HISTORY AND PHILOSOPHY OF LOGIC, 13 (1992), 149–177

A Theory of Truth Based on a Medieval Solution to the Liar Paradox

RICHARD L. EPSTEIN
P.O. Box 751, Cedar City, Utah 84721, U.S.A.

Received 1 August 1991

In the early part of the 14th century Jean Buridan wrote a book called *Sophismata*. Chapter 8 of that deals with paradoxes of self-reference, particularly the liar paradox. Modern discussions of the liar paradox have been dominated by the formal analysis of truth of Tarski, and more recently of Kripke, and Gupta. Each of those either denies that the sentence 'What I am now saying is false' is a proposition, or denies that the usual laws of logic hold for such sentences. In Buridan's resolution of the liar paradox that sentence is a proposition, every proposition is true or false though not both, and the classical laws of logic hold.

In this paper I present a formal theory of truth based on Buridan's ideas as exposited by Hughes, contrasting it with the analyses of Tarski, Kripke, and Gupta. I believe that Buridan's ideas form the basis for the most convincing resolution of the liar paradox in a modern formal theory of truth.

I first survey the theories of Tarski, Kripke, and Gupta. Then I state the principles on which the Buridanian theory is based. After a brief description of how these principles are used in analyzing the truth-values of propositions, I set out the formal theory. Following that I discuss a number of examples in which the informal principles and the technical methods are explained and tested for their aptness; in those discussions I often draw on Buridan's explanations.

1. Modern theories of truth

Little is known of the life of Buridan beyond his writings: a Frenchman, he was Rector of the University of Paris in 1328 and in 1340, and died sometime after 1358. For a short discussion of his life and work consult Hughes *1982*; the most complete study of his life is Faral *1949*. My theory presented here is not an historical reconstruction, but in this section I provide some context in which to place Buridan's analysis by describing several modern formal theories of truth in terms of the principles on which they are based. The reader interested in a technical comparison of these and other modern theories of truth can consult Yablo *1985*, Hellman *1985*, and Burgess *1986*.[1]

The liar paradox in its simplest form is the assertion, 'This sentence is not true'.[2] It presents a difficulty for any formal theory of truth because, on the face of it, if it is true then it is false, and if it is false then it is true.

Tarski *1934* held that the solution to giving a technical analysis of truth without involving the liar paradox is to restrict attention to a formal language from which the

1 Comparisons could also be made with modern theories of truth that depend on some form of indexicality to resolve the liar paradox, such as that of Burge *1979*. And other medieval resolutions of the liar paradox have some points of contact with my version of Buridan's views; Simmons *1987* is an example as well as a useful reference. But to incorporate such comparisons here would require a monograph.

2 I use single quotation marks to indicate a quote or to form quotation names, and double quotation marks as scare quotes.

word 'true' is expunged. Since it is easy to construct other self-referential paradoxes if predicates such as 'false' or 'is satisfied by' are allowed into the language, Tarski chose to exclude from the formal language all predicates about the syntax or semantics of the language. He believed that what was left was a language suitable for mathematical and scientific theories in which one could assert a proposition, but not assert that it is true. Relative to a given interpretation of its symbols, every sentence of such a language is either true or false but not both, and the "classical" laws of 2-valued logic hold. However, discussions of the syntax and semantics of the language must take place in a metalanguage.

Tarski's analysis is in essence a theory of types in which not just particular words but whole languages are typed. At the lowest level is the formal language of mathematics and science; at the next level we have a copy of the first language as well as syntactic and semantic terms applicable to it; at the next level is a copy of the language of the second level with syntactic and semantic terms applicable to that; and so on.

Kripke *1975* argues that Tarski's semantic analysis is an unsuitable solution to the liar paradox in that it fails to model certain important uses of the word 'true' which he takes to be intuitive and correct. There is nothing paradoxical or wrong in asserting '2 + 2 = 4 and what I have just said is true'. Moreover, he points out that Tarski's notion of levels of languages is counterintuitive: if John says, 'What Robert is saying is not true', and Robert says at the same time, 'What John is saying is true', then these must be assigned different levels of the hierarchy of languages lest we have a paradox. Yet there is no apparent reason to claim that one should be of a lower level than the other.

Kripke argues that we can and should incorporate 'true' into the formal language if we consider the grounds on which a sentence is deemed true or false. Some sentences such as 'Snow is white' which refer to the world external to the language are simply true or false. They are grounded. A sentence such as 'It is true that snow is white' is grounded in 'Snow is white' which is itself grounded, so that it, too, has a truth-value. The liar paradox, 'This sentence is not true', is ungrounded, for an analysis of it does not lead to any grounded sentence, and hence it cannot be assigned a truth-value.

Kripke converts these ideas into a semantical analysis of truth in a formal language which incorporates its own truth-predicate by establishing a hierarchy of models. He first takes a model in which the predicates which would have been suitable for a Tarskian model, e.g., 'is a man', have their usual interpretation, and names are interpreted in the usual way, so that 'Socrates is a man' is satisfied in the model. But he leaves it indeterminate whether other predicates apply, particularly 'is true', so that some sentences such as 'It is true that Socrates is a man' or 'This sentence is false' have what Kripke calls undefined true-value. He then uses Kleene's strong 3-valued logic to evaluate the truth-value of compound and quantified sentences. Thus some sentences in this model are true, some false, and some have undefined truth-value. From this model he builds another in the same way, except that 'is true' is interpreted to be the collection of sentences true in the first model. Thus 'It is true that Socrates is a man' is true in the second model. Continuing to build models by expanding the interpretation of 'is true' to be the sentences true in the previous model, a fixed point is eventually reached where the next model built is the same as the previous one.

Depending on which sentences with 'is true' in them are deemed true or deemed

false at the first stage, one can get different fixed points. There is a minimal fixed point contained in all others which arises by assuming at the first stage that no sentence is true and no sentence is false. It is this which Kripke suggests would be a natural interpretation for a language which has its own truth-predicate. The other fixed points, however, are important in classifying sentences as "intrinsically" true, false, or paradoxical, grounded or ungrounded, accordingly as they are true, false, or have undefined truth-value in some or all of the fixed point models. In this analysis the liar paradox is classified as paradoxical, for its truth-value is undetermined at every stage.

Kripke's theory is a massive departure from the classical laws of logic: e.g., the law of excluded middle fails. Moreover, there are no sentences involving the predicate 'true' which are necessarily true, i.e., true in all models, due to the use of Kleene's three-valued logic. Kripke demurs on this point, saying that he has used Kleene's logic because it seems technically most apt, but he understands his methodology as a schema of theories of truth, depending on what technical device one uses to evaluate truth in models which allow sentences to have undefined truth-value. Much of Kripke's discussion seems to me to be about how we come to know which sentences are true, and his theory could be taken as a formal explication of the epistemology of truth.

Gupta's theory, *1982*, is apprarently closer in spirit to classical logic. He builds levels of models as Kripke does, except that he argues that we do and should make a guess at the truth-value of sentences which Kripke would have left undefined at the first level. Thus every sentence in the first model is (guessed at) true or false. This allows him to use Tarski's method of determining truth at each level. The result is that in each model the laws of classical logic are valid, e.g., every sentence is true or false. However, in the long run some sentences such as 'This sentence is not true' are neither true nor false for they have no stable truth-value, oscillating from true to false, false to true from level to level. His theory, then, is only superficially classical, for it is the resulting classification of sentences into stably true if true from some level onward, stably false, paradoxical, and so on, which Gupta takes to be the goal of his theory. His analysis, too, seems primarily concerned with the way we come to know which sentences are true.

Tarski resolves the liar paradox by claiming that it is not a proposition, and this allows him to retain the classical laws of logic. Both Kripke and Gupta allow that the liar paradox is a proposition, but then deny one of the basic assumptions of classical logic by claiming that some propositions are neither true nor false. Kripke jettisons classical logic entirely in favor of 3-valued logic; Gupta retains classical logic for how we reason hypothetically, but denies it is applicable in reality. All three of these logicians use hierarchies to analyze truth: Tarski uses hierarchies of languages; Kripke and Gupta utilise hierarchies of models.

In contrast, I will propose here a theory which first accepts that the liar paradox is a proposition. Every proposition is true or false but not both, and it is a separate issue how we come to know that. One formal language suffices to do logic, and relative to an interpretation of its formal symbols it has only one model. No classical assumptions are denied. Only one idealization which we use to simplify applications of logic is shown to be wrong for languages which allow self-reference. That idealization is that equiform tokens in a formal language must be, or mean, the same proposition, and hence have the same truth-value in a model.

If I say, 'I am the author of *A theory of truth based on a medieval solution to the*

liar paradox' and then you say the same words, then what I say is true and what you say is false. Different propositions are uttered. I will argue below that if a language has the means to refer to itself, and in particular 'This sentence is not true' can be formalized in it, then the same kind of indexicality will occur. Distinct equiform tokens can have different truth-values.

On the other hand, if the language has no ambiguous (indexical) words such as 'I' or 'this', and no means to refer to itself, then it is entirely correct to assert that any two equiform tokens will have the same truth-value in a model, as I will argue below. In that case we may identify them, saying that they are, or mean, the same proposition. In such a language we are justified in taking sentence types to be true or false relative to a model. This, however, is not an assumption of classical logic, but a simplification for our use of classical logic.

In what follows I will argue that it is the token itself, a physical utterance or inscription, which is the proposition. However, you may understand the theory presented here as a theory of abstract or mental propositions which are represented by tokens. In essence, the same questions need to be dealt with in both interpretations, as explained in Example 3 below.

2. The principles

2.1 In this section I will present the principles on which I base the formal theory of truth of §3. It is these fundamental principles which I believe should be at the heart of any debate about a solution to the liar paradox. In the discussion following the formal theory I will show the significance of each of these and attempt to eliminate ambiguities in their interpretation.

These principles are derived from Buridan as exposited in the translation and commentary given in Hughes *1982*, with two important exceptions. First, I have replaced the medieval theory of suppositions by Tarski's notion of satisfaction in a model. And second, I do not invoke mental propositions to explain how tokens can be meaningful. Instead I argue for meaningfulness in terms of a notion of agreement.

Neither these principles nor the formal (i.e., technical) theory based on them should be understood as giving a definition or complete characterization of truth. Rather, I believe that I am bringing out further aspects of a common notion and investigating their consequences. I cannot begin to say to what extent this is descriptive as opposed to prescriptive; I believe that if the analysis is convincing enough it may come to be seen as descriptive even it it was originally prescriptive (see Epstein *1990*, ch. II). Nonetheless, I claim that this analysis takes into account all the aspects of truth necessary to set out the truth conditions for a wide class of sentences. I do not claim, however, that I have considered all aspects of this notion: for instance, it may be that to extend this class to include operators such as 'knows that' further aspects of truth will need to be taken into consideration. Consistency, too, is an informal notion which is at best partially described in the principles and formal theory.

Principle 1 A proposition is a specific linguistic entity, a sentence token which is uttered or written at a specific time and which we agree to view as being either true or false, but not both.

In Example 3 I show why different tokens can have distinct truth-values, and

hence why it is wrong to ascribe truth-values to sentence types. In that example I also explain briefly why I say we agree to view a proposition as being true or false; a fuller explanation can be found in Epstein, *1990*, ch. I and II.

Throughout I will use 'uttered' for 'written or uttered'.

Principle 2 Propositions come into existence and cease to exist just as any other objects in the world. Until they come into existence they are not among the objects under discussion.

Strictly speaking we are constrained by this principle to say of an uttered and not written proposition that it was true or false, since such propositions exist for only a moment. However, to simplify our discussions I will assume that once a proposition exists it continues to exist for the duration of our analysis, as would be the case with written propositions. Examples 8 and 9 are particularly concerned with examining some of the consequences of this principle.

Principle 3 If in a discussion we agree that equiform words are to be understood in the same way, then equiform propositions will have the same truth-conditions if they contain no semantic or syntactic terms or names which can engender self-reference. Hence we may identify them as being the same for the purposes of logic.

In Example 1 I argue for this principle and show why it allows us to use Tarski's notion of satisfaction in a model for languages which do not contain predicates which refer to the syntax or semantics of the language.

Principle 4 The classical laws of logic hold.

Principle 5 For a proposition to be true things must be the way it says they are; that is, the material conditions for the truth of a proposition must hold. For propositions in a formal first-order language the classical (Tarskian) analysis of truth is the correct interpretation of the truth-conditions of a proposition. Moreover, if the formal language does not contain its own truth-predicate or any other semantic or syntactic predicate applicable to itself, then we may take sentences to be types, and it suffices for the truth of a proposition that its material conditions hold.

Henceforth, I will call the Tarskian truth-conditions for a proposition A *the material conditions for*, or *of*, *A*. If we chose we could give an alternative theory of truth based on a nonclassical semantic interpretation of truth in a formal language which does not admit self-reference, such as intuitionistic first-order logic. However, it is more in the spirit of Buridan to take Principles 4 and 5, using Tarski's models instead of the medieval theory of suppositions. Moreover, this allows us to make as little departure as possible from the most commonly accepted forms of logic so that what is new in Buridan's analysis will be more apparent.

In Example 1 I explain why Tarski's theory is justified in terms of Principle 3. In Example 14 I discuss the disquotational aspect of Tarski's theory of truth.

Principle 6 If A is true, then any subsequent utterance of 'A is true' is true, and hence the material conditions of the latter holds. Colloquially, if A is true, then 'A is true' is true; and conversely, if 'A is true' is true, then A is true.

This is the principle of *truth entailment.*

Principle 7 If all the facts are in, then there are no arbitrary choices to be made in determining the truth-value of a proposition. In particular, a proposition cannot be true simply by assuming it to be true.

This is the principle that *truth is not arbitrary.*

The technical explanation of 'all the facts are in' will be that the predicates and names of the language are interpreted in a model, except possibly for whether 'true' applies to the sentence in question. This principle is needed to resolve the truth-teller, 'This sentence is true', which is Example 4.

Principle 8 Principles 5, 6, and 7 give necessary and sufficient conditions for a proposition to be true:

$$\text{A is true iff } \begin{cases} \text{(i) the material conditions for A hold} \\ \textit{and} \\ \text{(ii) it is consistent and not arbitrary that (i).} \end{cases}$$

The difficulty in formalizing Buridan's ideas is to explain what we mean by 'consistent' here. Example 2 is devoted to that, which is part of the burden of Principle 6; a colloquial description is given in the informal explanation below.

In what follows I understand 'A is not true' and 'A is false' to mean the same, though in Example 14 I consider applying the term 'false' to propositions only.

2.2 Having set out the principles, let me now describe informally and colloquially the method of determining the truth-values of propositions. I will describe the process in terms of how we come to know the truth-values, though that is only a convenient expository device.

Some propositions are uttered. I would like to know which are true and which are false. I know how an analysis of the material conditions of each proposition should be made. And I know that I must be consistent: if in analyzing A I conclude that A is true, I should not later have to retract that and say that A is false due to the principles I have set out.

I survey the possible ways I can assign truth-values to the propositions uttered simultaneously with A. For each one of these assignments I see if the facts, that is the material conditions of A relative to our linguistic conventions and model, force me after sufficient analysis to assert that A is true rather than false on pain of inconsistency. If I am always forced to assert that A is true, then A must be true and there is nothing arbitrary about that. On the other hand, if for even one of these possible assignments: (i) I would be forced to assert that A is false on pain of inconsistency; or (ii) the material conditions of A force me to alternately assert A and to retract my assertion, then A is false. It would be an arbitrary choice to avoid that case, and truth is not arbitrary.

Suppose that we have concluded that A is false, yet the material conditions of A

would force us to assert that A is true if we begin with that assumption. Then it is not inconsistent to assert that A is false, for the material condtions of A are only part of its truth-conditions. Since A is false, it must be because it is either arbitrary or inconsistent with the facts to assume that A is true.

3. The formal theory

I do not believe that formalizing the principles of the last section will make them more "scientific". Still, though we may be able to apply the principles to many of the examples in §4 without any technical apparatus, it is not clear that these principles can account for all propositions in a language which contains its own truth-predicate, and that they generate no contradiction. In particular, the notion of consistency seems unclear, and I have only been able to convice myself of the coherence of Buridan's ideas by employing a formal device for dealing with hypothesized truth-values suggested by the technical semantics of Gupta *1982*[3].

There will always be more than one way to convert informal principles into a formal theory. Various technical assumptions are needed, and for the theory here I have labeled four of them as pragmatic assumptions. In the context of the formal theory each is reasonable, but I view them as less fundamental than the principles of the last section.

3.1. *The formal language*

Pragmatic Assumption 1 In any discussion (model) in which we agree to understand equiform words in the same way, the differences between equiform inscriptions of words do not matter to logic. We thus take the vocabulary of the formal language to be types, where a type is understood not as an abstract object, but as an identification of distinct inscriptions as being the same for the purposes of logic.

From Principles 1 and 2 a language is not a completed infinite collection of sentences but rather a vocabulary and a way to generate sentences from the words. Unless I say otherwise, I will always mean by *a language* a (formal) vocabulary and formation rules.

Sometimes it is a justifiable abstraction to view a language as an infinite collection of sentences, or even sentence types; I discuss this in Example 1.

We take as our language the usual language of first-order logic **L**, supplemented by the predicate (symbol) **T**, which we intend to interpret as 'is true'. It is immaterial to the discussion that follows whether or not **L** has an equality predicate or names, other than the names of sentences of **L** described below in Pragmatic Assumption 2. The formation rules for sentences are the usual ones.

3 I also hope that this formal theory will help to rebut what seem to me to be misinterpretations of Buridan. Scott *1966*, 56*ff* believes that Buridan's solution of the liar paradox fails, but he apparently misunderstands that the condition for a proposition to be true is a conjunction of clauses, each of which can individually fail (Principle 8). Herzberger *1975* basing his work on Scott's translation, argues that a many-valued formal semantics is necessary to adequately represent Buridan's views on the relationship between the material conditions and the consistency of a proposition, but this contradicts Buridan's assertion that the classical laws of logic hold. Angelelli *1985* believes that the principle of truth entailment (Principle 6) adds nothing to the material conditions for a sentence to be true and is a misreading of Buridan by Hughes.

I understand the predicate and name symbols of the formal language as first being realized by English language word-types such as 'is a dog' for predicate symbol '$\mathbf{P_0}$', and 'Ralph' for the name symbol '$\mathbf{a_0}$'. The model described below is then given for this "semi-formal" language in which Pragmatic Assumption 1 does matter. In Epstein *1990* and *1993* I discuss the relationship between the formal language, the semi-formal language, and formal semantic models.

It has been suggested to me that it is necessary to give a theory of strings for tokens prior to discussing the formal language, as Tarski *1934* does for types. But this, I believe, assumes that informal mathematics is prior to and justifies formal logic, which I think is wrong. I discuss this further in Example 3 and Epstein *1990*, ch. 2.F.[4]

3.2 *The model.* I will define a model for **L** which is extended as sentences of **L** are uttered. It will determine the truth-value of every sentence of **L** uttered so far.

Let **L*** be the language of **L** with **T** deleted from its vocabulary. Let $\mathfrak{M}$ be a Tarski model for **L***. I understand the model for **L** described below as an extension of $\mathfrak{M}$ and the semantics as an extension of Tarski's (Principle 5). The interpretation in $\mathfrak{M}$ of the predicates and names of **L*** codes whether each predicate is or is not satisfied by each (sequence of) object(s) of the universe we are discussing. I will colloquially refer to this interpretation as *the given facts* or *the facts of the matter.*

It does not follow that we need all the machinery of Tarski's semantics along with the mathematical assumptions on which it is based. In Epstein *1990* I discuss the role of mathematics in formalizing logic and the extent to which infinitistic assumptions are necessary, and in Epstein, *1993* I show how it is possible to give a constructive, nominalist reading of the usual Tarskian semantics. But for the presentation here it is simplest if we assume the technical machinery of Tarskian semantics as well as whatever mathematics you might assume necessary for developing that. In particular, I will be assuming Tarski's notions of reference, satisfaction, truth in a model, etc.

Stage n, $n \geq 0$

The set of sentences $\mathfrak{S}_n$

Any sentences of **L** can be uttered as propositions at stage n. At this stage we will collect only some of them for analysis of their truth-values, essentially imposing a linear order on them except that we may need to consider several simultaneously, due to interlocking references. It is possible to consider a partial order on sentences and an analysis of various chains of that order simultaneously, but that complication adds nothing essential to the ideas here. The conditions we set out for what sentences are considered to be in $\mathfrak{S}_n$ can be viewed as simplifying pragmatic assumptions.

4 I am hardly the first modern logician to propose basing logic on tokens. Jaskowski *1934*, 235 says:
> In order to avoid any misunderstanding, we must always remember that, by an expression, a thesis etc., we shall treat a given inscription as a material object, just as Professor S. Lesniewski did in the explanations concerning his systems. Thus two inscriptions having the same appearance but written down in different places must never be taken as identical; they can only be said to be *equiform* with each other.

Markov *1954* bases his theory of algorithms on concrete letters, alphabets, and words, in contradistinction to abstract letters, alphabets, and words, and he gives a technical analysis adequate for the formal theory I present here.

Pragmatic Assumptions 2 At this stage or any later stage we may name any of these sentences and use those names in forming new propositions. In particular, those names may be used to form propositions in $\mathfrak{S}_n$. Quotation names, however, are not allowed.

It may be possible to use the methods of this paper to resolve paradoxes of ungrounded reference, such as Berry's paradox.[5] However, for simplicity of exposition here I have chosen to use the following simplification, already tacitly assumed for the Tarski model.

Pragmatic Assumption 3 All names refer. In particular, if $A \in \mathfrak{S}_n$ and '**b**' is a name of a proposition and '**b**' appears in A, then **b** has already been uttered or is uttered at this stage. That is, $\mathbf{b} \in \bigcup_{m \leq n} \mathfrak{S}_m$.

The next assumption allows us to treat those cases where the sentences uttered contain interlocking references which must be analyzed simultaneously:

Pragmatic Assumption 4 We impose a logical ordering on the sentences uttered at stage n by dividing the utterances into two parts. In the latter part goes any proposition B such that there is another proposition A and (i) the method of analysis of the truth-value of A described below gives the same truth-value regardless of whether B had been uttered or not, but (ii) the truth-values of B in the analysis below (or even whether B is a proposition), depends on whether A has been uttered and analyzed before B.

The propositions in the first part are all those not described above, and they are to be analyzed within the theory first. By 'first' I mean within the logical analysis below, which is not to be construed as temporal since the order of utterance of the propositions is fixed temporally by our stages. Nonetheless, rather than use 'stage n first part' and 'stage n second part' which would involve further confusing notational devices, I will assign the propositions in the second part to a later stage, as if we were temporally deferring them. This is only a convenient expository device (it accords well with written propositions that we could assume continue to exist throughout the analysis at hand).

In Example 3 I show that, while this last assumption may not be essential, it resolves some otherwise puzzling situations. That example will also demonstrate why quotation names are not allowed.

It follows from Principles 1 and 2 that we have only a finite number of propositions uttered at any one time. In Examples 16 and 17 I consider making the abstraction that infinitely many sentences can be uttered simultaneously, based on our ability to describe schematically a potentially infinite collection of sentences. Also, in the discussion of the applicability of Tarski's semantics to our formal language in Example 1 we will see that it is a reasonable simplification to take **L*** as a completed infinity of sentence-types.

5 A version of Berry's paradox arises if we let e = the least natural number not denoted by any English expression of thirty words or less. Such a number must exist since there are only a finite number of English expressions using less than thirty words. But then that number is denoted by just such an expression.

The truth-values of the propositions in $\mathfrak{S}_n$

Each proposition on being uttered is true or false, and its truth-value never changes. But how we come to see what its truth-value is can best be described in stages, by assigning hypothetical truth-values according to the rules below.

We have a model of all propositions in $\bigcup_{m \leq n-1} \mathfrak{S}_m$. I will denote by $\mathfrak{T}_{n-1}$ and $\mathfrak{F}_{n-1}$ the propositions of $\bigcup_{m \leq n-1} \mathfrak{S}_m$ which are, respectively, true and which are false in that model. At this stage the colloquial phrase *the given facts* now also refers to the truth-values of all propositions uttered so far as well as the information coded by $\mathfrak{M}$.

Informally, we first make some hypothesis about which propositions of $\mathfrak{S}_n$ are true. Those plus the propositions true so far, $\mathfrak{T}_{n-1}$, we collect and label as $\mathfrak{T}_n(O)$. We then consider the Tarski model $\mathfrak{M}$ augmented by interpreting **T** as $\mathfrak{T}_n(O)$. The set of sentences true in that model we call $\mathfrak{T}_n(1)$; now using those to interpret **T** we have another Tarski model. This process of considering the semantic consequences of the hypothetical truth-values $\mathfrak{T}_n(k)$ can be continued indefinitely. If the facts are such that A really is true, then eventually we will be forced to recognize this, for A will stabilize as being in $\mathfrak{T}_n(k)$ for all large k. On the other hand, we may have that (i) for all large k, A is not in $\mathfrak{T}_n(k)$, or (ii) for arbitrarily large j and k, A vacillates between being in $\mathfrak{T}_n(k)$ and out of $\mathfrak{T}_n(j)$. If that should happen on even one assumption about which propositions of $\mathfrak{S}_n$ are true, that is for even one choice of $\mathfrak{T}_n(O)$, then A is false. For it is not necessary, relative to the facts, that A be true. A is true if and only if for each such analysis A will be in $\mathfrak{T}_n(k)$ from some point onward.

For convenience, the sentences not in $\mathfrak{T}_n(k)$ will be labelled $\mathfrak{F}_n(k)$, for they are the ones false in the k^{th} hypothetical Tarski model.

In essence, relative to every possible assumption about which propositions of $\mathfrak{S}_n$ could be true, we construct a sequence of classical possible worlds in order to answer the question 'Would it be consistent to asset that A is true?' It might seem that to answer that question an infinite sequence of possible models will be required, but in Lemma 3 I show that a finite number suffice.

Formally, suppose $\mathfrak{S}_n = \{A_1, \ldots, A_m\}$. Let $\alpha_1, \ldots, \alpha_{2^m}$ be a list of all subsets of $\mathfrak{S}_n$. *For each* α_i we have the following *substage analysis*:

Substage 0: If we are at stage $n = 0$, then $\mathfrak{T}_o(0) = \alpha_i$, $\mathfrak{F}_o(0) = \mathfrak{S}_o - \alpha_i$. If $n > 0$, then $\mathfrak{T}_n(0) = \mathfrak{T}_{n-1} \cup \alpha_i$, and $\mathfrak{F}_n(0) = \mathfrak{F}_{n-1} \cup (\mathfrak{S}_n - \alpha_i)$.

Substage k + 1: $\mathfrak{T}_n(k+1) = \mathfrak{T}_{n-1} \cup \{A \in \mathfrak{S}_n$: using the Tarskian analysis, A is true in the model $\mathfrak{M}$ augmented by expanding its universe to include all propositions of $\bigcup_{m \leq n} \mathfrak{S}_m$ and interpreting **T** as $\mathfrak{T}_n(k)\}$. And $\mathfrak{F}_n(k+1) = \mathfrak{F}_{n-1} \cup \{A \in \mathfrak{S}_n$: $A \notin \mathfrak{T}_n(k+1)\}$.

An easier way to think of and to write the definitions for substage k + 1 is to let t denote the operator corresponding to what is in the brackets in the definition of $\mathfrak{T}_n(k+1)$, operating on the set of sentences which interpret **T** (throughout this stage, $\mathfrak{M}$ and $\bigcup_{m \leq n} \mathfrak{S}_m$ are fixed). Sometimes t is called 'Tarski's machine'. Then

$$\mathfrak{T}_n(k+1) = \mathfrak{T}_{n-1} \cup t(\mathfrak{T}_n(k))$$

and

$$\mathfrak{F}_n(k+1) = \mathfrak{F}_{n-1} \cup \{A \in \mathfrak{S}_n : \mathfrak{A} \notin t(\mathfrak{T}_n(k))\}$$

Note that for all substages k, $\mathfrak{T}_{n-1} \subseteq \mathfrak{T}_n(k)$ and $\mathfrak{F}_{n-1} \subseteq \mathfrak{F}_n(k)$. The truth-values of the propositions in $\cup_{m \leq n-1} \mathfrak{S}_m$ are not affected by any new propositions which are uttered.

The propositions in $\cup_{m \leq n} \mathfrak{S}_m$ which are *true under the hypothesis of* α_i, that is under the assumption that the true propositions in $\mathfrak{S}_n$ are those in α_i, are:

$$\mathfrak{T}_n(\alpha_i) = \mathfrak{T}_{n-1} \cup \{A \in \mathfrak{S}_n: \text{there is some m, such that for all } k \geq m,\ A \in \mathfrak{T}_n(k)\}$$

and *the false ones under that assumption* are:

$$\mathfrak{F}_n(\alpha_i) = \mathfrak{F}_{n-1} \cup \{A \in \mathfrak{S}_n : A \notin \mathfrak{T}_n(\alpha_i)\}$$

Finally, the propositions in $\cup_{m \leq n} \mathfrak{S}_m$ which are *true* are those which are true under any assumption about the truth-values of the propositions in $\mathfrak{S}_n$. That is:

$$\mathfrak{T}_n = \mathfrak{T}_{n-1} \cup (\cap_i \mathfrak{T}_n(\alpha_i))$$

The ones which are *false* are those which are not true, that is,

$$\mathfrak{F}_n = \mathfrak{F}_{n-1} \cup \{A \in \mathfrak{S}_n : A \notin \mathfrak{T}_n\} = \{A \in \cup_{m \leq n} \mathfrak{S}_m : A \notin \mathfrak{T}_n\}$$

We can view **L** as the collection of propositions we have whenever we wish to terminate our analysis, that is as $\cup_{m \leq n} \mathfrak{S}_m$ for some n. Or we may think of **L** as what we would get "if we went on forever". In that case we must have a way to specify $\mathfrak{S}_n$ for each n and agree to the idealization that there can be arbitrarily long inscriptions. With respect to the given facts the collection of true propositions is then $\mathfrak{T} = \cup_n \mathfrak{T}_n$, and the false propositions $\mathfrak{F} = \cup_n \mathfrak{F}_n$.

I will assume a fixed model $\mathfrak{M}$ throughout the following discussion, so I will say simply 'true' or 'false' now. Sometimes when the context makes it clear that I am talking about a substage analysis I will say that A is true when I mean that it is hypothetically true, and similarly for 'false'.

3.3 *Some observations*. In proving that several of the principles on which the theory is based are appropriately modeled, the questions arise: What is the language of our discussions? What is the metalogic? I am not about to give formal answers to these questions: you can assume that I am using "intuitive" logic, or better, classical logic for the (formalizable) metalanguage (as justified by Example 1).

Let us establish that Principles 4 and 6 hold in our formal theory.

Lemma 1 Every proposition is true or false but not both. If A is an instance of a classical tautology, then A is true. If A is an instance of a classical anti-tautology it is false.

Proof The first part is clear. For the second part, let $A \in \mathfrak{S}_n$. If A is a tautology, then for any choice of $\mathfrak{T}_n(0)$ we have that $A \in \mathfrak{T}_n(k)$ for all $k \geq 1$, so A is true. If A is an anti-tautology, then $A \in \mathfrak{F}_n(k)$ for all $k \geq 1$, and hence A is false.■

Lemma 2 The principle of truth-entailment holds: if A is true and '**a**' is a name of A, and B is a proposition of the form **T(a)** then B is true; and conversely, if B is true then A is true.

Proof If $A \epsilon \mathfrak{T}_n$ and $B \epsilon \mathfrak{S}_k$ for some $k > n$, then the lemma is straightforward. Pragmatic Assumption 4 rules out the case where A and B are both in $\mathfrak{s}_n$, unless B is A itself. In that case A is false, as I will show in Example 4 below.■

Lemma 2 would hold even were we not to employ Pragmatic Assumption 4: we would have that if A and B are both in $\mathfrak{S}_n$, then for some m, all $k \geq m$, $A \epsilon \mathfrak{T}_n(k)$, hence all for all $k \geq m + 1$, $B \epsilon \mathfrak{T}_n(k)$. So $B \epsilon \mathfrak{T}_n$, too.

Lemma 3 If there are exactly m propositions in $\mathfrak{S}_n$ and $A \epsilon \mathfrak{S}_n$ is true, then for each substage analysis A will be true from substage $2^m - 1$ onward.

Proof The truth-value of A at any substage of any analysis depends on only two things: which predicates apply to which objects in $\mathfrak{M}$, and which of the A_i's are true. The only other new objects introduced at stage n are parts of the propositions $A_1, \ldots, A_m$ and, as explained in Example 14, the predicate **T** is not satisfied by any of those. Therefore, the truth-values of $A_1, \ldots, A_m$ at any substage are completely determined by the truth-values assigned at the preceding substage. Thus if any combination of truth-values for $A_1, \ldots, A_m$ appears at a substage k and again at substage $k + r$, then substages $k, k + 1, \ldots, k + r - 1$ are forever repeated in that sequence, as nothing additional can enter into the calculations. There are 2^m different possible combinations of truth-values for $A_1, \ldots, A_m$, hence by substage $2^m - 1$ (recall we start at substage 0) every combination which can appear in a particular analysis will have appeared. Hence if A is true from some substage onward, it will be true from substage $2^m - 1$ onward.■

Corollary 4 if there are exactly m propositions in $\mathfrak{S}_n$ then the truth-value of every $A \epsilon \mathfrak{S}_n$ can be determined by calculating at most 2^{2m} different substages in various substage analyses.

Proof There are 2^m different possibilities for the hypothesis we make at substage 0. For each of these we need only calculate 2^m further substages in order to determine whether A will be true under that hypothesis.■

The bound in Corollary 4 can be improved: if in the substage analysis of $\mathfrak{T}_n(\alpha_i)$, $\mathfrak{T}_n(k) = \alpha_j$ for some $j \geq i$, then $\mathfrak{T}_n(\alpha_j)$ need not be calculated separately. Of more interest is whether the bound in Lemma 3 can be improved. In Example 7 I show that for every m there are m propositions such that one of them is true but does not settle down to appear so until substage m in at least one substage analysis.

The import of Lemma 3 is that relative to the model $\mathfrak{M}$ the truth-values of atomic propositions in this theory are constructively decidable. No infinite procedure must be accomplished before the truth-value of an atomic proposition can be determined. This is also the case for Tarski's theory, but apparently does not hold for Kripke's and Gupta's (see Burgess *1986*). I comment further on this in Example 10.

4. Examples

We can now begin our investigation of examples which will test whether this theory adequately models our intuitive conception of truth. Each example will be in English followed by a formalized version. The details of the formal technical analysis are usually easy, and I will sketch them only. What I think is important is how the examples help us to understand puzzling common language propositions, and what

Here 'A' must name a proposition other than the conclusion of the rule (see Lemma 2 and Example 4). Similarly, we have the derived "rule":

$$\frac{A \text{ is false}}{\begin{array}{c}\neg\mathbf{T(a)} \\ | \\ \mathbf{b}\end{array}}$$

where **a** is a name of A, so long as **b** is uttered after **a**.

In Example 15 I discuss the rule of *modus ponens*.

Example 10 *No proposition has the word 'true' in it.*
$\neg\exists x(P(x)\wedge W(x))$

Let us call the formal proposition **a**. Gupta *1982* has shown how we may include predicates such as 'has the word true in it' in Tarskian models of first-order logic: the self-reference they engender is apparently harmless.

Let us then suppose that we have the predicate 'has the word 'true' in it' in our model $\mathfrak{M}$, and our example is in the language **L*** (which is **L** with '**T**' deleted). Suppose also that **a** is uttered at stage 0. Then **a** is true. Suppose also that at the next stage the following proposition is uttered: **T(b)** where **b** names **a**. And then a proposition **c** equiform with **a** is uttered at stage 2. Then **c** is false. It is not that **a** has changed its truth-value: **a** is still false, for it refers to the time of its utterance, namely stage 0. If you like, all propositions are indexed by their time of utterance. At stage 2, **c** is false: the world has changed (Principle 2).

Self-reference is never really harmless if tokens are taken as the bearers of truth-values. For that reason I believe it best to assume that $\mathfrak{M}$, the model for **L***, has no syntactic or semantic predicates which apply to **L***. Predicates of that kind may be introduced in the same way as 'is true'.

Thus the example is true or false depending on its time of utterance. Its truth depends on what sentences have already been uttered (at an earlier stage or the same stage). Sentences are part of the world, and the order in which they are uttered matters, just as it matters whether Plato died after Socrates for the truth-value of 'Plato wrote a description of the death of Socrates'. Our formal theory has three parts: **L**, $\mathfrak{M}$, and $\{\mathfrak{S}_n: n \geq 0\}$.

Nonetheless, I claim that this theory allows us to dispense with hierarchies, unlike the theories of Tarski, Kripe, or Gupta. Tarski's resolution of the liar paradox imposes a hierarchy of languages on our utterances. The classification schemes of Kripke and Gupta depend on transfinite hierarchies of models.

But suppose you claim that there is a hierarchy here: the $\mathfrak{S}_n$'s. You say that I have not produced one language, but many. I reply that English is just one language, yet new sentences appear all the time. I have only one language, which can be extended by ostensive naming, and one set of rules for forming sentences.

Yet you persist and say that my theory has a hierarchy of models in the substages every bit as much as Kripke's or Gupta's. No, I reply, the substage analysis is there for my convenience, to understand Principles 1–8. Propositions are true or false when uttered (Principle 1), and the finite number of hypothetical models used to analyze that (Lemma 3) are no more a hierarchy than are the rows of a truth-table used to analyze a proposition in classical propositional logic.

Example 11 *The first proposition uttered in the 21st century is true.*

No problems unique to our theory are posed by this example.

If you wish to claim that this sentence is a proposition and thus is either true or false even though the descriptive name in it does not refer to an existing object, then I would say that we shall have to wait until the 21st century to know which, and hence to set out a model for it.

My preference is to say that the example has no truth-value if the name in it does not refer. This is reflected in Pragmatic Assumption 3. The same assumption is usually made when a model is given for a first-order language. This solution avoids the problem which would occur if this sentence were uttered at, say stage 1, and at stage 2 we have the sentence 'No proposition with the words '21st century' in it is true' whose truth-value would depend on the present example.

There are other possible solutions which would be consonant with the theory developed here (with Pragmatic Assumption 3 modified), for example treating the sentence as false using Russell's theory of descriptions. But then, is this sentence a proposition? I do not believe that there is a "fact of the matter" as to whether it is or is not. This is one of the borderline cases which we must agree to resolve, as discussed in Example 3.

Example 12 *Every proposition is true or false.*

We have two ways to formalize this:

$$\forall x(\mathbf{T}(x) \vee \neg \mathbf{T}(x)) \quad \text{or} \quad A\forall x(\mathbf{P}(x) \rightarrow [\mathbf{T}(x) \vee \neg \mathbf{T}(x)])$$

This is true, as we demonstrated in Lemma 1. But note that if it is uttered at stage n, then the propositions it refers to are those in $\mathsf{U}_{m \leq n} \mathfrak{S}_n$. If an equiform proposition is uttered at a later stage it will refer to a larger class of propositions. The proposition is a necessary truth: it is true no matter when uttered, it is true in all possible worlds.

Suppose you argue that this example demonstrates the need for abstract propositions: there is no sentence which can "mean" that every proposition, no matter when uttered, is true or false. That is, you would say, there is no sentence which timelessly captures the principle of *tertium non datur*. But I reply that our discussion has already pointed to such a sentence: '"Every proposition is true or false' is a necessary truth'. In Gupta's theory of truth this example is also true, but counter-intuitively so. In his analysis it is true in the short run because we use Tarski's machinery to evaluate truth. But in the long run, in the classification system which he claims is the point of his theory, this is not the case, for every proposition is not true or false, but stably true, or stably false, or paradoxical, etc. If his discussion of truth as a revision process is to be taken seriously, then the example is true only of hypothetical, guessed-at truth-values, and is false in reality.

In Kripke's theory the example is not true due to his use of a 3-valued logic to generate his classification scheme. This leads to a problem with what could be called the strengthened liar paradox. Suppose **a** is paradoxical, that is, **a** has undefined truth-value in the minimal fixed point model. Then clearly **a** is not true, which is further reflected in that $\mathbf{T}(\mathbf{a})$ has undefined truth-value in that model, too. But then so also does $\neg\mathbf{T}(\mathbf{a})$. Thus '**a** is not true' does not mean '**a** is false or has undefined truth-value', and indeed there is no way to express the latter in Kripke's theory.

Example 13 *This sentence is true or false*

$$\mathbf{T(a) \vee \neg T(a)}$$
$$|$$
$$\mathbf{a}$$

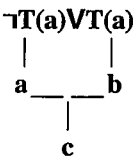
As a substitution instance of the previous example (a concept which can be made precise for sentence tokens just as easily as sentence types, since we are assuming that words are types), this must be true. And indeed it is: at whatever stage it is uttered, whether it is first assumed to be false or true at all further substages it is evaluated as true.

Example 14 *The first disjunct of this sentence is false or the first disjunct of this sentence is true.*

$$\mathbf{\neg T(a) \vee T(a)}$$
$$|$$
$$\mathbf{a}$$

Principle 1 asserts that sentences are the bearers of truth-values. A consequence of this is that only propositions, which are sentences, can be true, not parts of them. Thus the first disjunct, which we name '**a**', cannot be true. So our substage analysis will show that the entire proposition is true.

What would happen if parts of propositions were to be considered as propositions? Call the second disjunct **b**, and the entire proposition **c**.



If **c** is analyzed at stage n, then let us first suppose that **a** and **b** are false:

$$\begin{array}{lllllll} & \mathbf{a} \in \mathfrak{F}_n(0) & \text{and} & \mathbf{b} \in \mathfrak{F}_n(0), & & \\ \text{so} & \mathbf{a} \in \mathfrak{T}_n(1) & \text{and} & \mathbf{b} \in \mathfrak{F}_n(1) & \text{and} & \mathbf{c} \in \mathfrak{T}_n(1), \\ \text{so} & \mathbf{a} \in \mathfrak{F}_n(2) & \text{and} & \mathbf{b} \in \mathfrak{T}_n(2) & \text{and} & \mathbf{c} \in \mathfrak{T}_n(2), \end{array}$$

and at all further stages the hypothetical truth-values of **a** and **b** switch, and so are opposite one another. So $\mathbf{c} \in \mathfrak{T}_n(k)$ for all $k \geq 1$. The same conclusion is reached for any other initial assumption about the truth-values of **a** and **b**. Thus **a** would be false, yet **c** would be true. This is counterintuitive and unacceptable on the basis of Principle 8, as the material conditions of **c** would fail. Alternatively, if we were to analyze the truth-value of **c** after **a** and **b**, then we would have that **c** is false; contradicting Principle 4.

But you might argue that I am committed to parts of propositions being propositions by using the Tarskian machinery for evaluating truth-values. For instance,

(1) 'Snow is white or grass is green' is true
iff 'Snow is white' is true or 'grass is green' is true
iff snow is white or grass is green.

For this analysis to make sense, you might claim, we need that the disjuncts 'Snow is white' and 'grass is green' are propositions in their own right, with truth-values.

I argue that we do not need to assume that 'Snow is white' and 'grass is green' are propositions and have truth-values, where the quoted words are meant to name the parts of the original proposition. Indeed, all our experience of self-reference has shown us how risky it is to make the first step at (1) because of the ambiguity of quotation names. I view (1) as a (usually) harmless way of abbreviating the correct analysis:

Call the proposition: *Snow is white or grass is green*
by the name **d**. Then:

d is true iff the material conditions for **d** hold
and (*) { it is consistent that they hold and there is nothing arbitrary about that }
iff (the material conditions for the first disjunct of **d** hold) or (the material conditions for the second disjunct of **d** hold) and (*)
iff (snow is white or grass is green) and (that is consistent and there is nothing arbitrary about that)

If we ignore (*), which is redundant for non-self-referential propositions, then we have Tarski's *1934* disquotational analysis of truth, convention (T). But the way we have come to it is by the more cautious route, forced on us by our work with self-referential sentences, which talks of the material conditions expressed by, or for, the parts of a proposition. Of course, if you wish to deal only with non-self-referential sentences, as Tarski's analysis usually does, then it is perfectly harmless to identify the material conditions of a proposition with its truth-conditions, and hence to say that parts of a proposition are true or false. But if we wish to work with self-referential sentences then the question of their existence is sometimes crucial to the analysis of their truth-value, and parts of a proposition do not exist as completed utterances in their own right the way a propsition does.[10]

This example also points out that perhaps falsity is not adequately represented in our model if 'false' is to be identified with 'not true'. For in that case 'The Parthenon is not true' is true and hence the Parthenon is false, even though 'The Parthenon' does not name a proposition. So I suggest we incorporate an additional predicate 'is a proposition' into our model, representing it in our language as **P**. Then we can define $\mathbf{F(x)} \equiv_{\mathrm{Def}} \mathbf{P(x)} \wedge \neg \mathbf{T(x)}$ and **F** can be interpreted as 'is false'. Note then that $\forall \mathbf{x}(\mathbf{T(x)} \vee \mathbf{F(x)})$ will not then be true and would be an incorrect formalization of Example 12.

That a part of a proposition is not a proposition bears on how we use the rule of *modus ponens*.

10 Buridan gives a number of further arguments why a part of a proposition is not a proposition.