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.

0144–5340/92 $3.00

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, 56ff* 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 $\bigcup_{m \le n-1} \mathfrak{S}_m$. I will denote by $\mathfrak{T}_{n-1}$ and $\mathfrak{F}_{n-1}$ the propositions of $\bigcup_{m \le n-1} \mathfrak{S}_m$ 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 kth 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 $\bigcup_{m \le n} \mathfrak{S}_m$ and interpreting **T** as $\mathfrak{T}_n(k)\}$. And $\mathfrak{F}_n(k + 1) = \mathfrak{F}_{n-1} \cup \{A \in \mathfrak{S}_n\text{: } A \notin \mathfrak{T}_n(k + 1)\}$.

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 $\bigcup_{m \le n} \mathfrak{S}_m$ 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 $A \in \mathfrak{T}_n$ and $B \in \mathfrak{S}_k$ 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 $k \geq m$, $A \in \mathfrak{T}_n(k)$, hence all for all $k \geq m + 1$, $B \in \mathfrak{T}_n(k)$. So $B \in \mathfrak{T}_n$, too.

Lemma 3 If there are exactly m propositions in $\mathfrak{S}_n$ and $A \in \mathfrak{S}_n$ 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 $A \in \mathfrak{S}_n$ 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 they tell us about truth and logic. It is not my purpose here to prove a number of generalities about the theory, though the examples are chosen to exhibit general principles.

In most cases if a proposition in an example is false I will exhibit only one substage analysis to demonstrate that.

Example 1 *Ralph is a dog* *Ralph is a dog*

$\mathbf{P_0(a_0)}$ $\mathbf{P_0(a_0)}$

Even though we may not know the truth-value of either of these two propositions, still we know that in any given model they must have the same truth-value. Formally that is clear from our technical presentation above. But the reason they should have the same truth-value, and our formal model reflects that, is because the truth of either one does not depend on it itself satisfying some predicate. That is, if we do a Tarskian truth analysis of one of these, or of any other proposition of classical first-order logic, we are not led back to it again, but rather are led only to truth-conditions corresponding to its parts. That is why the inductive Tarskian analysis works, and why it fails to deal adequately with propositions such as 'This sentence is false'. Since any other proposition equiform to our examples will be analyzed identically with reference to the same objects and predicates, we know it will have the same truth-value as our examples. Hence we may identify these two propositions and any other proposition equiform to them; we may harmlessly speak of them as being the same proposition.

When, then, is 'Ralph is a dog' true? If you were to say that it is exactly when Ralph is a dog, you would be right. But you wouldn't be giving the truth-conditions for the proposition. Those are: Ralph is a dog, and it is consistent that Ralph is a dog and there is nothing arbitrary about that. It's just that for this proposition, as for any other which is not self-referential, the latter clause adds nothing. We know that if Ralph is a dog then it is consistent and not arbitrary that its material conditions hold. If you doubt that, then Tarski's theory of truth should be questionable to you.

Thus the definition of truth given in Principle 8 is not *ad hoc*, designed only for propositions with 'true' in them. The identification of equiform sentences of classical first-order logic (Principle 3) and the use of Tarski's analysis of truth for them (Principle 5) are justified. Moreoever, the abstraction that we make in doing classical first-order logic that the language is a completed infinite collection of sentence types is a reasonable simplification in light of Principles 3 and 5.

Example 2 *This sentence is false*

$$\begin{array}{c} \mathbf{\neg T(a)} \\ | \\ \mathbf{a} \end{array}$$

I will use the notation

$$\begin{array}{c} \mathrm{B} \\ | \\ \mathbf{b} \end{array}$$

to indicate that '**b**' is a name of the proposition B.

The example is false. Suppose it occurs at stage n > 0. Then let us begin by assuming it false and hence in $\mathfrak{F}_n(0)$. So at substage 1, **a**$\in\mathfrak{T}_n(1)$, as **a** asserts that **a** does not satisfy the predicate which interprets **T** and we are interpreting **T** as $\mathfrak{T}_n(0)$. That is, if we assume **a** to be false, then **a** is true. Similarly, **a**$\in\mathfrak{F}_n(2)$: if we assume **a** to be true, then it is false. And this vacillation continues. So it is inconsistent to take **a** to be true, and hence **a** is false (Principle 6). We have the same result if we begin our analysis by assuming **a** to be true.

Since **a** is false, its material conditions hold. Moreoever, they do not hold "after the fact", that is, after our analysis, but rather the moment that **a** is uttered. At that time, perhaps, we did not know that the material conditions of **a** held, and thus feel that we had recourse to our substage analysis because there is no "fact of the matter" as to **a**'s truth to begin with. But that way of considering things confuses the way we come to know that **a** is false, and explicate it, with whether **a** *is* false.

Our substage method allows us to conclude that **a** cannot be true. It does not, however, lead to a contradiction to take **a** to be false: that does not yield that **a** is true. The substage method merely investigates whether it is consistent to have the material conditions of **a** hold. For **a** to be true its material conditions must hold, but also it must be consistent and not arbitrary that they hold. For **a** to be false its material conditions may hold, too, if one of the other clauses of the truth-conditions for **a** fail. To summarize:

A is true $\Rightarrow$ 'A is true' is true

assuming 'A is true' to have been uttered subsequently. Similarly,

'A is true' is not true $\Rightarrow$ A is false,

but

'A is false' is not true $\not\Rightarrow$A is true

and

'A is false' is inconsistent $\not\Rightarrow$A is true

Principle 6 (Lemma 2) and Principle 1 (Lemma 1) yield the implications. The present example provides the counterexamples for the others. The next example deals with the implication: A is false $\Rightarrow$ 'A is false' is true.

Example 3	*This sentence is false*	*'This sentence is false' is false*
	$\neg$**T**(a)	$\neg$**T**(a)
	|	|
	a	**b**

The analysis of the previous example showed that **a** is false. Hence **b** is true, as it asserts that **a** is false. Formally, **b** is analyzed at a later stage than **a** by Pragmatic Assumption 4, as it is possible to analyze **a** without analyzing **b**, but not vice-versa. Since **a** is false it will be false at every stage and substage at which **b** is analyzed, hence **b** will be true. Here is a case where Pragmatic Assumption 4 matters, for if these propositions were analyzed simultaneously, then we would have that **b** is false, as it would vacillate between $\mathfrak{T}_n(k)$ and $\mathfrak{F}_n(k)$ in alternation with **a**. Yet **a** *is* false: no alternation of **b**'s truth-value in the substages arises once we recognize that fact (i.e., incorporate it into our model).

A similar argument should convince you that "'This sentence is false' is true', **T(a)**, is false, as is **¬¬T(a)**.

These formal versions provide an example of equiform sentences which, even if uttered at the same time (though analyzed sequentially), have opposite truth-values. To illustrate the point further I will draw on an example and explanation from the introduction to Hughes *1982*.

box **Z** ——→ $\boxed{\text{The sentence in box } \mathbf{Z} \text{ is false}}$

The sentence in box **Z** is false

The two sentences are equiform. But on the face of it these sentences are very different. One refers to itself, one does not. To test the truth of the left-hand one we need to consider whether it *itself* satisfies a predicate, here 'false', whereas that is not the case with the right-hand one. There is a clear and perceptible difference which is obscured by identifying them. That the difference *makes* a difference we have already seen, as the left-hand one is false, the right-hand one is true. Identifying equiform propositions leads to paradoxes:

> Call the sentence in box **Z** by the name **c**. We know that
>
> (1) **c** is false.
>
> That is,
>
> (2) The sentence in box **Z** is false.
>
> So by the principle on truth entailment (Principle 6, Lemma 2),
>
> (3) 'The sentence in box **Z** is false' is true.
>
> Hence,
>
> (4) **c** is true.

The fallacy enters here when we identify **c** with the sentence at (2). The step from (2) to (3) is justified, but only if what is named by the quoted words at (3) is the sentence at (2). The step from (3) to (4) is justified only if what is named by the quoted words at (3) is **c**. But (2) and **c** are not the same: (2) is true and **c** is false.

The use of quotation names can lead to confusions: it is a convention designed for sentence types and is inherently ambiguous with respect to equiform sentence tokens. I have only used quotation names in these examples where there is little chance of misunderstanding and the examples are more easily understood in that form. I eschew quotation names entirely in the formal model.

The present example also illustrates how to form the contradictory of a sentence. Given a sentence **c** we know that if we form *another* proposition **¬T(c)** it will have the opposite truth-value to that of **c**, even if it is equiform to **c**. On the other hand, simply negating a false proposition may yield one which is also false: **T(a)** and **¬¬T(a)** are both false as we observed above; they are only apparently contradictory. We cannot, though, have two apparent contradictories both true, as I shall discuss in Example 4.

So if you want to say that the proposition over there is wrong, not true, you will have to say that. Don't say 'not A', at least if A has 'true' in it, or has names which via some chain of references lead to a proposition with 'true' in it, for 'not' negates the material conditions of a proposition. Only if the material conditions suffice for the consistency conditions does the use of 'not' have the same effect as 'it's not true that' which negates all the truth-conditions.

Much of our discussions can be recast to apply equally to abstract propositions, that is, to those "things" which some philosophers believe are the bearers of truth and yet are not perceptible to our senses. In those terms I would say that the two sentences of our example represent (or perhaps mean) different abstract propositions. Just because two sentence tokens are equiform, even in a formal language, need not entail that they must represent the same abstract proposition. Strict criteria for when two sentence tokens mean the same abstract proposition must be laid out, with reasons why. To say that it is convenient to take equiform sentences to mean the same abstract proposition is not sufficient, especially as we have seen that doing so leads to paradoxes. The alternative, to expunge 'true' from our object language, is hardly convenient and seems *ad hoc*. Besides, if abstract propositions are to be defended as anything more than (inconvenient) fictions then their nature and the nature of our language must determine when two sentences refer to the same abstract proposition. Since the relationship between abstract propositions and our language is so unclear, it seems wise to me to deal with what we have: sentence tokens, specific linguistic entities (Principle 1).[6]

It may be possible to recast much of this theory for sentence types, but each sentence type would have to be fully indexed. Any aspect of sentence tokens which can be predicated of one and not another within a sentence type could, I believe, be used to reformulate the paradoxes. The only way I can see to ensure that there is no (logically) discernible difference among sentence tokens in a type is if there is only one token in each.

Still, you may wish to argue that the questions: What is a sentence? What constitutes a use of a sentence? When has one been used assertively or even put forward for discussion? are not questions allowing unambigious answers. You might argue that these questions can be avoided by taking things inflexible, rigid, timeless as propositions. But that only pushes back these same problems to: How do we use logic? What is the relation of these formal theories of mathematical symbols to our arguments, discussions, and search for truth? How can we tell if this utterance is an instance of that abstract proposition? It's not that taking sentence tokens as propositions (Principle 1) brings up questions which can be avoided. The emphasis on formalism and mathematics in logic has gone too far if we cannot relate our work to its intended use, its original motivation. Logic is not a formal mathematical theory; it is, at its best, formal theories used to explain difficult problems about how to reason correctly.

I would suggest that the answer to the question of how to characterize the use of a sentence is simple, though not rigid. We must have agreement to do logic, we must

---

[6] Gupta *1982*, 4 argues that we need abstract propositions because there are unexpressed truths just as there are unnamed objects. But this is question begging. I know that there are unnamed objects because I can see and feel them; I do not know any unexpressed truths until they are expressed (barring mysticism from the scope of logic).

speak the same language, be able and willing to follow each other's arguments when uttered or written. As well, we must agree in any particular instance whether a sentence which has a truth-value has been put forward. If we cannot agree on certain cases then we cannot reason together about them. That does not mean that we then do different logics, or that logic is psychological; it means only that we differ on borderline cases. If you like, you may say that I take as a primitive notion the utterances of sentences which may be assigned truth-values. That notion may be further explicated, and has been done so by others (Buridan does so by reference to mental propositions). But agreement by cases, or in general as with a first-order language, is what logic needs, and all that it needs.

In Example 11 I will briefly return to the question of how agreements are needed in logic. In the first and last chapters of Epstein *1990*, I discuss this question at length, explaining that by 'agreement' I do not mean only active, explicit agreements, but also implicit ones, what may be said to be our common background. Most of our agreements are implicit, and not necessarily freely made: whether for metaphysical, psychological, physical, or pragmatic reasons we cannot in general determine.

Example 4 *This sentence is true*

$$\begin{array}{c}\mathbf{T(a)}\\ | \\ \mathbf{a}\end{array}$$

The example is false. At whatever stage **a** is analyzed we may begin by assuming it false. Then at all further substages it will be false, too. Hence, it is false.[7]

You might argue that we could just as well have assumed that **a** is true, and then it would have been true. But is it in our power to make a proposition true simply by calling it so? Relative to our linguistic conventions and the facts of the matter, there should be nothing arbitrary about a proposition being true.

Yet neither is **a** false by an arbitrary choice: **a** is false because it fails to be true. The false propositions are those which are not true: those whose material conditions fail, or which are inherently contradictory, or which are arbitrary. The odd, anomalous, paradoxical sentences, ones which we cannot proceed on as a basis for good reasoning, are all false: *tertium non datur*. If a proposition offend thee, cast it out.

I grant that one could set up a theory exactly like the one here except requiring that it not be arbitrary that A is false, calling a proposition true if it is not false. Formally, one could define $A \in \mathfrak{F}_n(\alpha_i)$ if it is in $\mathfrak{F}_n(k)$ for all large k,

$$\mathfrak{F}_n = \mathfrak{F}_{n-1} \cup (\cap_i \mathfrak{F}_n(\alpha_i)), \quad \text{and then} \quad \mathfrak{T}_n = \mathfrak{T}_{n-1} \cup \{A \in \mathfrak{S}_n : \mathfrak{A} \notin \mathfrak{F}_n\}.$$

Thus truth would be taken as the default truth-value. Perhaps one could justify that generous notion of truth, but I have argued in Epstein *1990* that, both mathematically and philosophically, modern logics such as intuitionistic logic, the many-valued logics of Łukasiewicz and of Kleene, modal logics, and many others

---

[7] There is evidence in Sophism 10 of Buridan, as explained in the commentary by Hughes, that Buridan would have agreed that the example is false.

can and should be understood as taking falsity as the default truth-value. The only logics that I know that might be understood as taking truth as the default truth-value are paraconsistent logics in which a non-trivial theory may contain both a proposition and its negation. In Epstein *1990* I present an example of a paraconsistent logic in which it is possible for apparent contradictories both to be true, but not both false.

We may still wish to classify sentences as to why they are false. That could result in a classification system similar to Kripke's or Gupta's.

Example 5 *Plato says 'What Socrates is saying is false' and Socrates says 'What Plato is saying is false' and that is all that either says.*

$$\begin{tikzcd}
\neg \mathbf{T}(\mathbf{b}) \arrow[d, dash] & \neg \mathbf{T}(\mathbf{a}) \arrow[d, dash] \\
\mathbf{a} & \mathbf{b}
\end{tikzcd}$$

Both propositions are false. If they are analyzed at stage n, then let us begin by assuming them both false. So at the first substage they are both (hypothetically) true, and hence at the next substage false, etc.

Kripke has argued that this kind of example cannot be handled by Tarski's theory, since there is no reasonable choice to be made as to which of these propositions is to be in the object language and which in the metalanguage. Yet they must be in separate levels for Tarski's analysis. Kripke's analysis classifies these are paradoxical, without truth-value.

The colloquial version of this example is Sophism 8 of Buridan. Buridan analyzes these by what Hughes calls the principle of parity of reasoning: **a** and **b** must have the same truth-value, for any analysis of the one must apply equally to the other, as there is no logical basis for distinguishing them. Another way of thinking of Buridan's analysis, however, is that were we to say that one of the propositions of the example is true and the other is false, which it seems would be consistent, then the choice of which one is true is completely arbitrary. Principle 7, that truth is not arbitrary, then deals with this case. I believe it is a good interpretation of Buridan's views.

Example 6 *Plato says 'What I am saying is true, and what Socrates is saying is true if and only if what Socrates is saying is true', and Socrates says 'What Plato is saying is false'.*

$$\begin{tikzcd}
\mathbf{T}(\mathbf{a}) \wedge (\mathbf{T}(\mathbf{b}) \longleftrightarrow \mathbf{T}(\mathbf{b})) \arrow[d, dash] & \neg \mathbf{T}(\mathbf{a}) \arrow[d, dash] \\
\mathbf{a} & \mathbf{b}
\end{tikzcd}$$

The proposition **a** is the truth-teller in a new guise, but it must be analyzed simultaneously with **b**.

Let us begin by assuming that **a** and **b** are both false. Then at substage 1 **a** is false and **b** is true, and at every subsequent substage **a** is false and **b** is true. So **a** is false. But we may not therefore conclude that **b** is true, for the analysis of **b** is not independent of that of **a**. If we begin an analysis by assuming both **a** and **b** to be true, then at all subsequent substages **a** is true and **b** is false. So **b** is not necessarily true. Therefore, **b** is false.[8]

Example 7 *Plato says 'What Socrates is saying is false', and Socrates says 'What Critias is saying is false', and Critias says 'What Plato is saying is false or God exists', and that is all that any of them say.*

$$\begin{array}{ccc} \neg \mathbf{T(b)} & \neg \mathbf{T(c)} & \neg \mathbf{T(a)} \vee \exists \mathbf{xG(x)} \\ | & | & | \\ \mathbf{a} & \mathbf{b} & \mathbf{c} \end{array}$$

Suppose we analyze these at stage n. At substage 0 we may begin by assuming them false. So at substage 1 we must conclude that each is true. At this point we have two cases depending on whether God exists (i.e., is in your model). Let's first suppose that you're an atheist. Then since the material condition of the second disjunct is never satisfied, **c** will be validated at a substage if and only if **a** is not true. So at any substage **a**, **b**, and **c** will all have the same (hypothetical) truth-value, which will alternate from substage to substage. So they are all false.

The analysis is more interesting if God exists. Regardless of what we take for $\mathfrak{T}_n(0)$, at all substages $k \geq 1$, $\mathbf{c} \in \mathfrak{T}_n(k)$. So at all substages $k \geq 2$, $\mathbf{b} \in \mathfrak{F}_n(k)$, and therefore at all substages $k \geq 3$, $\mathbf{a} \in \mathfrak{T}_n(k)$, though $\mathbf{a} \in \mathfrak{F}_n(2)$. So **a** and **c** are true while **b** is false. Here we have an example of a proposition, **a**, which is true but does not settle down to appear so until the third substage.

Note that **b** is false regardless of whether God exists, so **b** is necessarily false, though **a** and **c** are (formally) contingent. Buridan, on the other hand, would have used **c** as an example of a necessary truth.

For any m we can find a proposition which is true but in some substage analysis does not settle down until at least stage m. Let $\mathfrak{S}_n$ be $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$ where for $i = 1, \ldots, m - 1$ $\mathbf{a}_i$ is $\neg T(\mathbf{a}_{i+1})$ and $\mathbf{a}_m$ is a proposition in $\mathbf{L}^*$ whose material conditions are validated by $\mathfrak{M}$. If we first assume each $\mathbf{a}_i$ to be false, then $\mathbf{a}_1$ has hypothetical truth-value 'false' at substage $m - 1$, but is assigned 'true' at every subsequent stage. In no other substage analysis does $\mathbf{a}_1$ settle down later than stage m.

The next four examples deal with questions of possibility, necessity, and validity, as well as the temporal nature of propositions.

Example 8 *No proposition is negative*

$$\neg \exists \mathbf{x}(\mathbf{P(x)} \wedge \mathbf{N(x)})$$

This sentence is false no matter when uttered, as it itself must satisfy the predicate 'negative'. Hence it is impossible that it be true: it is necessarily false.

---

[8] In the original version of the formal theory presented here I used only one substage analysis for $\mathfrak{S}_n$ in which every proposition in $\mathfrak{S}_n$ was initially assumed to be false. My reasoning was that for A to be true its truth must be forced on us even if we assume the contrary. Principle 7 was then that 'A is not true' is inconsistent. Using that analysis every example in this paper comes out the same except for the present one, in which we would have that a is false and b is true. This may seem more intuitive, but only at the cost of ignoring that a and b *must be analyzed simultaneously*. To use only the one substage analysis now seems to me to involve an arbitrary assumption about the possible truth-values of the propositions under discussion.

This should not be confused with the fact that "things may be as the proposition asserts them to be", that is, the material conditions for the proposition may hold. (As Buridan suggests, God may annihilate all negative propositions; the modern version, alas, is that the world and everything on it may be destroyed by nuclear war.) But then the proposition itself could not exist.

We must distinguish between a proposition being *possibly true*, which this example is not, and a proposition being *possible*, which the example is. A proposition is possible if there is a possible world in which its material truth-conditions are satisfied. A proposition is possibly true if there is a possible world in which it is true, and hence exists. We have that **a** is true entails that **a** exists.

An inference is valid if it is not possible that the material truth-conditions of the premises hold and those of the consequent fail. It is wrong to define validity of an inference as: it is impossible for the premises to be true without the conclusion being true. That would invalidate, e.g., 'Every proposition is affirmative. Therefore: no proposition is negative' for the conclusion is never true, because for it to be true it must exist. See Buridan's Sophisms 1 and 2 for a fuller discussion.[9]

But, concerning the temporal nature of propositions, what if we want to analyze some proposition of Aristotle for the first time? Don't we say that it's true or false? Yes, but I would argue only in the same sense in which we assume that Socrates *is* a man when we say that 'Socrates is a man' is true. In general, we analyze propositions which we believe to have the same truth-conditions as Aristotle's: those which are equiform to his purported utterances or inscriptions, or are translations of those. So long as these involve no self-reference it is harmless to make the claim that they *are* Aristotle's, hiding by this that we are making an identification.

Example 9 *Plato was Athenian entails 'Plato was Athenian' is true.*

The ambiguity of quotation names requires that this example be stipulated more clearly. I shall examine the formal version:

$$\mathbf{A}(\underset{\substack{| \\ \mathbf{a}}}{\mathbf{p}}) \text{ entails } \mathbf{T(a)}$$

This is false: for Plato was Athenian, yet it is possible for **a** not to exist. For a sentence to be true it must exist; for the material conditions for **T(a)** to hold, **a** must exist. Something has to have been uttered.

What, then, about the principle of truth entailment? This says that if A is true, then 'A is true' is true. Or in the notation of a rule:

$$\frac{\underset{\substack{| \\ \mathbf{a}}}{\mathrm{A}}}{\mathbf{T(a)}}$$

---

[9] Prior *1968* also gives a good presentation of Buridan's ideas on this, though much else in that paper seems confused (see also Scott *1966*, 28, n. 58 and p. 56, n. 98).

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":

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

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:

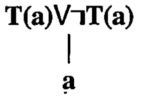
$$\forall x(T(x) \vee \neg T(x)) \quad \text{or} \quad A\forall x(P(x) \rightarrow [T(x) \vee \neg T(x)])$$

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 $\bigcup_{m \leq n} \mathfrak{S}_n$. 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.

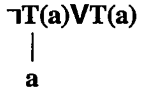
In Kripke's theory the example is not true due to his use of a 3-valued logic to generate his classification scheme. This leads to a problem with what could be called the strengthened liar paradox. Suppose **a** is paradoxical, that is, **a** has undefined truth-value in the minimal fixed point model. Then clearly **a** is not true, which is further reflected in that **T(a)** has undefined truth-value in that model, too. But then so also does **¬T(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*



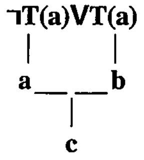
As a substitution instance of the previous example (a concept which can be made precise for sentence tokens just as easily as sentence types, since we are assuming that words are types), this must be true. And indeed it is: at whatever stage it is uttered, whether it is first assumed to be false or true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



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:

$\mathbf{a}\in\mathfrak{F}_n(0)$ and $\mathbf{b}\in\mathfrak{F}_n(0)$,
so $\mathbf{a}\in\mathfrak{T}_n(1)$ and $\mathbf{b}\in\mathfrak{F}_n(1)$ and $\mathbf{c}\in\mathfrak{T}_n(1)$,
so $\mathbf{a}\in\mathfrak{F}_n(2)$ and $\mathbf{b}\in\mathfrak{T}_n(2)$ and $\mathbf{c}\in\mathfrak{T}_n(2)$,

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:

$$\begin{array}{lll} \textbf{d}\text{ is true} & \text{iff} & \text{the material conditions for } \textbf{d} \text{ hold} \\ & & \text{and } (*) \left\{ \begin{array}{l} \text{it is consistent that they hold and} \\ \text{there is nothing arbitrary about that} \end{array} \right. \\ & \text{iff} & \text{(the material conditions for the first disjunct of } \textbf{d} \text{ hold) or (the} \\ & & \text{material conditions for the second disjunct of } \textbf{d} \text{ hold) and } (*) \\ & \text{iff} & \text{(snow is white or grass is green) and (that is consistent and there} \\ & & \text{is nothing arbitrary about that)} \end{array}$$

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 $\mathbf{F}(\mathbf{x}) \equiv_{\mathrm{Def}} \mathbf{P}(\mathbf{x}) \wedge \neg \mathbf{T}(\mathbf{x})$ and **F** can be interpreted as 'is false'. Note then that $\forall \mathbf{x}(\mathbf{T}(\mathbf{x}) \vee \mathbf{F}(\mathbf{x}))$ 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.

Example 15 *No proposition has the word 'Socrates' in it.*
If no proposition has the word 'Socrates' in it,
then Socrates is not Socrates.
Therefore,
Socrates is not Socrates.

$$\neg\exists x(P(x) \wedge S(x))$$
$$\neg\exists x(P(x) \wedge S(x)) \rightarrow s \neq s$$

Therefore, $s \neq s$.

The inference is invalid. We may imagine a situation in which the first proposition is true (as described in Example 8, with the usual convention on quotation names of words). If the second is then uttered, it is true because the material conditions for its antecedent fail. Yet the conclusion is false.

Nonetheless, this is not a counterexample to the rule of *modus ponens*. Though the first proposition is equiform with the antecedent of the second, the material conditions of the one may be satisfied, while those of the other are never satisfied.

The correct statement of *modus ponens* for a language which contains its own truth-predicate is, I think:

> From a proposition A and a proposition of the form C → D where C is equiform with A, conclude a proposition B equiform with D if the material conditions of A are identical with those of C and the material conditions of D are identical with those of B.

If A is self-referential, even indirectly, then the material conditions of A and the antecedent of C → D cannot be the same, though it may be that for every possible world they both hold or they both fail; similarly for B and the consequent of C → D. So in effect this restricts *modus ponens* to non-self-referential propositions, which we can be sure include at least those in the language L*. But if we restrict the rule to L*, we might as well state it in its simplified form $\frac{A, A \rightarrow B}{B}$ since, as argued in Example 1, the material conditions of A and the antecedent will be the same, as will those of B and the consequent.

It is not the case that *modus ponens* fails: to use that rule we have always been under an obligation to check that A and the antecedent of A→B have the same truth-conditions.

In the next two examples we make the abstraction that $\mathfrak{S}_n$ can be infinite.

Example 16 We assume that we have a sequence of sentences all uttered at the same time each of which asserts that the next one in the sequence is true.

$$\mathfrak{S}_n = \{\underset{\substack{| \\ a_0}}{T(a_1)}, \underset{\substack{| \\ a_1}}{T(a_2)}, \underset{\substack{| \\ a_2}}{T(a_3)}, \ldots\}$$

If we begin by assuming that each a_i is false, then at all further substages each a_i is false. Hence each a_i is false.

Example 17 We assume that we have a sequence of sentences all uttered at the same time which alternately assert that the one following it is true, or that the one following it is false.

$$\begin{array}{ccccccc} \mathfrak{S}_n = \{ & \mathbf{T(a_1)}, & \neg\mathbf{T(a_2)}, & \mathbf{T(a_3)}, & \neg\mathbf{T(a_4)}, & \mathbf{T(a_5)}\ , & \ldots\ \} \\ & | & | & | & | & | & \\ & \mathbf{a_0} & \mathbf{a_1} & \mathbf{a_2} & \mathbf{a_3} & \mathbf{a_4} & \end{array}$$

If we begin by assuming that each of these is false, then:

at substage 1: each $\mathbf{a}_{2i}$ is false; each $\mathbf{a}_{2i\,+\,1}$ is true,
at substage 2: each $\mathbf{a}_{2i}$ is true; each $\mathbf{a}_{2i\,+\,1}$ is true,
at substage 3: each $\mathbf{a}_{2i}$ is true; each $\mathbf{a}_{2i\,+\,1}$ is false,
at substage 4: each $\mathbf{a}_{2i}$ is false; each $\mathbf{a}_{2i\,+\,1}$ is false.

Substage 4 is the same as substage 0, so the cycle begins again. Hence no $\mathbf{a}_i$ is in $\mathfrak{T}_n(k)$ for all large k, and thus each is false.

Note that were $\mathfrak{S}_n$ finite, say with $\mathbf{a}_3$ or any $\mathbf{a}_{2i\,+\,1}$ being $\neg\mathbf{T}(\mathbf{a}_0)$, the same analysis would apply.

It is hard for me to say whether these last two examples adequately reflect intuitions about truth. When propositions are sufficiently simple and regular that we can identify equiform propositions and abstract to sentence types, then it is a harmless idealization to assume that we have a completed language with infinitely many propositions all of which are "uttered" simultaneously. However, if it really matters when propositions are uttered, as it does when self-reference is possible in a language, the idealization that $\mathfrak{S}_n$ could be infinite no longer seems so harmless. Oddities which arise seem as likely to me to be because we have assumed that infinitely many propositions are asserted all at once as due to the definition of truth in the theory.

5. Toward one language for logic

In the writings of Buridan there is no distinction made between logic and metalogic; one language and system of reasoning is to suffice. Nor did Russell and Whitehead employ a distinction between logic and metalogic; their goal was to create one precise perspicuous language for reasoning in which logical relations among propositions would be evident.

The self-referential paradoxes seemed to make that goal unattainable. Frege's system was inconsistent, and Russell and Whitehead had recourse to a stratified language. Since Tarski's analysis of truth, most modern logicians have taken metalogic to be distinct from logic.

A satisfactory analysis of self-referential paradoxes in a language with its own truth-predicate might suggest that a re-incorporation of logic and metalogic is possible. The next examples illustrate difficulties and directions for further research in the pursuit of that goal.

Example 18 *No sentence equiform with this sentence is true.*

$$\begin{array}{c}\neg\exists x(E(x,a)\wedge T(a))\\ | \\ \mathbf{a}\end{array}$$

The proposition is false, for to assume it true would lead to an inconsistency. Moreover, every proposition equiform with **a** is also false. But then how are we to say that all those propositions are indeed false?

There are various ways, avoiding equiformity. For example, the following is true:

$$\begin{array}{c}\neg\exists x(T(a)\wedge E(x,a))\\ | \\ \mathbf{c}\end{array}$$

But this is odd, for in classical logic are not **a** and **c** equivalent? It is not because we have abandoned classical logic that they are not equivalent here, but because the use of tokens limits the conclusions we can draw from rules such as *modus ponens*, as discussed in Example 15. The exact form of propositions matters more than we might previously expect when we begin allowing for self-reference. Synonymity is more difficult to establish, as noted already in Example 3.

George Hughes (private communication) thinks that Buridan would not be prepared to claim that all sentences equiform with this (informal) example are false. 'At least in the *Sophismata*, Chapter VI, Sophism 7, he says that although I can be sure that *this* proposition 'Homo est asinus', which I now form, is false, since I know the meaning it has when I state it, yet I do not know whether an arbitrary token equiform with it is false, since such a token might be true in some other language with which I am not familiar'. So Hughes suggests the following example:

Example 19 *No sentence with the same meaning as this sentence is true.*

Any worry about the truth or falsity of this proposition (if proposition it be) awaits a clearer understanding of the word 'meaning' than we now have. To the extent that meaning and synonymity are related, this semantic analysis of truth has illustrated the need for sharper distinctions. Perhaps, though, consideration of the truth-conditions or even the material conditions of propositions might suffice.

Example 20 *The material conditions of this sentence do not hold.*

Informally, the proposition is false, since to assume it true would lead to an inconsistency.

But then do the material conditions of the example hold? This depends in part on how we are to extend the formalization of our principles when more than one new predicate is added to a usual first-order language and model. Technical difficulties arise in the substage analysis, and in this particular case we need to be clear about exactly what part of the analysis is to be identified with the material conditions of the proposition being considered. In summary, I do not view this proposition as illustrating a flaw or strengthened liar paradox, but as a challenge to extend our methods to incorporate more semantic predicates than 'is true'.

Example 21 *The collection of all collections each of which does not belong to itself belongs to itself.*

Let $\mathbf{z} = \{\mathbf{x}\colon \mathbf{x} \notin \mathbf{x}\}$. Then:

$$\begin{array}{c} \mathbf{z} \in \mathbf{z} \\ | \\ \mathbf{a} \end{array}$$

One particular predicate which can lead to paradoxes in the presence of self-reference is the set-membership relation.

The semantic principles of §2 classify the proposition as false, for to assume it true would lead to a contradiction. And, similarly, the apparent contradictory $\mathbf{z} \notin \mathbf{z}$ is also false; the real contradictory is $\neg\mathbf{T}(\mathbf{a})$. But what conclusions should we draw from this about formalizing the extension of this predicate?

Set theory can also be formulated as part of second-order logic. What further principles, if any, would be needed to analyze quantification over predicates in the presence of self-reference?

Example 22 *This sentence is not provable.*

A formalization is needed of the notion of proof corresponding to the semantics presented here. Some of the problems associated with that have been raised in Examples 9, 15, and 18. Given such a formal notion of proof, if we specify the order in which sentences of the language are to be generated we could Gödel number them. What, then, would be the status of Gödel's incompleteness theorems?

Example 23 *The name of this sentence is 'Rudolfo'.*

The role of naming in reasoning needs to be further clarified. In the theory presented here, the naming of propositions is essential to logic but is not part of the theory itself. Consider

$$\begin{array}{c} \neg\mathbf{T}(\mathbf{a}) \\ | \\ \mathbf{a} \end{array}$$

The part above the vertical line is the proposition. The vertical line and the letter below it represent the act of pointing to the proposition and uttering the name '**a**' (a type), or the act of sticking a label onto the proposition. It cannot be that the entire complex of signs is a proposition which says '**a** is the name of $\neg\mathbf{T}(\mathbf{a})$' for then where would the proposition be that '**a**' names? It cannot be *part* of this latter proposition, for a part of a proposition is not a proposition (Example 14).

It seems to me we have several choices in how we reason about such a complex:

(i) We could argue that a part of a proposition can be a proposition. But I do not see how to reconcile that with our original principles, as I explained in Example 14.

(ii) Labeling is an incomplete process: we implicitly assume a description of what it is we are pointing to when we name. By making this explicit we could restrict ourselves to using only descriptive phrases as names, for example, $\neg\mathbf{T}$(the single sentence in $\mathfrak{S}_{47}$). Such phrases can be incorporated into the logic.

(iii) Naming is metalogical and cannot be incorporated within our logic. Naming is pointing; pointing is prior to but not part of language. We can of course write **cNd** for 'that which is named by **c** names **d**' where, e.g., **c** names '**a**' and **d** names **a** above. If we want to express that **c** and **d** name the same thing we can write **c** = **d**. But we cannot assert that '**b**' names the proposition: $\neg$**T(b)**.

Even if one language were sufficient for logic, is naming not an act essential to logic but not part of it?

Acknowledgements

I began reading Hughes's translation of Buridan with Benson Mates in January 1983, and the discussions we had then were the impetus for me to write this paper. His encouragement and critical comments have been valuable. It has been good working with a skeptic.

In the summer of 1983 I was struggling with the formalization of Buridan's ideas. Peter Eggenberger gave me much of his time then and throughout this project. I feel that this paper would not have been written were it not for his interest and insights. Example 9 is directly from our conversations.

The comments of the participants of a seminar at the University of California, Berkeley in March 1984, were important in revising the presentation of this work; the discussion of naming in Example 23 are due to Walter Carnielli's queries.

I am grateful to several people who read the first draft of this paper and suggested useful revisions, particularly Newton Da Costa, John Corcoran, and anonymous referees. Example 6 was contributed by Howard Blair.

Bibliography

Angelelli, Ignacio *1985* [Review of Hughes *1982*], *The Journal of symbolic logic*, **50**, 859–860.

Burge, Tyler *1979* 'Semantical paradox', *Journal of philosophy*, **76**, 169–198.

Burgess, John P. *1986* 'The truth is never simple', *The journal of symbolic logic*, **51**, 663–681.

Epstein, Richard L. *The semantic foundations of logic.*

—— *1990 Volume 1: Propositional logics*, Dordrecht (Kluwer).

—— *1993 Volume 2: The basic assumptions of predicate logic*, Typescript, to appear.

Faral, Edmond *1949* 'Jean Buridan: maître ès arts de l'Université de Paris', in *Histoire littéraire de France*, **28**, pt. 2, 462–605.

Gupta, Anil *1982* 'Truth and paradox', *Journal of philosophical logic*, **11**, 1–60.

Hellman, Geoffrey *1985* [Review of papers by Martin and Woodruff, Kripke, Gupta, and Herzberger], *The journal of symbolic logic*, **50**, 1068–1071.

Herzberger, Hans *1975* 'Dimensions of truth', in *Contemporary research in philosophical logic and linguistic semantics* (ed. Hockney *et alii*), Dordrecht (Reidel), 71–92.

Hughes, G. E. *1982 John Buridan on self-reference*, Cambridge (University Press).

Jaskowski, Stanisław *1934* 'On the rules of suppositions in formal logic', *Studia logica*, **1**, 5–32. [Reprinted in *Polish logic, 1920–1939*, (ed. Storrs McCall), Oxford (Clarendon Press), 1967.]

Kripke, Saul *1975* 'Outline of a theory of truth', *Journal of philosophy*, **82**, 690–716.

Markov, A. A. *1954* 'The theory of algorithms', *Trudi. Mat. Inst. Steklov*, **42**. [English translation: Israel Program for Scientific Translations Ltd., Jerusalem, 1971.]

Prior, A. N. *1968* 'The possibly true and the possible', *Mind*, n.s. **78**, 481–492.

Scott, Theodore Kermit *1966 John Buridan: sophisms on meaning and truth* (Appleton-Century-Crofts).

Simmons, Keith *1987* 'On a medieval solution to the liar paradox', *History and philosophy of logic*, **8**, 121–140.

Tarski, A. *1934* 'The concept of truth in formalized languages', in his *Logic, semantics, metamathematics*, second ed., Indianapolis (Hackett), 1983, 152–278.

Yablo, Stephen *1985* 'Truth and reflection', *Journal of philosopical logic*, **14**, 297–349.