Notre Dame Journal of Formal Logic
Volume XIX, Number 4, October 1978
NDJFAM

JOHN BURIDAN ON THE LIAR: A STUDY AND RECONSTRUCTION

PAUL VINCENT SPADE

This paper is a partial study of the position taken by the 14th century logician and philosopher John Buridan (see [4]) on the Liar and related paradoxes—the "insolubles" (*insolubilia*) as they were called then. Buridan's position is most extensively set out in the eighth chapter of his Sophismata ([2], translated in [3]. There is also a discussion in [1].). Although at least three brief treatments have appeared in English ([3], introduction; [7], [9]), and although the *Sophismata* has recently been translated into English (in [3]), no study has yet examined Buridan's view in depth. This paper attempts a partial rectification of that situation.

We shall begin by setting out some preliminary notions and notation. Then we shall cite some passages from Buridan that set out his position. On the basis of these passages, we shall extract several theses which may be regarded as "principles" of his approach. Our reconstruction of Buridan's position will have to conform as much as possible to these theses. I said that this paper is a "partial" study of Buridan's position. Our overall policy will be to reconstruct that position within a very simple framework, allowing only the most straightforward kinds of paradox. Accordingly, we shall set out a very limited syntax **SYN**—containing only singular terms, two predicates, and a negation operator—and a very restricted set of models, which allow only the simplest and most direct kind of vicious self-reference. Insolubility will then be defined for this context. Such a narrow approach permits us to abstract from certain kinds of cases that Buridan simply does not discuss, and to put off to another occasion a treatment of some of the more complicated cases he does discuss. An initial obstacle is the fact that the Buridanian theses to be extracted from his text will turn out to be inconsistent even within this simple framework. Some compromise must be made if our study is to get off the ground. We shall discuss some possible ways of making this compromise, and choose from among them. Our choice will involve giving up the rules of double negation in certain cases. It will also involve adopting the principle that the

Received April 22, 1974

correspondence theory of truth fails exactly in the cases of the insolubles. Once all this has been done, we shall set up a language $\mathcal{L}_1$ within the simple context provided by **SYN** and $\mathcal{L}_1$-models. $\mathcal{L}_1$ will satisfy all the Buridanian theses extracted from his text, except for the rules of double negation, which we resolved to sacrifice. We shall look at $\mathcal{L}_1$ in some detail.

1 For simplicity, we shall not consider languages with polyadic predicates. If ϕ is a primitive monadic predicate of some language $\mathcal{L}$, and if $V_{\mathfrak{M}}$ is a valuation function assigning a truth value 0, . . ., 1 to the sentences of $\mathcal{L}$ with respect to a model $\mathfrak{M}$, and if $f_{\mathfrak{M}}$ is the denotation function for $\mathfrak{M}$, assigning denotations to the singular terms x of $\mathcal{L}$, then we shall say by definition that $\mathsf{EXT}_{\mathfrak{M}}(\phi) = \{f_{\mathfrak{M}}(x): V_{\mathfrak{M}}(\phi x) = 1\}$, and we shall call $\mathsf{EXT}_{\mathfrak{M}}(\phi)$ the "extension of ϕ with respect to $\mathfrak{M}$". This definition will be extended later to cover predicates formed by prefixing a string of negations to a primitive predicate ϕ. The definition demands that everything contained in the extension of any predicate of $\mathcal{L}$ have a name. We accept this condition for the purposes of our discussion. (A further consequence of the definition is that, where T is a truth-predicate, $\mathsf{EXT}_{\mathfrak{M}}(T)$ is defined *not* as the class of true sentences, but rather as the class of sentences with true truth-sentences. In virtue of Theorem 3 below, however, these two classes are identical.)

2 Buridan's suppositional truth-rules ([3], pp. 90-93), with which the reader is assumed to be familiar, provide a reasonably good account of the correspondence theory of truth according to which a sentence is true iff it corresponds to reality, to the facts (see [13], §3). We shall say that a sentence *corresponds* (to reality, to the facts) iff the world is as the sentence "signifies according to formal significations" ([3], p. 195). That is, a sentence corresponds iff the *normal* suppositional truth-conditions for such a sentence are met ([3], pp. 90-93). If the sentence is an insoluble one, then that condition is not in fact its actual truth-condition, as we shall see. But that is of no account here.

We shall be discussing only atomic sentences and sentences formed from atomic sentences by negation. Within this context we can say that an atomic sentence corresponds in a model $\mathfrak{M}$ iff its subject and predicate "stand for the same" ([3], pp. 90-93, conclusions 9 and 10)—that is, iff the denotation of its subject term is contained in the extension of its predicate term in $\mathfrak{M}$. The negation $n(s)$ of a sentence s corresponds in $\mathfrak{M}$ iff s itself does not correspond in $\mathfrak{M}$. Thus, if ϕ is a primitive predicate, then by definition $\phi_x \in \mathsf{CORR}_{\mathfrak{M}}$ iff $f_{\mathfrak{M}}(x) \in \mathsf{EXT}_{\mathfrak{M}}(\phi)$, and $n(s) \in \mathsf{CORR}_{\mathfrak{M}}$ iff $s \notin \mathsf{CORR}_{\mathfrak{M}}$. The correspondence theory of truth may now be expressed as follows: A sentence is true in $\mathfrak{M}$ iff it corresponds in $\mathfrak{M}$: $V_{\mathfrak{M}}(s) = 1$ iff $s \in \mathsf{CORR}_{\mathfrak{M}}$.

3 In order to provide a framework for the discussion to follow, I want here to set out eight "theses", which, I shall argue, are in one way or another theses of Buridan's position. I say "in one way or another", because they do not all have the same status. Some of them Buridan explicitly holds; some can be inferred from things he explicitly holds. Some, finally, seem to be only tacitly accepted. The status of each will be discussed briefly in turn. For a fuller discussion, see [11]. The eight are these:

(1) *"From every proposition, together with the condition that it exists, there follows the conclusion that it is true."* ([3], p. 196). This is explicitly held by Buridan. The context is this ([3], pp. 195f.):

> Thus, it is otherwise said, nearer the truth, that every proposition virtually implies another proposition, so that of the subject standing for it, there is affirmed this predicate "true". I say it implies virtually just as an antecedent implies that which follows from it. Thus, any proposition is not true, if in this consequent affirmation, the subject and predicate do not stand for the same.

Buridan agrees with this view, but adds the proviso that the "proposition" must exist. For him, a sentence (what he calls a "propositio") is a sentence-*token*, an individual utterance or inscription. As such, its existence is a contingent affair. Utterances die away, and inscriptions can be erased. For Buridan, it is only *existing* sentences (tokens) that can be said to have a truth-value. (See [3], p. 17.) For simplicity, we shall confine ourselves in this paper by *fiat* to contexts in which the existential condition is satisfied for all sentences. (Compare [3], p. 57.)

Let us call (1) the "virtual implication principle" (VIP). It, or a version of it, was a characteristic of a long line of mediaeval views on the paradoxes. (See [10], items [IV], [VIII], [XII], [XXIV], [XXVII], [XXXI], [XXXVIII], [XXXIX], [XLIX], [LIII], [LVI], [LVIII], [LXIV].) In the form above, the principle says in effect that if $f_{\mathfrak{M}}(x) = s$, then if s has in $\mathfrak{M}$ whatever properties are preserved under the consequence relation, so too does Tx. (See thesis (7) for one choice of such a property.)

(2) *Insolubles are false.* This too is explicitly held by Buridan, in his explanation of the suppositional rules of truth ([3], p. 92). We call this thesis (IF).

(3) *The normal suppositional rules of truth fail exactly in the cases of the insolubles.* This too is more or less explicitly held by Buridan. He does explicitly say that the usual rules hold for all non-insoluble cases ([3], p. 93). He also says that the rules fail in the case of affirmative insolubles ([3], p. 198; compare [3], pp. 90-93). It does not seem illegitimate to add to this that they also fail for insoluble negatives, although Buridan does not stipulate this explicitly. Since the normal suppositional rules give us an account of correspondence, as observed above, thesis (3) amounts to saying that *the correspondence theory of truth fails exactly in the cases of the insolubles*: $s \in \mathsf{INS}_{\mathfrak{M}}$ *iff it is not the case, that* $s \in \mathsf{CORR}_{\mathfrak{M}}$ *iff* $V_{\mathfrak{M}}(s) = 1$. Hence we call it (CT-INS).

(4) *Bivalence*: *Every* (*existing*) *sentence is either true or false.* Buridan does not explicitly hold this. Nevertheless, he mentions only two values, and we ought to follow him as closely as possible. I know of only two mediaeval authors ([10], items [LXIII] and [X], 4th previous opinion) who allowed for failure of bivalence in the case of insolubles. We call this thesis (BIV).

(5) *If it is true that a sentence s is true, then s is true.* Buridan does not explicitly hold this either. But he says nothing that indicates that he would

reject it, and indeed maintains some things that lead directly up to it. (See [11], pp. 41-42.) I know of no mediaeval who questioned it. Call this thesis (TT).

(6) *If x does not stand for a sentence, then* ⌜*x is true*⌝ *is false*. This principle is the basis for Buridan's solution of a sophism in [3], p. 189f. Call this thesis (NS).

(7) *Between sentences that exist, the valid consequences are just the truth-preserving transformations.* Buridan explicitly holds one half of this: "It is true that of every valid consequence, it is impossible for the antecedent to be true without the truth of the consequent formed at the same time as it." ([3], p. 183) On the other hand, in the same place he holds that

> it is not sufficient for a consequence to be valid that it is impossible for the antecedent to be true without the consequent formed at the same time as it. . . . So more is required, namely, that it could not be as the antecedent signifies unless it were as is signified by the consequent.

The point of this revision is to accomodate certain problems raised by his view that only existing sentence-tokens have truth-values. If valid consequences were just the truth-preserving transformations, then consequences such as 'Every proposition is affirmative, so none is negative' ([3], p. 183) would not be valid. Whenever the antecedent is true, the consequent cannot exist, and so cannot be true.

Buridan resolves this problem by observing that the world may be as a certain sentence-token would signify it to be, even though that sentence-token might not in fact exist. Thus, if Socrates is running, then the world is as would be signified by a sentence-token of the type 'Socrates is running', even if that sentence-token did not in fact exist. We shall say that the sentence-token is "secure" in this case. If in addition, the sentence-token happens to exist, then it is true. A sentence-token is true iff it is secure and exists (see [6] and [9]. The term 'secure' is taken from [6]). Buridan's revised notion of consequence, then, amounts to saying that the valid consequences are just the transformations that always preserve security. As long as we confine ourselves to contexts in which all sentence-tokens exist, truth and security collapse, and principle (7) holds. We call the principle (TC).

(8) *The valid consequences are just the traditional ones.* Buridan nowhere explicitly say this. But if we look closely at his text ([3], pp. 180-185), it appears that the problem which led him to base consequence on security rather than on truth was just that certain traditionally valid argument-forms would have to be given up, and certain traditionally invalid forms would have to be admitted—once the bearer of a truth-value was made the existing sentence-token—if the usual notion of consequence as truth-preserving were retained. Thus, Buridan's goal seems to have been to revise the notion of consequence in such a way that just the traditionally valid argument-forms could be retained. Since the traditional consequences which will concern us will be the double-negation rules, we call this thesis (DN).

4 The eight theses above must be kept in mind as we attempt to reconstruct Buridan's position. As we shall see, they are mutually incompatible. Let us begin with the simple syntax **SYN** of [12], with a primitive vocabulary consisting of denumerable singular terms, two primitive predicates P and T, and a negation operator. If a sentence s has an atomic constituent of the form Tx, then we shall call s a *semantic sentence*. Let $n(s)$ be the negation of s. This allows us to write '$n^i(s)$' for the i-fold negation of s. If i is odd and s is of the form Tx, we call $n^i(s)$ an *odd semantic sentence*. A *model* for **SYN** will be a triple $\mathfrak{M} = \langle X_{\mathfrak{M}}, f_{\mathfrak{M}}, g_{\mathfrak{M}}\rangle$, where $X_{\mathfrak{M}}$ includes the set of all sentences of **SYN**, $f_{\mathfrak{M}}$ is a one-one function whose domain is the set of singular terms of **SYN** and whose range is $X_{\mathfrak{M}}$, and $g_{\mathfrak{M}}$ assigns a subset of $X_{\mathfrak{M}}$ to P. These restrictions allow us to define for each model $\mathfrak{M}$ a function $t_{\mathfrak{M}}$ assigning to each sentence s its unique *truth sentence*, the sentence Tx such that $f_{\mathfrak{M}}(x) = s$. (Compare [12].) Finally, for any set Y of sentences of **SYN**, form the smallest set Y^* including Y and such that if $n(s)$ is in Y^*, so is s. (Y^* is thus the smallest set containing every constituent sentence of every member of Y.) Then a *naive valuation* over a set Y of sentences with respect to a model $\mathfrak{M}$ for **SYN** is a function $V_{\mathfrak{M}}$ with domain Y^* and range $\{0, 1\}$, such that: (a) $V_{\mathfrak{M}}(Px) = 1$ iff $f_{\mathfrak{M}}(x) \in g_{\mathfrak{M}}(P)$; (b) $V_{\mathfrak{M}}(n(s)) = 1 - V_{\mathfrak{M}}(s)$; (c) $V_{\mathfrak{M}}(t_{\mathfrak{M}}(s)) = V_{\mathfrak{M}}(s)$. Observe that (a) and (b) follow from the Buridanian thesis of bivalence (BIV) and the correspondence theory of truth. On the other hand, (c) follows from the theses (VIP), (BIV), (TT), and (TC).

5 We introduce the auxiliary notion of the syntactic relation G which a semantic sentence bears to its grammatical subject. Then, with respect to a model $\mathfrak{M}$, we define the relation $R_{\mathfrak{M}}$ of "semantic descent" to be the proper ancestral of the relative product of G into $f_{\mathfrak{M}}$. Now an $\mathcal{L}_1$*-model* will be a model $\mathfrak{M}$ for **SYN** such that: (1) if s is $R_{\mathfrak{M}}$-ungrounded (see [5]), then for some s' (possibly $s = s'$), $R_{\mathfrak{M}}(s, s')$ and $R_{\mathfrak{M}}(s', s')$. That is, there are no non-cyclic infinite chains of semantic descent in $\mathfrak{M}$. (2) $R_{\mathfrak{M}}$ is antisymmetric. There are no $R_{\mathfrak{M}}$-cycles of length greater than one. If a sentence is a semantic descendant of itself, it is an *immediate* semantic descendant of itself. (3) Only odd semantic sentences are semantic descendants of themselves. Cases such as $s = t_{\mathfrak{M}}(s)$ are ruled out. Thus in $\mathcal{L}_1$-models a sentence is $R_{\mathfrak{M}}$-ungrounded iff it either is or leads by an $R_{\mathfrak{M}}$-path to a directly self-referential odd semantic sentence. Self-reference of this kind is paradoxical. The paradigm is: $s = n(t_{\mathfrak{M}}(s))$, but we also have in general $s = n^i(t_{\mathfrak{M}}(s))$ for odd i. Condition (2) above is imposed in order to allow us to study this simple kind of paradox in isolation, without worrying about more complicated cases in which reference is passed from hand to hand, as it were, around a cycle with several members. They are a topic for another paper. Conditions (1) and (3), on the other hand, are imposed in order to rule out cases that Buridan simply does not treat.

6 For a language whose syntax is **SYN** and whose admissible models are $\mathcal{L}_1$-models, we define: *A sentence* s *is an insoluble in an admissible model* $\mathfrak{M}$ (*i.e.*, $s \in \mathsf{INS}_{\mathfrak{M}}$) iff $R_{\mathfrak{M}}(s, s)$. This definition appears to capture just the sentences that Buridan would want to call insoluble within our restricted framework. He several times mentions an insoluble's "reflection on itself'' ([3], p. 180 and *passim*). Yet, he makes it clear that not all such ''reflection'' or self-reference generates paradox ([3], pp. 192f.). The ''reflection'' in question appears to be the kind in which a sentence is its own semantic descendant. That Buridan would want to call *all* such cases within our context insoluble follows from Lemma 2, below. Where $R_{\mathfrak{M}}(s, s)$, there is no naive valuation over $\{s\}$. Hence s requires an exception to (VIP), (BIV), (TT), (TC) or the correspondence theory of truth. But since Buridan holds the first four theses, it is the correspondence theory that is violated. Hence, by (CT-INS), s is insoluble. Within the context provided by **SYN** and $\mathcal{L}_1$-models, we can prove the following preliminary results.

Lemma 1 *The insolubles are just the odd semantic sentences that refer directly to themselves*: $s \in \mathsf{INS}_{\mathfrak{M}}$ *iff for some odd* i, $s = n^i(t_{\mathfrak{M}}(s))$.

The proof follows by the definition of $\mathsf{INS}_{\mathfrak{M}}$ and conditions (2) and (3) on $\mathcal{L}_1$-models.

Lemma 2 *Is s is insoluble in a model* $\mathfrak{M}$, *then there is no naive valuation* $V_{\mathfrak{M}}$ *over* $\{s\}$.

For suppose $s \in \mathsf{INS}_{\mathfrak{M}}$. Then for some odd i, $s = n^i(t_{\mathfrak{M}}(s))$, by Lemma 1. Suppose further that there is a naive valuation $V_{\mathfrak{M}}$ over $\{s\}$. Then $V_{\mathfrak{M}}(s) = V_{\mathfrak{M}}(n^i(t_{\mathfrak{M}}(s)))$. Hence, by condition (b) on naive valuations, $V_{\mathfrak{M}}(s) \neq V_{\mathfrak{M}}(t_{\mathfrak{M}}(s))$, since i is odd by hypothesis. Thus by condition (c) on naive valuations, $V_{\mathfrak{M}}(s) \neq V_{\mathfrak{M}}(s)$. The Lemma follows by *reductio*.

Notice that for models (not $\mathcal{L}_1$-models) in which $s = n^i(t_{\mathfrak{M}}(s))$ for i *not* odd, there are naive valuations over $\{s\}$. $\mathcal{L}_1$-models thus rule out all but ''vicious'' kinds of direct self-reference, in which cases the naive rules of valuation break down.

Where s is atomic, let us call the sequence: $\langle s, n(s), n^2(s), n^3(s), \ldots \rangle$ a *negation sequence*. Then:

Lemma 3 *A negation sequence contains at most one insoluble in a given model*: *if* $n^i(s)$ *and* $n^j(s)$ *are both insoluble in* $\mathfrak{M}$, *then* $i = j$.

The proof is straightforward, from Lemma 1.

7 It turns out that even within the simple context of **SYN** and $\mathcal{L}_1$-models, the Buridanian theses extracted above are mutually inconsistent. In fact, we need only three of the theses to derive a contradiction:

Theorem 1 *The theses* (CT-INS), (TC), *and* (DN) *are mutually inconsistent*.

For (1), suppose that $s \in \mathsf{INS}_{\mathfrak{M}}$. Then (2), $n^2(s) \notin \mathsf{INS}_{\mathfrak{M}}$ by Lemma 3. Now (3), suppose further that $s \in \mathsf{CORR}_{\mathfrak{M}}$. Then (4), $n^2(s) \in \mathsf{CORR}_{\mathfrak{M}}$. Moreover (5), $V_{\mathfrak{M}}(s) \neq 1$, by (1), (3) and (CT-INS). Thus (6), $V_{\mathfrak{M}}(n^2(s)) \neq 1$, by (5), (TC) and (DN). But nevertheless (7), $V_{\mathfrak{M}}(n^2(s)) = 1$, by (2), (4) and (CT-INS). Hence (8), $s \notin \mathsf{CORR}_{\mathfrak{M}}$, from (3)-(7) by *reductio*. Therefore (9), $n^2(s) \notin \mathsf{CORR}_{\mathfrak{M}}$. Now (10), $V_{\mathfrak{M}}(s) = 1$, by (1), (8) and (CT-INS). And so (11), $V_{\mathfrak{M}}(n^2(s)) = 1$, from (10), (TC) and (DN). But (12), $V_{\mathfrak{M}}(n^2(s)) \neq 1$, by (2), (9) and (CT-INS). Steps (11) and (12) are of course contradictory.

What Theorem 1 shows is that Buridan's position is incoherent. The tension can be resolved only by departing from one or more of the three theses involved in Theorem 1. Several modern approaches make the compromise at (CT-INS). For instance, any solution in which the Liar sentence fails of bivalence and in which negation is choice-negation (see [6], p. 27) is of this kind. Such an approach departs from the correspondence theory not only in the case of insolubles, but also in the case of their double negations, which are *not* insoluble, by Lemma 3. However much such approaches have to offer, we shall not follow them in sacrificing (CT-INS). For it seems that Buridan himself would have chosen to make the compromise elsewhere, as I shall argue presently. Accordingly, we can give up either (TC) or (DN). If we give up (TC), we must ask whether any interesting property will replace truth as being preserved under valid consequence. The most plausible candidate is perhaps correspondence. Thesis (TC) would then be replaced by (CC): *Between sentences that exist, the valid consequences are just the correspondence-preserving transformations*. But it is easy to show that this maneuver forces us to give up at least one more of our Buridanian theses. For suppose that $s \in \mathsf{INS}_{\mathfrak{M}}$. Then $V_{\mathfrak{M}}(s) = 0$ by (IF). Hence $s \in \mathsf{CORR}_{\mathfrak{M}}$ by (CT-INS). From this two things follow. First, $t_{\mathfrak{M}}(s) \in \mathsf{CORR}_{\mathfrak{M}}$, using (VIP) and (CC). But also $t_{\mathfrak{M}}(s) \notin \mathsf{CORR}_{\mathfrak{M}}$, using Lemma 1 and the fact that $s \in \mathsf{INS}_{\mathfrak{M}}$, together with the definition of $\mathsf{CORR}_{\mathfrak{M}}$. Hence we would have to give up either (IF) or (VIP). For the sake of keeping as close as possible to Buridan, we shall not adopt this policy. Instead, we shall make our compromise at (DN)—negation is going to behave in a peculiar way.

There is some indication that Buridan himself would make the compromise here. For despite his wanting to retain the traditionally valid argument-forms, he appears also to have recognized that his negation did peculiar things. In [3], ch. 8, sophism 8, Buridan has Socrates say only 'Plato speaks falsely' and Plato say only 'Socrates speaks falsely'. This is not a situation that can be represented in the limited context of **SYN** and $\mathcal{L}_1$-models, and indeed we should not normally think of the case as generating a paradox. But Buridan does; he holds that both Socrates and Plato utter insolubles. For our purposes, the interesting point in this discussion is that Buridan allows that an insoluble and its negation are both false ([3], p. 199).

8 In the remaining sections of this paper, we shall study the language $\mathcal{L}_1$ whose syntax is **SYN** and whose admissible models are just the $\mathcal{L}_1$-models. The rules of valuation for $\mathcal{L}_1$ are these: (A) $V_{\mathfrak{M}}(Px) = 1$ iff $f_{\mathfrak{M}}(x) \in g_{\mathfrak{M}}(P)$, and $V_{\mathfrak{M}}(Px) = 0$ otherwise. (B) $V_{\mathfrak{M}}(t_{\mathfrak{M}}(s)) = V_{\mathfrak{M}}(s)$. (C) If $f_{\mathfrak{M}}(x)$ is not a sentence of **SYN**, then $V_{\mathfrak{M}}(Tx) = 0$. (D) If $s \in \mathsf{INS}_{\mathfrak{M}}$ or $n(s) \in \mathsf{INS}_{\mathfrak{M}}$, then $V_{\mathfrak{M}}(n(s)) = 0$. (E) If $s \notin \mathsf{INS}_{\mathfrak{M}}$ and $n(s) \notin \mathsf{INS}_{\mathfrak{M}}$, then $V_{\mathfrak{M}}(n(s)) = 1 - V_{\mathfrak{M}}(s)$. It can be shown that for each admissible model $\mathfrak{M}$ for $\mathcal{L}_1$, these rules assign exactly one of 0, 1 to each sentence of **SYN**. For the proof, which is straightforward but rather lengthy, see [11]. (The definition of insolubility is slightly different there, but equivalent.)

Theorem 1 showed that Buridan's position cannot be consistently represented as a whole. But making the necessary sacrifices, $\mathcal{L}_1$ is a good reconstruction within the limits of its own syntax and models. The Buridanian theses (IF), (BIV) and (NS) are satisfied directly by the rules (A)-(E) above. Thesis (TT) follows at once from (B). (CT-INS) will be proven below as Theorem 4. Since we agreed that valid consequences were to be truth-preserving, thus satisfying (TC), we also thereby satisfy (VIP), in view of (B). But the double negation rules (DN) fail whenever they lead to or from an insoluble, in virtue of Theorem 8 below. Conversely, given a language with the syntax and models of $\mathcal{L}_1$, and given the definition of $\mathsf{INS}_{\mathfrak{M}}$, rule (A) follows from (CT-INS) and (BIV). One half of (B), namely, $V_{\mathfrak{M}}(s) \leq V_{\mathfrak{M}}(t(s))$, follows from (VIP), (TC) and (BIV). The other half follows from (TT) and (BIV). Rule (C) is just (NS). Rule (D) follows from (IF), (CT-INS) and (BIV). Rule (E) follows from (CT-INS) and (BIV). Thus, given the syntax and models of $\mathcal{L}_1$, and given the definition of an insoluble, Buridan's own principles entail (A)-(E). (DN) is of course suspended.

Buridan does not discuss infinite non-cyclic chains, nor cycles of length one that are not vicious (e.g., $s = t_{\mathfrak{M}}(s)$). His silence on these matters is honored by the fact that $\mathcal{L}_1$-models do not admit such cases. He does discuss cycles of length greater than one, and sentences formed with binary connectives, but these are beyond the resources of $\mathcal{L}_1$, which takes no stand on them. They are matters for another study.

9 Theorem 2 *The extension of P is given by* $g_{\mathfrak{M}}$: $\mathsf{EXT}_{\mathfrak{M}}(P) = g_{\mathfrak{M}}(P)$.

For $\mathsf{EXT}_{\mathfrak{M}}(P) = \{f_{\mathfrak{M}}(x): f_{\mathfrak{M}}(x) \in g_{\mathfrak{M}}(P)\}$, by rule (A) and the definition of $\mathsf{EXT}_{\mathfrak{M}}$. The theorem follows by the conditions on models.

Theorem 3 *The extension of T in a model is the class of sentences true in that model*: $\mathsf{EXT}_{\mathfrak{M}}(T) = \{s: V_{\mathfrak{M}}(s) = 1\}$.

This follows from rules (B) and (C) and the definition of $\mathsf{EXT}_{\mathfrak{M}}$.

We shall say that ϕ is a *predicate* of $\mathcal{L}_1$ iff ϕ is either a primitive predicate of $\mathcal{L}_1$, i.e., either P or T, or else formed from a primitive predicate of $\mathcal{L}_1$ by prefixing a string of i negations ($i \geq 0$). We let $N^i(\phi)$ be the result of prefixing i negation operators to ϕ. We now extend the definition of $\mathsf{EXT}_{\mathfrak{M}}$ to all predicates of $\mathcal{L}_1$. If ϕ is a primitive predicate of $\mathcal{L}_1$, then $\mathsf{EXT}_{\mathfrak{M}}(N^i(\phi)) = \mathsf{EXT}_{\mathfrak{M}}(\phi)$ for i not odd, and $\mathsf{EXT}_{\mathfrak{M}}(N^i(\phi)) = X_{\mathfrak{M}} - \mathsf{EXT}_{\mathfrak{M}}(\phi)$ for odd i. (It might have seemed more natural to allow $\mathsf{EXT}_{\mathfrak{M}}(\phi)$ to be defined at the outset as $\{f_{\mathfrak{M}}(x): V_{\mathfrak{M}}(\phi x) = 1\}$ for *all* predicates ϕ, primitive or not. But this would have had the effect of excluding insolubles from the extension of the non-truth predicate $N(T)$, even though they are non-truths. In terms of personal supposition theory—where a predicate stands for (*supponit pro*) the members of its extension—this would mean that in the insoluble $s = n(t_{\mathfrak{M}}(s))$, $N(T)$ would not stand for the whole sentence of which it is a part. Buridan unequivocally rejects that approach. (See [3], pp. 192f.)

Lemma 4 *The correspondence theory of truth holds for non-insolubles*: *if* $s \notin \mathsf{INS}_{\mathfrak{M}}$, *then* $s \in \mathsf{CORR}_{\mathfrak{M}}$ *iff* $V_{\mathfrak{M}}(s) = 1$.

For suppose $s \notin \mathsf{INS}_{\mathfrak{M}}$. Then if s is a non-semantic sentence $n^i(Px)$, $Px \in \mathsf{CORR}_{\mathfrak{M}}$ iff $f_{\mathfrak{M}}(x) \in \mathsf{EXT}_{\mathfrak{M}}(P)$. Hence, by Theorem 2 and rule (A), $Px \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(Px) = 1$. Thus the Lemma follows by i applications of rule (E) and the definition of $\mathsf{CORR}_{\mathfrak{M}}$. If, on the other hand, s is a semantic sentence $n^i(Tx)$, then

Case 1: $f_{\mathfrak{M}}(x)$ is not a semantic sentence. Then $f_{\mathfrak{M}}(x) \notin \mathsf{EXT}_{\mathfrak{M}}(T)$ by Theorem 3. Hence $Tx \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(Tx) = 1$ by rule (C) and the definition of $\mathsf{CORR}_{\mathfrak{M}}$. Thus again, the Lemma follows by i applications of rule (E) and the definition of $\mathsf{CORR}_{\mathfrak{M}}$.

Case 2: $f_{\mathfrak{M}}(x)$ is a sentence s', and s is $R_{\mathfrak{M}}$-grounded. Then $s = n^i(t_{\mathfrak{M}}(s'))$. Now $t_{\mathfrak{M}}(s') \in \mathsf{CORR}_{\mathfrak{M}}$ iff $s' \in \mathsf{EXT}_{\mathfrak{M}}(T)$, by the definition of $\mathsf{CORR}_{\mathfrak{M}}$. By Theorem 3 and rule (B), it follows that $t_{\mathfrak{M}}(s') \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(t_{\mathfrak{M}}(s')) = 1$. Hence the Lemma follows by i applications of rule (E) and the definition of $\mathsf{CORR}_{\mathfrak{M}}$.

Case 3: s is $R_{\mathfrak{M}}$-ungrounded. Again, for some s', $s = n^i(t_{\mathfrak{M}}(s'))$. If for no $0 \leqslant j \leqslant i$ is $n^j(t_{\mathfrak{M}}(s')) \in \mathsf{INS}_{\mathfrak{M}}$, the argument is just as for Case 2. If the contrary, j is odd by Lemma 1. Thus $n^{j+1}(t_{\mathfrak{M}}(s')) \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(s') = 1$, by the definition of $\mathsf{CORR}_{\mathfrak{M}}$ and rule (B). Now since $n^j(t_{\mathfrak{M}}(s')) \in \mathsf{INS}_{\mathfrak{M}}$, Lemma 1 yields $n^j(t_{\mathfrak{M}}(s')) = s'$. Thus $V_{\mathfrak{M}}(s') = 0$ and $V_{\mathfrak{M}}(n^{j+1}(t_{\mathfrak{M}}(s'))) = 0$ by rule (D). From (BIV) and the above, we have $n^{j+1}(t_{\mathfrak{M}}(s')) \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(n^{j+1}(t_{\mathfrak{M}}(s'))) = 1$. Now $j \neq i$, since by the hypothesis of the Lemma $s \notin \mathsf{INS}_{\mathfrak{M}}$. Hence by Lemma 3, there will be no insolubles from $n^{j+1}(t_{\mathfrak{M}}(s'))$ to $n^i(t_{\mathfrak{M}}(s'))$. So from the above, by $i - (j+1)$ applications of rule (E) and the definition of $\mathsf{CORR}_{\mathfrak{M}}$, the Lemma follows. This completes the proof.

Lemma 5 *Every insoluble corresponds*: $s \in \mathsf{INS}_{\mathfrak{M}}$ *only if* $s \in \mathsf{CORR}_{\mathfrak{M}}$.

For suppose $s \in \mathsf{INS}_{\mathfrak{M}}$. Then for some odd i, $s = n^i(t_{\mathfrak{M}}(s))$, by Lemma 1. Now the definition of $\mathsf{CORR}_{\mathfrak{M}}$ and Theorem 3 yield $t_{\mathfrak{M}}(s) \in \mathsf{CORR}_{\mathfrak{M}}$ iff $V_{\mathfrak{M}}(s) = 1$. Hence by rule (D), $t_{\mathfrak{M}}(s) \notin \mathsf{CORR}_{\mathfrak{M}}$. So $s \in \mathsf{CORR}_{\mathfrak{M}}$ by the definition of $\mathsf{CORR}_{\mathfrak{M}}$, since i is odd.

Theorem 4 *The correspondence theory of truth fails exactly for insolubles*: (CT-INS).

This follows from Lemmata 4 and 5 and rule (D).

If ϕ is a primitive predicate of $\mathcal{L}_1$, we shall say that $N^i(\phi)$ is *safe* just in case for all admissible models $\mathfrak{M}$, $\mathsf{EXT}_{\mathfrak{M}}(N^i(\phi)) = \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^i(\phi x)) = 1\}$. Then:

Lemma 6 *If ϕ is a primitive predicate and i is not odd, then $N^i(\phi)$ is safe.*

For in that case $n^i(\phi x) \notin \mathsf{INS}_{\mathfrak{M}}$, by Lemma 1, and so $V_{\mathfrak{M}}(n^i(\phi x)) = 1$ iff $f_{\mathfrak{M}}(x) \in \mathsf{EXT}_{\mathfrak{M}}(\phi)$ by Theorem 4. Now $\mathsf{EXT}_{\mathfrak{M}}(N^i(\phi)) = \mathsf{EXT}_{\mathfrak{M}}(\phi) = \{f_{\mathfrak{M}}(x)\colon f_{\mathfrak{M}}(x) \in \mathsf{EXT}_{\mathfrak{M}}(\phi)\}$, from the (extended) definition of $\mathsf{EXT}_{\mathfrak{M}}$. The Lemma follows by substitution.

Lemma 7 *If i is odd, then $N^i(P)$ is safe.*

For in that case $n^i(Px) \notin \mathsf{INS}_{\mathfrak{M}}$ by Lemma 1, and so $V_{\mathfrak{M}}(n^i(Px)) = 1$ iff $f_{\mathfrak{M}}(x) \notin \mathsf{EXT}_{\mathfrak{M}}(P)$ by Theorem 4 and the definition of $\mathsf{CORR}_{\mathfrak{M}}$, since i is odd. Now $\mathsf{EXT}_{\mathfrak{M}}(N^i(P)) = X_{\mathfrak{M}} - \mathsf{EXT}_{\mathfrak{M}}(P) = \{f_{\mathfrak{M}}(x)\colon f_{\mathfrak{M}}(x) \notin \mathsf{EXT}_{\mathfrak{M}}(P)\}$, by the (extended) definition of $\mathsf{EXT}_{\mathfrak{M}}$, since the range of $f_{\mathfrak{M}}$ is $X_{\mathfrak{M}}$. The Lemma follows by substitution.

Theorem 5 *The unsafe predicates of $\mathcal{L}_1$ are just the predicates $N^i(T)$ for odd i.*

That the other predicates are safe follows from Lemmata 6 and 7. Let $s = n^i(t_{\mathfrak{M}}(s))$ for odd i. Then $s \in \mathsf{INS}_{\mathfrak{M}}$ by Lemma 1, and so $s \in X_{\mathfrak{M}} - \mathsf{EXT}_{\mathfrak{M}}(T)$ by rule (D) and Theorem 3. Thus $s \in \mathsf{EXT}_{\mathfrak{M}}(N^i(T))$, by the (extended) definition of $\mathsf{EXT}_{\mathfrak{M}}$. Now $V_{\mathfrak{M}}(n^i(t_{\mathfrak{M}}(s))) = 0$, by rule (D), and so $s \notin \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^i(Tx)) = 1\}$. Thus by extensionality, the model $\mathfrak{M}$ shows that $N^i(T)$ is unsafe.

An unsafe predicate $N^i(T)$ will be said to *fail in* those models $\mathfrak{M}$ and *for* those sentences s such that $s \in \mathsf{EXT}_{\mathfrak{M}}(N^i(T)) - \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^i(Tx)) = 1\}$. Then:

Theorem 6 *No two unsafe predicates of $\mathcal{L}_1$ fail in exactly the same models for exactly the same sentences: If i and j are odd, and $i \neq j$, then there is and s and $\mathfrak{M}$ such that $N^i(T)$ fails in $\mathfrak{M}$ for s and $N^j(T)$ does not.*

For, given the conditions, suppose that $s = n^i(t_{\mathfrak{M}}(s))$ and $s' = n^j(t_{\mathfrak{M}}(s))$. Then $s \in \mathsf{EXT}_{\mathfrak{M}}(N^i(T))$ and $s \notin \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^i(Tx)) = 1\}$, as in the proof of Theorem 5. Thus $N^i(T)$ fails in $\mathfrak{M}$ for s. Now $s' \notin \mathsf{INS}_{\mathfrak{M}}$ by Lemmata 1 and 3. Thus $V_{\mathfrak{M}}(s') = 1$ iff $s' \in \mathsf{CORR}_{\mathfrak{M}}$ by Theorem 4. Hence $V_{\mathfrak{M}}(s') = 1$ iff $V_{\mathfrak{M}}(s) = 0$ by the definition of $\mathsf{CORR}_{\mathfrak{M}}$, Theorem 3 and (BIV), since j is odd. Then since $V_{\mathfrak{M}}(s) = 0$ by Lemma 1 and rule (D), $V_{\mathfrak{M}}(s') = 1$. Hence $s \in \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^j(Tx)) = 1\}$, and so $N^j(T)$ does not fail in $\mathfrak{M}$ for s.

On the other hand, there are admissible models in which every unsafe predicate fails for some sentence. Let $s_1 = n(t_{\mathfrak{M}}(s_1))$, $s_2 = n^3(t_{\mathfrak{M}}(s_2))$, and in general $s_i = n^{2i-1}(t_{\mathfrak{M}}(s_i))$. Nevertheless, in a given model and for a given sentence, at most one predicate fails:

Theorem 7 *If i and j are odd and $i \neq j$, then if $s \notin \mathsf{EXT}_{\mathfrak{M}}(T)$ and $s \notin \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^i(Tx)) = 1\}$, then $s \in \{f_{\mathfrak{M}}(x)\colon V_{\mathfrak{M}}(n^j(Tx)) = 1\}$.*

For given the conditions, $V_{\mathfrak{M}}(t_{\mathfrak{M}}(s)) = 0$ by Theorem 3, and also $V_{\mathfrak{M}}(n^i(t_{\mathfrak{M}}(s))) = 0$. Suppose $n^i(t_{\mathfrak{M}}(s)) \notin \mathsf{INS}_{\mathfrak{M}}$. Then $n^i(t_{\mathfrak{M}}(s)) \notin \mathsf{CORR}_{\mathfrak{M}}$ by Theorem 4, and $t_{\mathfrak{M}}(s) \in \mathsf{CORR}_{\mathfrak{M}}$, since i is odd. But since $V_{\mathfrak{M}}(t_{\mathfrak{M}}(s)) = 0$, Theorem 4 requires by *reductio* that $n^i(t_{\mathfrak{M}}(s)) \in \mathsf{INS}_{\mathfrak{M}}$ after all. Hence $n^j(t_{\mathfrak{M}}(s)) \notin \mathsf{INS}_{\mathfrak{M}}$ by Lemma 3, and so $V_{\mathfrak{M}}(n^j(t_{\mathfrak{M}}(s))) = 1$ iff $s \notin \mathsf{EXT}_{\mathfrak{M}}(T)$, by Theorem 4 and the definition of $\mathsf{CORR}_{\mathfrak{M}}$, since j is odd. This and the given conditions yield $V_{\mathfrak{M}}(n^j(t_{\mathfrak{M}}(s))) = 1$, and so the Theorem holds.

Theorems 5-7 indicate some of the peculiar behavior of negation in $\mathcal{L}_1$. The following Theorem shows the basis of this peculiarity. Let s be an atomic sentence. If for some $i \geq 0$, $V_{\mathfrak{M}}(n^i(s)) = V_{\mathfrak{M}}(s)$ iff i is not odd, then we say that negation *behaves classically in $\mathfrak{M}$ for $n^i(s)$*). Then:

Theorem 8 *Negation behaves classically in $\mathfrak{M}$ except just for the insolubles in $\mathfrak{M}$.*

First, if s is atomic and for no $i \geqslant 0$ is $n^i(s) \in \mathsf{INS}_{\mathfrak{M}}$, then negation behaves classically for $i = 0$, trivially. For $i > 0$, $V_{\mathfrak{M}}(n^i(s))$ is calculated on the basis of $V_{\mathfrak{M}}(s)$ by i applications of rule (E), and so behaves classically. On the other hand, if s is atomic and for some $j \geqslant 0$, $n^j(s) \in \mathsf{INS}_{\mathfrak{M}}$, then $V_{\mathfrak{M}}(s) = 0$ and for all $i \geqslant 0$, negation behaves classically in $\mathfrak{M}$ for $n^i(s)$, *except* where $i = j$, as we shall now show. $V_{\mathfrak{M}}(n^j(s)) = 0$ by rule (D). Also, $n^j(s) = n^j(t_{\mathfrak{M}}(n^j(s)))$ and j is odd by Lemma 1. Thus $V_{\mathfrak{M}}(s) = V_{\mathfrak{M}}(t_{\mathfrak{M}}(n^j(s))) = 0$, by rule (B). So $V_{\mathfrak{M}}(s) = V_{\mathfrak{M}}(n^j(s))$. But since j is odd, this means that negation does not behave classically in $\mathfrak{M}$ for the insoluble $n^j(s)$. On the other hand, it does behave classically for the atomic s, trivially. By Lemma 3, there are no insolubles between s and $n^j(s)$. So for $0 < i < j$, $n^i(s)$ is evaluated by rule (E), guaranteeing that negation behave classically in these cases. By rule (D), $V_{\mathfrak{M}}(n^{j+1}(s)) = 0$, and since $j + 1$ is even, negation behaves classically for $n^{j+1}(s)$. By Lemma 3 again, for no $i > j$ is $n^i(s) \in \mathsf{INS}_{\mathfrak{M}}$. Thus $n^i(s)$ for $i > j + 1$ is evaluated on the basis of $V_{\mathfrak{M}}(n^{j+1}(s))$ by rule (E), and so negation behaves classically in these cases too.

From Theorem 8, we see that in the "normal" case, negation marches through the sequence $\langle s, n(s), n^2(s), \ldots \rangle$ by *alternating* between the values 0 and 1 or 1 and 0. But in the case of such a sequence that contains an insoluble, the insoluble forces negation out of step, as it were, so that a 0 is followed by a 0. Recovery occurs at once, and the next step is a 0, as it should be. From here on, the march is back "in step", and proceeds as if nothing had happened.

In future papers, I shall examine some of the complications that arise when Buridan considers situations in which two or more sentences are arranged in an $R_{\mathfrak{M}}$-cycle, and in which sentences with binary connectives are taken into account.

REFERENCES

[1] Buridan, J., *In Metaphysicen Aristotelis quaestiones argutissimae magistri Joannis Buridani*, Paris, Jodocus Badius Ascensus, 1518.

[2] Buridan, J., *Sophismata Buridani*, Paris: Antoine Denidel and Nicole de la Barre, [c. 1495-1500] (*Incunabula in American Libraries*, 3rd census, B-1295). Copy at Harvard University Library, incunabula 8325.5.

[3] Buridan, J., *Sophisms on Meaning and Truth*, Theodore Kermit Scott, tr., "Century Philosophy Sourcebooks"; New York: Appleton-Century-Crofts (1966).

[4] Faral, E., "Jean Buridan: Maître ès arts de l'Université de Paris" in *Histoire littéraire de la France*, vol. 38, Paris, Imprimerie nationale (1949), pp. 462-605.

[5] Herzberger, H. G., "Paradoxes of grounding in semantics," *The Journal of Philosophy*, vol. 67 (1970), pp. 145-167.

[6] Herzberger, H. G., "Truth and modality in semantically closed languages," in Robert L. Martin, ed., *The Paradox of the Liar*, New Haven, Connecticut: Yale University Press (1970), pp. 25-46.

[7] Moody, Ernest A., *Truth and Consequence in Mediaeval Logic*, "Studies in Logic and the Foundations of Mathematics"; Amsterdam, North-Holland Publishing Company (1953).

[8] Prior, A. N., "The possibly-true and the possible", *Mind*, vol. 78 (1969), pp. 481-492.

[9] Prior, A. N., "Some problems of self-reference in John Buridan," *Proceedings of the British Academy*, vol. 48 (1962), pp. 281-296.

[10] Spade, P. V., *The Mediaeval Liar: A Catalogue of the Insolubilia-Literature,* Subsidia Mediavalia, vol. 5, Toronto, Pontifical Institute of Mediaeval Studies (1975).

[11] Spade, P. V., *The Mediaeval Liar: A Study of John Buridan's Position on the Paradox with a Catalogue of the Insolubilia-Literature of the Middle Ages*, Ph.D. dissertation, University of Toronto (1972).

[12] Spade, P. V., "On a conservative attitude toward some naive semantic principles," *Notre Dame Journal of Formal Logic*, vol. XVI (1975), pp. 597-602.

[13] Tarski, A., "The semantic conception of truth," in Herbert Feigl and Wilfrid Sellars, eds., *Readings in Philosophical Analysis*, New York, Appleton-Century-Crofts (1949), pp. 52-84.

Indiana University
Bloomington, Indiana