot be true, which implies that not all Cretans unfailingly utter falsehoods. That is to say, there is some Cretan, Epimenides or someone else, who on one occasion at least utters something true. It is astonishing that Epimenides' mere utterance of what he uttered can have such an outcome.

The inverse of this paradox is Curry's paradox (Goldstein (1986)), a sentence such as 'If this sentence is true then God exists'. If this conditional sentence is false, its antecedent is true, and hence the conditional is not false but true, and, what is more, has a true antecedent. Using the rule of *modus ponens* (or the truth table for the conditional) we may conclude that the consequent is true also; that is, that God exists. In such a way we may demonstrate *ab initio* any sentence that we wish to demonstrate.

I should mention also the so-called paradox of the *truth-teller*, a sentence that attributes truth, and nothing more, to itself. The problem here is that there seems to be no consideration whatever that could help us to assign a truth value to this sentence.

In this paper, in order to avoid problems associated with the use of demonstrative pronouns and adjectives, indexical expressions, and other idiomatic locutions that introduce into the discussion elements that are far from formal and at times undeniably empirical, I shall

mean by the liar paradox some version generated by the method of diagonalization invented by Gödel in 1931.

What follows is an adaptation of a construction by Tarski that is hidden in note 11 of his popular article (1944), and rarely mentioned. If  $y$  is a sentence of the form ‘ $b$  has  $B$ ’, let the *transform*  $y^*$  of  $y$  be the sentence ‘“ $b$  has  $B$ ” has  $B$ ’. For example,  $y$  may be the true sentence ‘The first sentence of the first published English translation of Wittgenstein’s *Tractatus* has exactly eight words’, in which case  $y^*$  is the false sentence ‘“The first sentence of the first published English translation of Wittgenstein’s *Tractatus* has exactly eight words” has exactly eight words’. Let  $z$  be the sentence ‘ $b$  has a false transform’. Then  $z^*$ , the transform of  $z$ , is ‘“ $b$  has a false transform” has a false transform’. It is intuitively evident that  $z^*$  is true if & only if the transform of  $z$  is false, that is, if & only if  $z^*$  is false. We may construct likewise an example of the truth-teller paradox. Let  $x$  be the sentence ‘ $b$  has a true transform’. Its transform  $x^*$ , that is, ‘“ $b$  has a true transform” has a true transform’ is equivalent to the sentence ‘ $x^*$  is true’.

Henceforth  $z^*$  (or any other similar sentences constructed by the method of diagonalization) will be given the name  $U$ , and called a *liar*. A sentence like  $x^*$  will be called a *truth teller*. Note, however, that the grammatical subject of the sentence  $U$ , that is, the clausal name ‘“ $b$  has a false transform”’ does not refer to  $U$  (but to  $z$ ).  $U$  is not a self-referential sentence in the usual sense, though it says something about itself.

We must note also a point of great importance: that if we assume the classical truth table for the operation of negation, that is, that a sentence  $y$  is true if & only if its negation  $\neg y$  is false, then the negation  $\neg U$  of  $U$  is equally a liar, since  $\neg U$  is true if & only if  $U$  is false, that is, if & only if  $U$  is true, that is, if & only if  $\neg U$  is false. This sentence  $\neg U$ , that is, ‘“ $b$  has a false transform” has a true transform’, may seem to be even less self-referential than is  $U$ , since we cannot

identify  $\neg U$  with any expression mentioned in  $\neg U$ . Each of  $U$  and  $\neg U$  says something about  $U$ , but only one of them actually is  $U$ . But this asymmetry is an illusion. To see this, let the *countertransform*  $y^\dagger$  of a sentence  $y$  of the form ‘ $b$  has  $B$ ’ be the sentence ‘the sentence ‘“ $b$  has  $B$ ” does not have  $B$ ’. Let  $z$  be the sentence ‘ $b$  has a false transform’, as before, and let  $z^\dagger$ , the countertransform of  $z$ , be called  $W$ . Then  $W$  is interdeducible with  $\neg U$ , and  $\neg W$  is interdeducible with  $U$ . Each of  $W$  and  $\neg W$  says something about  $W$ , but only one of them actually is  $W$ . In short,  $\neg U$ , which is equivalent to  $W$ , is just as self-referential as  $U$  is. We shall sometimes say that  $U$ , and other liars such as  $\neg U$ ,  $W$ , and  $\neg W$  exhibit the phenomenon of *reflexivity*.

An old solution to the liar paradox, formulated a century before Bradwardine, is that of *cassationism* (Goldstein (2008)), which judges that reflexive locutions are often, or always, senseless, saying nothing, and should be excluded from language. A superficially similar approach is *restrictionism*, which uses less drastic means to ban self-reference in specified cases (Panaccio (2008)). That cassationism, at least, goes too far is obvious when we recall that, in Great Britain for example, many paperback books impose the following condition of sale:

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out or otherwise circulated without the publisher’s prior consent in any form of binding or cover other than that in which it is published and without a similar condition *including this condition* being imposed on the subsequent purchaser.

In the absence of the emphasized self-referential phrase, Daphnis could legally re-sell the original intact book to Chloe, requiring of her only that she does not re-bind it. Chloe could at once re-sell it to Daphnis unconditionally. He could then re-bind it, and sell it back to Chloe. The intended condition of sale would have been sidestepped. A related suggestion is that all those supposed sentences that directly or

indirectly state their own truth values are senseless (or, at least, have no truth value), and must be discarded. This proposal too excludes too much. By diagonalization we may construct, for example, two sentences  $J$  and  $K$  that are respectively interdeducible with ‘ $J$  and  $K$  have the same truth value’ and ‘ $J$  and  $K$  have different truth values’. Since  $J$  and  $K$  are mutually contradictory, they have different truth values; and therefore  $J$  is false and  $K$  is true. The truth values can be determined definitively without inconsistency. For my part, I am reluctant to reject as senseless such sentences and other reflexive expressions.

### 3. THE TRUTH SCHEME

We are accordingly under pressure to identify a false premise in the derivation of each contradiction. It seems that it was Stanisław Leśniewski, Tarski’s teacher, who was the first to point out that each derivation assumes implicitly at least some biconditionals of the form ‘“Snow is white” is a true sentence if & only if snow is white’ (see Tarski (1944, note 7)). The majority of these biconditionals are uncontroversial (Quine (1951) could have given them as examples of supposedly analytic statements, more interesting than ‘All bachelors are unmarried’), but some of them are contradictory. Given which sentence  $U$  is, one example is the biconditional ‘ $U$  is true if & only if “ $b$  has a false transform” has a false transform’.

These biconditionals can all be combined into the infinite scheme

$$\mathbb{T}(X) \leftrightarrow p, \tag{T}$$

where the letter ‘ $X$ ’ is replaced by a name of some sentence, and ‘ $p$ ’ is replaced by the sentence whose name replaces ‘ $X$ ’. More precisely, it is necessary, in order that the scheme (T) contain exclusively instances that are valid (or invalid) solely on logical and linguistic grounds, to restrict the possible substituends for ‘ $X$ ’ to structural-descriptive names

of sentences. I mean that we should consider only those names of sentences that reveal how the sentence is constructed from its syntactic components. For our purposes, the most important aspect of a structural-descriptive name  $a$  of a sentence is that we can decide in the background theory (usually a fragment of set theory) the formula  $a \in L$ , that is, whether or not the sentence whose structural-descriptive name is  $a$  belongs to the language whose structural-descriptive name is  $L$ . In this paper I shall, for simplicity, try to downplay the distinction between structural-descriptive names and others, and not emphasize it pedantically.

A quotation name of a sentence, such as ‘“Snow is white”’, which names the sentence ‘Snow is white’, is a structural-descriptive name. Observe, however, that the scheme (T) cannot be written in the quantified form  $\forall p[\mathbb{T}('p') \leftrightarrow p]$ . The dangers of substitution within quotation marks are well known. Those who prefer to attribute truth and falsity to propositions usually resort to something rather similar; that is, they combine the biconditionals of (T) in the universal statement ‘For each proposition  $p$ , it is true that  $p$  if & only if  $p$ ’. Since the variable ‘ $p$ ’ is not used uniformly here, it is doubtful whether this statement is well formed. We shall, however, suppose that propositional quantification makes good sense.

In whatever way the truth conditions of meaningful sentences are stated, the scheme (T) needs some adjustment, despite its expressing what Wright (1992, p. 27) calls ‘the correspondence platitude’. I shall here consider three approaches to putting (T) in order: those of the medieval luminaries Thomas Bradwardine and Jean Buridan, expounded by Read (2008a), and that of Alfred Tarski. The third approach is well known, and the second is not unknown. Although his presentation (2008a) is at times somewhat opaque, in presenting Bradwardine’s theory I shall rely heavily on Read’s exposition, which itself makes use of, and modifies, the pioneering exposition of Spade (1981). But

I shall try to eliminate various infelicities and unclarities from it, and to present it in the most effective way. Each approach has the great merit of working entirely within the confines of classical logic; this is, at least, the intention of each approach, though in reality, as we shall see, none of the three yields a theory of truth that is truly classical.

#### 4. THE THEORIES OF THOMAS BRADWARDINE (c.1290–1349)

Thomas Bradwardine, appointed Archbishop of Canterbury in 1349, only to be struck down by the plague a few weeks later, was involved in many spheres of learning, including physics (Grant (1965)), mathematics (Boyer (1949, Chapter III)), probability theory (Bellhouse (2000, §3)), theology (Leff (1957)), and logic (Read (2010)). Chaucer mentions him by name in *The Nun's Priest's Tale*.

His solution to the liar paradox really consists of two distinct theories, a rudimentary doctrine of signification or of *saying that*, presented axiomatically, which associates each sentence with those propositions that it signifies (in a sense that fulfils the postulates), and a definition, in those terms, of truth.

Apart from his use of the category of propositions (as well as sentences), which some people will consider extravagant, this doctrine of signification may seem by no means extraordinary. According to Read (2008a), it contains at least three postulates. There is first a postulate of *explicitness*, which lays down that each sentence signifies what it says explicitly. Using the colon ‘:’ to abbreviate indifferently the verb ‘signify’ in this specialized sense, and all its inflexions, we may express the postulate formally by the scheme

$$X : p, \tag{E}$$



where (as in the restricted scheme (T) above) ‘ $X$ ’ is replaced by a structural-descriptive name of some sentence, and ‘ $p$ ’ is replaced by the sentence whose name replaces ‘ $X$ ’. The restriction to structural-descriptive names is needed only in order to maintain the analytic or linguistic status of (E). For of course the first sentence of the first published English translation of Wittgenstein’s *Tractatus* signifies that the world is everything that is the case as much as the sentence ‘The world is everything that is the case’ signifies it, but the former is a matter of historical fact, rather than an analytic or linguistic truth, and ought not to be included in a semantical postulate such as the scheme (E).

It is to be noted that, like instances of the scheme (T), instances of the scheme (E) do not profess a relation between a sentence and a proposition but something more abstract. The symbol ‘ $\bullet$ ’ may be thought of as a parametrized propositional operator; ‘ $X \bullet p$ ’ is on a par with the expression ‘ $\mathcal{K}_A p$ ’, which is sometimes used in epistemic logic to mean ‘the agent A knows that  $p$ ’.

There are two further postulates that govern the pseudo-relation  $\bullet$  of signification; in these postulates, of *closure* (K) under the relation of logical implication, and of *adjunction* (C), the variables ‘ $x$ ’, ‘ $p$ ’, and ‘ $r$ ’ have a more general function:

$$\forall x \forall p \forall r ((x \bullet p \ \& \ p \Rightarrow r) \rightarrow x \bullet r), \quad (\text{K})$$

$$\forall x \forall p \forall r ((x \bullet p \ \& \ x \bullet r) \rightarrow x \bullet p \wedge r). \quad (\text{C})$$

The variable ‘ $x$ ’ here ranges over sentences, while the variable ‘ $p$ ’ ranges over propositions. The symbol ‘ $\Rightarrow$ ’ represents the relation of logical implication between propositions. Read scarcely notices the scheme (E), and Spade *ibidem*, note 26, combines it with the closure scheme (K). In §8 below we shall do the same, producing a stronger postulate (E+), but we shall want to extend (K) also in another direction. For the mo-

ment it is preferable to keep (E) and (K) apart. However Bradwardine's theory of signification is presented, it certainly requires some assumption that prevents the pseudo-relation  $\bullet$  from being void. This is the purpose of (E), which, despite its triteness, is not as uncontroversial as are (K) and (C). The latter postulate plays little part in what follows.

Taken together, the postulates (E), (K), and (C) require that the set of propositions signified by a sentence  $x$  constitutes a closed system, or a deductive system in the sense of Tarski (1935–1936). It is convenient also to write  $x \bullet\bullet p$  to mean that  $x$  signifies  $p$  and nothing more than  $p$  and its logical consequences (as the closure postulate (K) requires). So far, so unexciting.

Independent of these postulates, but making use of the pseudo-relation  $\bullet$  of signification, are the definition of truth proposed by Bradwardine, and his analysis of some self-referential sentences. His proposal was that a sentence  $x$  is true if & only if (in Read's words) 'things are wholly as  $x$  says they are' (2008a, p. 6); more formally, the quantified statement

$$\forall x(\mathbb{T}(x) \leftrightarrow \forall p(x \bullet p \rightarrow p)) \quad (\text{A})$$

replaces the scheme (T). As in (K) and (C), the variable ' $x$ ' here ranges over sentences, and the variable ' $p$ ' ranges over propositions. (A) may be read in this way: a sentence  $x$  is true if & only if for each proposition  $p$ , if  $x$  says that  $p$ , then  $p$ . The double use of the propositional variable ' $p$ ' ought to arouse some anxiety, but I shall try here to set that aside.

It should be noted that, in Read's presentation, (A) is a statement of logical equivalence, using the double-shafted (and double-headed) arrow ' $\leftrightarrow$ ' in place of the single-shafted arrow ' $\leftrightarrow$ '. But since not even the material biconditional (A) is able to deflect the liar paradox, as we shall see, nothing is gained (and something may be lost) by replacing it with a strict biconditional.

The definition (A), according to Read, constitutes an improvement and an extension of the scheme (T). To put the matter rather loosely, a sentence may signify some proposition that it does not imply, but there is in general no presumption that it does so; and if it does not do so we may simplify the right side of (A), replacing it by ‘ $p$ ’. For any such sentence  $x$  named by a structural-descriptive name, the definition (A) and the scheme (T) then say exactly the same thing. On the supposed superiority of (A) to (T), see the discussions in Miller (2010, p. 434), and in §5 below.

Something unexpected happens, however, when we substitute for ‘ $x$ ’ in (A) a structural-descriptive name ‘ $L$ ’ of a sentence  $L$  that signifies its own falsehood, a sentence  $L$  for which  $L \bullet \neg \mathbb{T}(L)$ . Let us suppose that everything extra that  $L$  signifies can be bundled up into the proposition  $q$ . In this case, we may write  $L \bullet \bullet \neg \mathbb{T}(L) \wedge q$ . Read understands the universal quantifier in the definition (A) in such a way that is possible to deduce from it the equivalence

$$\mathbb{T}(L) \leftrightarrow \neg \mathbb{T}(L) \wedge q, \quad (\text{AL})$$

from which it follows both that  $\neg \mathbb{T}(L)$  and that  $\neg q$ . In other words, the sentence  $L$  may be demonstrated to be false (not true). It is important to observe that it is not possible to demonstrate  $\mathbb{T}(L)$ , even though it is possible to demonstrate  $L$ . In order to establish this impossibility, let the pseudo-relation  $x \bullet p$  hold for every  $x$  and  $p$ , and  $\mathbb{T}(x)$  fail for every  $x$ .

Of the two propositions that  $L$  was explicitly assumed to signify, one, namely  $\neg \mathbb{T}(L)$ , turns out to be demonstrable, given (A), and the other, namely  $q$ , to be refutable. Assuming postulate (K), the assumption that signification is closed under logical implication, Bradwardine went on to show that  $L$  signifies also  $\mathbb{T}(L)$ . This result is given prominence by Read, and described by Restall (2008), p.142, as ‘Brad-

wardine's most interesting contribution to the discussion'. Yet because the proposition  $\mathbb{T}(L)$ , like the proposition  $q$ , is contradictory (given the definition (A)), the proposition that  $L \bullet \mathbb{T}(L)$  may be proved from the weaker assumption

$$\forall x \forall p \forall r ((x \bullet p \ \& \ p \Leftrightarrow r) \rightarrow x \bullet r), \quad (\text{KK})$$

according to which signification is closed under logical equivalence.

A word has to be said about Read's understanding of the universal quantifier in the definition (A). If we apply carefully to the case of  $L$  the postulate

$$\forall x (\mathbb{T}(x) \leftrightarrow \forall p (x \bullet p \rightarrow p)), \quad (\text{A})$$

we obtain

$$\begin{aligned} \mathbb{T}(L) \leftrightarrow & (L \bullet \neg \mathbb{T}(L) \rightarrow \neg \mathbb{T}(L)) \wedge (L \bullet q \rightarrow q) \quad (\text{AL-}) \\ & \wedge \forall p ((p \neq \neg \mathbb{T}(L) \wedge p \neq q) \rightarrow (L \bullet p \rightarrow p)), \end{aligned}$$

from which Read excises the final conjunct, on the grounds that  $L$  has been assumed to signify only the two propositions  $\neg \mathbb{T}(L)$  and  $q$  (and their logical consequences, which need not be separately considered). If (A) were a strict biconditional, this move would be open to question (Mills (2008, § 6.3)), on the grounds that the assumption  $L \bullet \bullet \neg \mathbb{T}(L) \wedge q$  is not evidently something demonstrable, but it is unexceptionable if (A) is a material biconditional. The argument now may be continued by simplifying the first conjunct of (AL-), but not the second, yielding instead of (AL),

$$\mathbb{T}(L) \leftrightarrow [\neg \mathbb{T}(L) \wedge (L \bullet q \rightarrow q)]. \quad (\text{AL=})$$

By the previous argument, we may conclude that  $\neg\mathbb{T}(L)$  and that  $\neg(L \bullet q \rightarrow q)$ ; that is, that  $\neg\mathbb{T}(L)$  and that  $L \bullet q$  and that  $\neg q$ . After all, it is possible to demonstrate that  $L$  signifies a contradiction, namely  $q$ . If (A) is a material biconditional, Bradwardine's argument that  $L \bullet \mathbb{T}(L)$  can therefore be brought safely to its conclusion.

But in any case, it is possible to derive from (AL<sup>-</sup>), strictly interpreted, the material conditional

$$\mathbb{T}(L) \rightarrow \neg\mathbb{T}(L) \wedge (L \bullet q \rightarrow q), \quad (\text{AL}\equiv)$$

from which  $\neg\mathbb{T}(L)$  is again derivable. But I see no way to the conclusions that  $L$  signifies a contradiction, and that it signifies also its own truth.

A further question, raised by Sandu (2007, p. 139), can also be put to rest (italics have been suppressed): 'why should we assume that everything the Liar says is expressible by one single proposition?' Sandu goes on to assert that since 'there are infinitely many propositions expressed by the Liar ... Read's argument ... can be properly carried out only in a (n infinitary) metalanguage'. This is not wholly correct. The argument may be repeated with an arbitrary deductive theory  $\mathbf{Q}$  taking the place of the proposition  $q$ , and if  $\mathbb{T}(L)$  is a proposition expressed by a single sentence then Bradwardine's proof shows that the theory  $\mathbf{Q}$  is contradictory, and therefore axiomatizable. But it is only in this sense that the axiomatizability of  $\mathbf{Q}$  is part of the assumption that  $\mathbb{T}(L) \wedge \mathbf{Q}$  is axiomatizable.

Be that as it may, in (A) alone we do not have a solution to the traditional paradox of the liar, since (A) does not guarantee that there exists a sentence  $L$  such that  $L \bullet \neg\mathbb{T}(L)$ . Without the postulate (E), or something like it, the above analysis is in vain. But in the presence of the postulate (E), the liar  $U$  is effectively neutralized by Bradwardine's

treatment, however the quantifiers are handled. This is, to be sure, a notable achievement. As we shall see, it is not enough.

## 5. THE APPROACH OF JEAN BURIDAN (c.1295–1358)

It is instructive to compare Bradwardine's little known approach with an approach that can be associated with his contemporary Jean Buridan. Buridan's opinions changed during his lifetime, but he seems to have held at one time that every sentence  $x$  signifies two propositions (and their logical consequences) and nothing more: what  $x$  says explicitly (this is (E)), and the proposition  $\mathbb{T}(x)$  that states the truth of  $x$ . This suggestion of Buridan's, stated more carefully, amounts to extending the scheme (E) to

$$X \bullet\bullet \mathbb{T}(X) \wedge p, \quad (\text{EB})$$

where, as in (E), the letter ' $X$ ' is replaced by a structural-descriptive name of some sentence, and ' $p$ ' is replaced by the sentence whose name replaces ' $X$ '. Combining this enriched postulate of signification with the definition (A) of truth proposed by Bradwardine, we obtain

$$\mathbb{T}(X) \leftrightarrow \mathbb{T}(X) \wedge p, \quad (\text{B})$$

which is logically equivalent to one half of the scheme (T), namely

$$\mathbb{T}(X) \rightarrow p. \quad (\text{TB})$$

Buridan counts the liar  $U$  as false (not true), since  $\mathbb{T}(U) \rightarrow \neg \mathbb{T}(U)$  by (TB), and hence  $\neg \mathbb{T}(U)$ . The approaches of Buridan and of Bradwardine are in agreement that  $\neg \mathbb{T}(U)$  and that  $U \bullet\bullet \mathbb{T}(U)$ ; that is, that  $U$  is contradictory.

Read is eager to emphasize (2008a, §1.6) that there is, despite this similarity, a great difference, between the two approaches. (See also Miller (2010, p. 436).) While Buridan assumes that  $X \bullet \mathbb{T}(X)$  for each sentence  $X$ , and that  $U \bullet \mathbb{T}(U)$  in particular, Bradwardine deduces  $U \bullet \mathbb{T}(U)$  from (A) together with the hardly controversial postulate (K) that signification is closed under logical implication. To resolve the liar paradox, he assumes nothing more than two principles concerning signification, namely (E) and (K), that are hardly revolutionary, and a definition, namely (A). There is no overt assumption that there exists any sentence that signifies more than it implies, nor that there exists no such sentence. As we have noted, not even (K) is needed, and the weaker postulate (KK) suffices.

I suppose that it is because of this lack of commitment that Bradwardine's solution seems so attractive; they may prefer other solutions, but the contributors to Rahman, Tulenheimo and Genot (2008), for example, variously describe it as 'novel' (Armour-Garb), '[of] outstanding merit' (Dutilh Novaes), 'princely' (Goldstein), 'intriguing' (Klima), 'admirable in ambition and attitude' (Mills), 'plausible' (T. Parsons), 'intriguing and fruitful' (Restall), 'a new breakthrough' (Sagüillo), and 'bold' (Serény).

A paradox consists of an unacceptable conclusion validly inferred from acceptable premises, and in order to disable the conclusion it is sufficient to weaken the premises, without upsetting their acceptability. Most known solutions, nonetheless, for example those of Buridan (EB) and of Russell, adopt also controversial new premises. Bradwardine's solution seems to avoid all such controversy, by not relying on any premise that was not already present.

But what Bradwardine's derivation reveals is not that Bradwardine assumed nothing about what  $U$  might signify in addition to the proposition that  $\neg \mathbb{T}(U)$ , but that some assumption to this effect was latent in the application of the definition (A) to the sentence  $U$ . In professional

language, the definition (A) is *creative*, an assumption, not a genuine definition, precisely because it allows us to deduce from the postulates (E) and (KK) a proposition that cannot be deduced without it. One example is the proposition that  $\exists q\exists r((U \bullet q \ \& \ U \bullet r) \ \& \ (q \Leftrightarrow \neg r)$ ; that is, that  $U$  signifies two propositions  $q$  and  $r$  (namely  $\mathbb{T}(U)$  and  $\neg\mathbb{T}(U)$ ) that are mutual contradictories. It is evident that this proposition is not deducible from (E) and (KK) together. In like manner it may be shown that the extension of Bradwardine's full theory of signification (the theory made up of the postulates (E), (K), and (C)) by his definition (A) of truth is not a conservative extension.

The hope that Bradwardine (or perhaps Buridan) has opened the gates to a logical paradise must now be further disappointed. Serious problems emerge when we compare Bradwardine's approach with that of Tarski.

## 6. THE THEORY OF ALFRED TARSKI (1901–1983)

The semantical conception of truth offered by Tarski is well known and does not need extended treatment here. Tarski undertook to provide for the sentences of any elementary language  $L$  (the *object language*) an explicit definition of a predicate  $\mathbb{T}$  from which may be deduced ' $\mathbb{T}(X) \Leftrightarrow p$ ' whenever any structural-descriptive name of a sentence of  $L$  replaces the letter ' $X$ ', and the sentence whose name replaces ' $X$ ' replaces the letter ' $p$ '. It is only for the sentences of the object language  $L$  that the predicate  $\mathbb{T}$  of truth is guaranteed to define truth. Such a definition is said to be *materially adequate* for  $L$ . The language in which the definition of the predicate  $\mathbb{T}$  is formulated is called the *metalanguage*  $ML$  (for  $L$ ). The details of the definition, and why the metalanguage  $ML$  is necessarily richer than the object language  $L$  do not concern us here. What is important is that, for two reasons, Tarski's method cannot provide a universal definition of truth: (a) to



define truth for the sentences of ML, it is necessary to employ a meta-metalanguage, and so on for ever; (b) the method of definition given by Tarski is applied differently to each elementary language, and cannot be generalized to a variable language. A list of philosophers (beginning with Black (1949, p. 104)) who have endorsed this criticism is given in note 14 of Mou (2001).

## 7. A COMPARISON OF THE THREE APPROACHES

Tarski's attitude to the scheme (T) is superficially very different from that of Bradwardine. The former appears to limit the scope of the scheme (T), while the second rejects and replaces it completely. For Tarski the modified scheme does not constitute a definition of truth, but a test of a definition's adequacy. His definition of the truth predicate  $\mathbb{T}$  for the sentences of the language  $L$ , given in terms of the relation of satisfaction, recursively defined, is utterly different from (A), the universal definition with which Bradwardine replaces (T). It is obvious indeed that Bradwardine's definition is not materially adequate. Their solutions to the liar paradox, nonetheless, and that of Buridan too, are structurally rather similar. What Tarski says is that only sentences of the object language  $L$  can be true, which means that the correct form of the (T) scheme is equivalent to

$$\mathbb{T}(X) \leftrightarrow (X \in L) \wedge p. \quad (\mathbb{T}_L)$$

To apply this scheme ( $\mathbb{T}_L$ ) to the liar  $U$ , we replace ' $X$ ' by ' $U$ ', and ' $p$ ' by ' $\neg \mathbb{T}(U)$ ', with the result

$$\mathbb{T}(U) \leftrightarrow (U \in L) \wedge \neg \mathbb{T}(U), \quad (\mathbb{T}_L U)$$

from which we may deduce both that  $U \notin L$  and that  $\neg \mathbb{T}(U)$  (that is,  $U$ ). Since ' $U$ ' is a structural-descriptive name of a sentence that

does not belong to the object language  $L$ , the sentence ' $U \notin L$ ' can be deduced in the background theory. In other words,  $\mathbb{T}(U)$ , according to  $(\mathbb{T}_L U)$ , is equivalent to the conjunction of  $\neg \mathbb{T}(U)$  with a contradiction. This is precisely the situation when we apply Bradwardine's definition (A) to the liar  $U$ , with the result

$$\mathbb{T}(U) \leftrightarrow (U : q \rightarrow q) \wedge \neg \mathbb{T}(U), \quad (\text{AU})$$

from which we may deduce both that  $\neg q$  and that  $\neg \mathbb{T}(U)$ . Formulas (AL=) and (AU) are analogous. The less sophisticated solution given by Buridan has a similar structure, since we obtain

$$\mathbb{T}(U) \leftrightarrow \mathbb{T}(U) \wedge \neg \mathbb{T}(U). \quad (\text{BU})$$

Once again, the additional clause, namely  $\mathbb{T}(U)$ , is contradictory.

We ought to note that, despite these similarities, Tarski's definition provides attributions of truth values in cases in which Bradwardine and Buridan are silent. Consider an indirect liar:  $G = \neg \mathbb{T}(H)$  and  $H = \mathbb{T}(G)$ . In Bradwardine's and Buridan's theories, it is deducible that  $\neg \mathbb{T}(H)$ , that is  $G$  is deducible, but the truth value of  $G$  is indeterminate. Tarski's theory, which withholds the predicate  $\mathbb{T}$  from any sentence outside the language  $L$ , implies that  $\neg \mathbb{T}(G)$  and that  $\neg \mathbb{T}(H)$ , and therefore implies both  $G$  and  $\neg H$ . A more recusant example was presented in §1 above: the two sentences  $J$  and  $K$ , which are interdeducible with ' $J$  and  $K$  have the same truth value' and ' $J$  and  $K$  have different truth values' respectively. At an intuitive level  $J$  is false and  $K$  is true. Bradwardine and Buridan say that  $\neg \mathbb{T}(J)$ , and are mute about the truth value of  $K$ . If we put  $J$  and  $K$  in the form

$$\begin{aligned} J &:: [\mathbb{T}(J) \wedge \mathbb{T}(K)] \vee [\neg \mathbb{T}(J) \wedge \neg \mathbb{T}(K)], \\ K &:: [\mathbb{T}(J) \wedge \neg \mathbb{T}(K)] \vee [\neg \mathbb{T}(J) \wedge \mathbb{T}(K)], \end{aligned}$$

we see that Tarski's theory implies the second disjunction, and contradicts the other three; and so it implies  $J$  and  $\neg K$ . That is, the poor semantical theory hands down two incorrect decisions, that  $\mathbb{T}(J)$  and that  $\neg\mathbb{T}(K)$ . I do not know what the moral of this example is.

This is the appropriate point to note that, despite their obedience to the logical law of *tertium non datur* (and all other laws of classical logic) none of the three definitions considered can deliver a central law of the classical theory of truth, namely the law  $\mathbb{T}(x) \vee \mathbb{T}(\neg x)$  (which is equivalent, given the classical truth tables, to the law of bivalence  $\mathbb{T}(x) \vee \mathbb{F}(x)$ ). This is immediate in Tarski's case, since both  $\neg\mathbb{T}(X)$  and  $\neg\mathbb{T}(\neg X)$  can be deduced if ' $X$ ' is replaced by a structural-descriptive name for a sentence that does not belong to the object language  $L$ . In §2 we noted that if the negation operation conforms to the classical truth table, then the negation  $\neg U$  of the liar  $U$  is a liar too. As we saw above, Buridan's definition (B) implies that all liars are false, and hence we can deduce from (B) that  $\neg\mathbb{T}(U)$  as well as that  $\neg\mathbb{T}(\neg U)$ , given that  $\neg U$  is a liar; one application of the rule of *reductio ad absurdum* shows that the law of bivalence cannot hold. We obtain the same result for Bradwardine's definition (A) when it is strengthened with the explicitness postulate (E). A more delicate analysis shows that Bradwardine's theory can retain the law of bivalence (in whose validity Bradwardine himself believed) if & only if it sacrifices the postulate (E). Buridan's theory has no such escape route.

These results may appear to be troublesome if we wish to develop a classical theory of truth. But there is another point of view from which to look at the matter. Since we are adopting a new theory of truth (Tarski's or Bradwardine's or Buridan's), why do we suppose that the classical deductive rules of inference satisfy the principle of transmission of truth? The opposite seems more correct, since by means of the classical rules (in fact, the intuitionistic rules suffice) we can deduce from each definition of truth the sentence  $U$ , that is that  $\neg\mathbb{T}(U)$ . For

my part, it is not obvious that even the rule of  $\wedge$ -introduction transmits Bradwardine's truth since, without appropriate instructions, there is no way to exclude the possibility that the conjunction  $x \wedge z$  signifies some proposition  $q$  that is not a consequence of any conjunction of propositions  $p$  and  $r$  where  $x \dot{\vdash} p$  and  $z \dot{\vdash} r$ . (Here it is assumed that the pseudo-relation  $\dot{\vdash}$  of signification obeys both the closure postulate (K) and the adjunction postulate (C).) I regret that it is not possible to pursue further this interesting point.

## 8. AN ALTERNATIVE TO THE EXPLICITNESS SCHEME

The presentation in §§6f. above of some aspects of the definition of truth — more precisely, the method of defining truth — proposed by Tarski is more or less standard. The predicate  $\mathbb{T}$  is defined in such a way that it applies (demonstrably) only to the sentences of the object language  $L$ ; that is, if ' $X$ ' is a structural-descriptive name for a sentence outside  $L$ , we can deduce from the definition that  $\neg\mathbb{T}(X)$ . This may be summarized in

$$\mathbb{T}(X) \leftrightarrow (X \in L) \wedge p, \quad (\mathbb{T}_L)$$

where ' $p$ ' and ' $X$ ' may be replaced respectively by any grammatical sentence and by one of its structural-descriptive names. It is possible to treat the predicate  $\mathbb{F}$  of falsity in a strictly parallel fashion, producing additionally the scheme

$$\mathbb{F}(X) \leftrightarrow (X \in L) \wedge \neg p. \quad (\mathbb{F}_L)$$

An alternative way of proceeding is to decree that  $\mathbb{T}$  and  $\mathbb{F}$  are absolute contradictories, that  $\mathbb{F}(x)$  if & only if  $\neg\mathbb{T}(x)$ , in which event there are sentences outside the language  $L$  that are false. What we do does not matter much, provided that we know what we are doing.

After all, what Tarski wanted to do was to define the predicates  $\mathbb{T}$  and  $\mathbb{F}$  adequately for the sentences of  $L$ ; all other sentences could be thought of as waste cases, for which it matters little whether the predicates  $\mathbb{T}$  and  $\mathbb{F}$  apply or not. In order to achieve this goal other styles of definition would have been equally effective, if less intuitive. Tarski could, for example, have formulated an explicit definition of *untruth* of the sentences of  $L$ , via the relation of *dissatisfaction*, and its contradictory *unfalsity*, giving

$$\mathbb{T}(X) \leftrightarrow p \leftarrow (X \in L), \quad (\mathbb{T}^L)$$

$$\mathbb{F}(X) \leftrightarrow \neg p \leftarrow (X \in L). \quad (\mathbb{F}^L)$$

Under such a definition, any sentence outside  $L$  is both true and false (in the sense that, if a structural-descriptive name for that sentence replaces ' $X$ ' and the sentence itself replaces ' $p$ ', it can be demonstrated both that  $\mathbb{T}(X)$  and that  $\mathbb{F}(X)$ ). It is evident, however, that for the sentences of  $L$ , a definition yielding  $(\mathbb{T}^L)$  is materially adequate.

This possibility suggests an alternative approach to the pseudo-relation of signification employed by Bradwardine. It is clear that the intention of his theory is to distinguish (in principle) what a sentence implies, its logical content, from what it signifies. He said that '[e]very sentence signifies or means everything which follows from it' (Read (2008b, § 13.1), Spade (*ibidem*, p. 120)), which may be written schematically as:

$$X \models Z \rightarrow X \bullet r. \quad (\text{E+})$$

Here the letters ' $X$ ' and ' $Z$ ' are replaced by structural-descriptive names of sentences, ' $r$ ' is replaced by the sentence whose name replaces ' $Z$ ', and  $\models$  represents the relation of logical implication between sentences. In short, a sentence signifies all the propositions that it implies.

It is easily shown that, in the presence of the closure postulate (K), the scheme (E+), which is very close to what Spade calls the Bradwardine principle (BP), is logically equivalent to the explicitness postulate (E). The formulation (E+) shows plainly that (E) permits, but does not require, a sentence to signify more than it logically implies. I made no attempt above to defend the scheme (E), preferring to judge it by its effects. Now is the time to be more inquisitorial and to consider an alternative.

Suppose that Bradwardine and Read, instead of being English, and usually prone to understatement, had been fishermen, and prone to exaggeration; not Angles but anglers. When speakers exaggerate, what an utterance means may be less than what it says explicitly. One mode of exaggeration is the precise declaration of something meant only approximately ('I never use a big big D'). Another mode is the unconditional announcement of something meant only conditionally (the sign 'Dogs must be carried' often to be seen adjacent to escalators on the London Underground). In the latter case, the intended meaning or signification of the uttered sentence is a proper consequence of what the sentence says explicitly. We noted above that, in the presence of the closure postulate (K), the scheme (E+) expresses the theory of signification that condones, but does not insist on, understatement. The inverse theory, which condones (but does not insist on) exaggeration, can be formulated in the scheme

$$X \bullet r \rightarrow X \models Z, \quad (\text{E-})$$

with the same substitution conventions as before. (E-) is very close to what Spade calls the converse Bradwardine principle (CBP).

These two schemes (E+) and (E-) are symmetrically placed on either side of their conjunction, that is, that a sentence signifies precisely its logical consequences:  $\forall r(X \bullet r \leftrightarrow p \Rightarrow r)$ , where 'X' and 'p' are

related in the usual way. This is the theory of signification that is implicit in the scheme (T). The symmetry is not perfect, but it extends to their interaction with the liar  $U$ . Suppose that it is only under the condition  $q$  that  $U$  says or signifies what it asserts explicitly; that is, that  $U \bullet\bullet q \rightarrow \neg\mathbb{T}(U)$ . According to Bradwardine's definition (A),  $\mathbb{T}(U) \leftrightarrow (q \rightarrow \neg\mathbb{T}(U))$ , which implies both that  $\mathbb{T}(U)$  and that  $\neg q$ . Although we can demonstrate that  $U$  is true (given (A)), we cannot demonstrate  $U$  (that is, that  $\neg\mathbb{T}(U)$ ). It is not hard to show that (A) is creative in the presence of (E-), (K), and (C), as it was in the presence of (E+), (K), and (C).

Bradwardine's definition (A) of truth is not, after all, enough to take the sting out of the traditional liar paradox. The postulates (K) and (C), although obviously acceptable constraints, are not enough either. Both the scheme (E+) and the scheme (E-) are consistent with (A), (C), and (K), but it is impossible to adopt the two schemes simultaneously without a return to the scheme (T) and inconsistency. (This is the line of reasoning that Spade *ibidem*, p. 122, 124, rather hesitatingly ascribes to Bradwardine, for which he is castigated by Read 2008b, §13.1.) If (E+) is adopted, then the liar  $U$  is a theorem, and the law of bivalence is violated. If (E-) is adopted, then the negation  $\neg U$  of  $U$  is a theorem, and we can show, by a dual argument, that the law of non-contradiction is violated. A theory of truth that behaves classically is therefore available only if one of  $U$  and  $\neg U$  satisfies (E+) and the other satisfies (E-). Read tries to avoid this conclusion by claiming that the negation  $\neg U$  of the liar sentence  $U$  is not self-referential in the way that  $U$  is, and may be accounted true even when  $U$  is accounted false (*ibidem*, §13.4). We have seen in §2 above that this move is unavailing.

Unlike the purely conventional choice between the Tarski-style definitions that provide the two schemes ( $\mathbb{T}_{\perp}$ ) and ( $\mathbb{T}^{\perp}$ ), the choice between (E+) and (E-) is a substantial choice between two competing theories of signification. For this reason, and others, the present out-

look for Bradwardine's theory of truth, and for his resolution of the liar paradox, is unsettled.

## REFERENCES

- BELLHOUSE, D.R. "De Vetula: A Medieval Manuscript Containing Probability Calculations". *International Statistical Review / Revue Internationale de Statistique* 68, p. 123-136, 2000.
- BLACK, M. *Language and Philosophy*. Ithaca NY: Cornell University Press, 1949.
- BOYER, C.B. *The History of the Calculus and Its Conceptual Development*. New York: Hafner Publishing Company, Inc. Reprinted 1959. New York: Dover Publications, Inc, 1949.
- GOLDSTEIN L. "Epimenides and Curry". *Analysis* 46, p. 117-121, 1986.
- . "Doubting Thomas: from Bradwardine back to Anon". In Rahman, Tulenheimo and Genot (2008), p. 65-85, 2008.
- GRANT, E. "Bradwardine and Galileo: Equality of Velocities in the Void". *Archive for History of Exact Sciences* 2, p. 344-364, 1965.
- LEFF, G. *Bradwardine and the Pelagians*. Cambridge & elsewhere: Cambridge University Press, 1957.
- MILLER, D.W. Review of Rahman, Tulenheimo and Genot (2008). *Philosophy* 85, 3, p. 433-436, 2010.
- MILLS, E. "Scheming and Lying". In Rahman, Tulenheimo and Genot (2008), p. 113-128, 2008.
- MOU, B. "The Enumerative Character of Tarski's Definition of Truth and its General Character in a Tarskian System". *Synthese* 126, 1/2, p. 91-122, 2001.



- PANACCIO, C. "Restrictionism: A Medieval Approach Revisited". In Rahman, Tulenheimo and Genot (2008), p. 229-253, 2008.
- QUINE, W. V. O. "Two Dogmas of Empiricism". *The Philosophical Review* 60, p. 20-43. Reprinted in W. V. O. Quine (1953), p. 20-46. *From A Logical Point of View*. Cambridge MA: Harvard University Press, 1951.
- RAHMAN, S., TULENHEIMO, T., & GENOT, E., editors. *Unity, Truth and the Liar: The Modern Relevance of Medieval Solutions to the Liar Paradox*. Heidelberg: Springer Verlag, 2008.
- READ, S. L. "The Truth Schema and the Liar". In Rahman, Tulenheimo and Genot (2008), p. 3-17, 2008a.
- . "Further Thoughts on Tarski's T-scheme and the Liar". In Rahman, Tulenheimo and Genot (2008), p. 205-225, 2008b.
- . *Thomas Bradwardine, "Insolubilia"*. Leuven: Peeters, 2010.
- RESTALL, G. "Models for Liars in Bradwardine's Theory of Truth". In Rahman, Tulenheimo and Genot (2008), p. 135-147, 2008.
- SANDU, G. "Read on the Liar". In S. Pihlström, P. Raatikainen, & M. Sintonen, editors (2007), p. 129-140. *Approaching Truth. Essays in Honour of Ilkka Niiniluoto*. London: College Publications, 2007.
- SPADE, P. V. "*Insolubilia* and Bradwardine's Theory of Signification". *Medioevo: Revista di storia delle filosofia medievale* 7, p. 115-134. Reprinted as Chapter IV of P. V. Spade (1988). *Lies, Language and Logic in the Late Middle Ages*. London: Variorum Reprints, 1981.
- TARSKI, A. "Grundzüge des Systemenkalkül". *Fundamenta Mathematicae* 25, 4, p. 503-526, and 26, 2, p. 283-301, 1935-1936. References are to the English translation. "Foundations of the Calculus of Systems", Chapter XII, p. 342-383, of Tarski (1956).

- . “The Semantic Conception of Truth and the Foundations of Semantics”. *Philosophy and Phenomenological Research* 4, p. 341-375, 1944.
- . *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press. Translated by J. H. Woodger. 2nd edition 1983. Edited by J. Corcoran. Indianapolis IN: Hackett Publishing Company, 1956.
- WRIGHT, C. J. G. *Truth and Objectivity*. Cambridge MA: Harvard University Press, 1992.