BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 20, Number 2, April 1989

0273-0979/89 $1.00 + $.25 per page

The Liar: An essay in truth and circularity, by Jon Barwise and John Etchemendy, Oxford University Press, New York and Oxford, 1987, xii + 185 pp., $19.95. ISBN 0-19-505072-x

Consider the classic Liar sentence: "This sentence is false." It claims that it is false. So if we assume that a sentence is true if and only if what it claims is the case, then the Liar is true if and only if it is false. People have thought about this paradox for centuries. Despite this, there is no single standard "solution." An attempted resolution of the paradox would tell us which of our intuitions are sound and which need further clarification. It would point out where and why our naive reasoning leads us to a contradiction.

Modern logic applies mathematical methods to the modeling and study of *truth*, *proof*, *computation*, and *infinity*. The paradoxes of semantics and set theory were important in the development of the field. The reason for working on the paradoxes of any field is not only to secure a foundation. The deeper reason is that by introducing, discarding, and clarifying the concepts that lead to paradox we are lead to the central ideas and questions of the field.

We see from **The Liar** that the paradoxes are still a source of inspiration in logic. The book is a new, exciting contribution to the study of truth. Its ideas might become important in the intense discussion of foundational issues in semantics by philosophers, linguists, logicians, and computer scientists. It can be read not only as a contribution to the philosophy of language, but also as an interesting application of a new theory of sets. It contains interesting theorems, and in turn it will stimulate purely mathematical work. What I plan to do here is to discuss some earlier responses to the Liar and also Barwise and Etchemendy's proposal concerning it. Of course, I will be unable to do either of these in much detail. In addition, I will comment extensively on the set-theoretic foundations of Barwise and Etchemendy's work because they are mathematically significant.

The basic elements of the Liar sentence are *self-reference, negation*, and *truth*. Each of these concepts is delicate and controversial; interaction of all three makes the Liar the challenge that it is. Furthermore, each could conceivably motivate its own treatments of the paradox. For example, one might feel that the central problem with the Liar is its self-reference. (Barwise and Etchemendy use the term *circularity*.) Of course, to do this effectively, one has to face the fact that there are sentences making use of self-reference, for example this one, which are nonparadoxical. Still, the self-referential aspect of the Liar seems to be part of what makes it humorous. And one can envision treatments of the Liar in which the self-referential aspect is prominent. However, it should be mentioned that there are Liar-like paradoxes in which the self-referential aspect is not so plainly syntactic as in the Liar sentence. An example, derived from Kripke [7] might be where two people engaged in an argument each say, "Everything you say is false." The circularity here is a feature of the entire discourse, not of any single sentence.

It is possible to look instead to negation as the source of paradox. The idea here is that negation is more complicated than it seems at first glance, much more complicated than the simple treatment of standard logic would have it. Early on, Bochvar **[4]** gave a treatment of the Liar which used three-valued logic to say that the Liar has the truth value "meaningless." A problem with this is that the paradox could resurface in the form of the *Strengthened Liar*: "This sentence is either false or meaningless." (Think about it.) More seriously, it would be necessary to state clearly why the Liar should be meaningless. Ad-hoc reasoning concerning the Liar is disapointing because one has the feeling that a deeper and more useful analysis of the problematic concepts is possible. Despite the difficulties with Bochvar's proposal, we will see other ways in which facts concerning negation can be profitably applied to the Liar.

The most influential work on the Liar is based on the assumption that truth is the problematic concept. Tarski **[10]** holds that the lesson of the Liar Paradox is that if natural languages contain their own truth predicate, then they are incoherent. (Despite this negative assessment, Tarski made proposals concerning the definition of truth in *formal* languages. His positive contribution in this area is fundamental to modern logic.) Nevertheless, his work allows one to isolate a fragment of language which is big enough to be interesting but small enough to be unaffected by the Liar paradox. Simplifying and distorting his position somewhat, the idea is that declarative sentences of a language come in an infinite hierarchy. The bottom level is the "object language" consisting of simple sentences which don't involve a predicate of truth. Next comes the "metalanguage" containing the object language together with sentences which assert that object language sentences are true or false. And after that comes the "metameta-language" and so on. This way of looking at things leads to the view that the Liar sentence is ambiguous (or rather has representatives on all of the levels). For example, the first representative is the sentence "This sentence is not true-in-the-object-language." This sentence does not even belong to the object language; therefore it is true. It should be noted that the Liar as we normally think of it simply would not occur in the hierarchy since it involves the truth predicate for a class of which it is a member. This approach avoids the paradox by a device which is artificial (though clever). This is a reason to be dissatisfied with it. For other reasons, cf. **[6, 7]**, and The Liar.

Kripke **[7]** also proposes a solution in which an analysis of truth is central. He consider a single language which contains a truth predicate. Now a language always has many possible interpretations. In the case at hand, an interpretation involves a space of sentences and a partial function from the sentences to truth values. (This use of partiality is a novelty.) Kripke makes use of *sequences of interpretations*. The first interpretation has the truth function completely undefined. The farther out in the sequence we get, more and more sentences are true, and more and more are false. Kripke argues that our intuitions are best modeled by an interpretation which is the *limit* of some such sequence. He constructs various sequences, including a natural one in which the Liar sentence has no truth value. And there is no predicate of "having no truth value" threatening a strengthened Liar. Kripke's work also enables one to give an explanation of what makes the Liar different from the Truth-teller sentence "This sentence is true."

I should mention two good sources which greatly expand on what I have so far only touched on. For a general survey of theories of truth and for an examination of the work of Tarski and Kripke, one should see Chapters 6–8 of the fine book by Haack [6]. In addition, Martin [9] is a collection of papers which were written after and in many cases influenced by Kripke [7].

Here is how Barwise and Etchemendy diagnose the Liar paradox:

> However described, the reasoning that makes the Liar look paradoxical has three distinctive stages. We first engage in a piece of metalevel argumentation which shows that the Liar cannot be true. Second, we objectify this conclusion and assume it to be a feature of the world, a feature that can influence truth and falsity. There is a clear move here from the realm of semantic facts to the typically nonsemantic domain that our statements describe. The third step involves using this newly discovered feature of the domain of discourse as premise for a further piece of metalevel reasoning, reasoning that shows the Liar to be true. Whence the paradox. (p. 175)

The cure they propose is to systematically keep track of the relevant facts as during the stages of the argument. At each stage we have a *situation*, a collection of facts. Before we consider the Liar sentence, we might suppose that we are in some initial situation s. When we then consider the Liar, we do not reason about some eternal sentence. What the Liar says to us at that moment is, in effect, "On the basis of s, this is false." Then we make the first step above, and we see that this is false. The second step is the subtle one. In it, we notice that our previous reflection has created a new fact. Once again, this fact is not that the Liar sentence itself is false, but rather that sentence quoted above is false. And in doing this we advance to a new situation s' which includes s and contains this new fact. At this time, we can reconsider the Liar. But now there is no paradox because this time it really means "On the basis of s', this is false." Our reasoning has changed the situation.

Now I turn to a short examination of the proposals that Barwise and Etchemendy actually make. They work with a collection of conceptual primitives which comes from work done in situation semantics. I shall describe the web of concepts informally, but I make an important caveat. Barwise and Etchemendy do not attempt in this book to justify their overall view of the world. Instead, they build a model of their basic concepts. Readers reluctant to embark on metaphysical excursions can happily stay at home with precise definitions of mathematical objects. This makes the book very enjoyable to read. However, at times I wish the authors had made a side trip into, among other things, the relation of situations to knowledge and belief. (There are long voyages in this direction in **[3]**, but one should be aware that situation semanticists may have changed their course in the years since it was written.) Even more, I would like to have seen an overall defense of "objectivist" model-theoretic work in semantics since there are many who do not feel that mathematical model building has anything whatsoever to do with human language. (A sharp criticism of this type is made, for example, by Lakoff **[8]**. His book also claims to synthesize an entirely different "cognitive semantics" with a different orientation to truth and meaning.) My criticism here should be interpreted constructively. Since **The Liar** shows how to model so many of our intuitions in such a nice way, it is a shame that the authors did not anticipate criticism that surely concerns them.

I shall not follow Barwise and Etchemendy's style of offering a set-theoretic model of the basic concepts. Instead, I give an informal presentation. I do this partly because I think the authors should have included something like this in **The Liar**, partly because it allows me to quickly get to the paradox, and partly to point out why a nonstandard set theory plays so prominent a role in the book.

The world contains objects which stand in different relations. Among the objects are people like you and me, and (significantly) abstract objects like *propositions.* Temporarily, think of a proposition as a complete meaning of a sentence. An *infon* is an abstract object which encodes the information that a given tuple of objects do or do not have a certain property. (In the book, the term *state-of-affairs* is used. "Infon" comes from [2].) For example, given two people a and b and the relation of seeing, there are two possible infons. One is that a sees b and the other infon, its *dual,* says that a does not see b. Another example concerns the truth predicate. Given a proposition p, and the relation of truth, we have two infons. One says that p is true, and the other says that p is not true. An infon is taken to be an object in its own right.

A *situation* is a set of infons. It is certainly possible for a situaton to contain both an infon and its dual. And the infons in a situation concerning the truth of propositions might be wrong. Only when neither of these undesired possibilities hold can a situation be *actual.* The authors are reluctant to elaborate on this in plainer terms, and in particular they don't address questions of how the concept of 'situation' relates to the informal concept. The most critical gap is an explanation of how situations change in the course of reasoning. In any case I think it is fair to say that a typical, intended example of an actual situation (in this technical sense) might be obtained by taking an actual situaton (in the nontechnical sense), fixing a point of view, and collecting together all of the salient facts.

For published examples see **[2]**. That paper is an important spin-off from the book, a situation-theoretic modeling of the notion of *shared information.*

A situation might contain *semantical facts.* These are infons which say that a given proposition is true or that it is not true. However, if an infon does not belong to a given situation, we have no right to conclude that its dual does belong. For example, there is a situation whose only infon is the fact that you are now reading this sentence. That situation is undecided about, for example, whether the President slept well last night.

The appropriate thing to ask about a situation is whether or not it is of a given *type*. Mathematically speaking, the types are modeled by the complete distributive lattice generated freely by the infons. If σ is a type corresponding to an infon, then a situation s is *of type* σ if σ belongs to s. This definition extends naturally to the case where σ is a conjunction or disjunction of simpler types. Given a situation and a type, we get a *proposition*. Thus a proposition says of a situaton that it is of a certain type.

I should stress that this describes propositions and it also defines truth. A proposition p says of a situation s that it is of a type σ. This proposition p is true just in case s is indeed of type σ. So every proposition is either true or false. The key point is that the truth or falsity of p is a new infon, and that infon *need not belong to s*. Furthermore, we can consider the proposition q that says about s that p is true. Even if p is true, q might well be false. In other words, even though p might be about s, its truth or falsity might not be not salient to s.

This understanding of propositions and truth is attributed to Austin. **The Liar** also models a different concept of a proposition influenced by the ideas of Russell, but I will not develop this point. The Austinian concept is clearly preferred by the authors.

Of course there are *sentences* and language more generally. Sentences here are taken as more or less syntactic, physical objects. They come without an interpretation. As such, it is not appropriate to ask what a given sentence means, or whether one is true or not. When a sentence is used, there is always a background situation. This is the situation that the sentence is *about*. Note that this means that the Liar as such is neither true nor false since it is a sentence. Nevertheless, for any situation s, there is a Liar proposition λ_s about s. It says that the fact that λ_s is false is part of s. Different situations give rise to different Liar propositions.

With all of the informal machinery in place, we can go on to discuss the Liar. The upshot on the Liar is that it is ambiguous. As a sentence, we repeat that it has no meaning until supplied with a situation. If s is an actual situation, then λ_s as defined above is false. For if not, λ_s would be true but the infon that says that λ_s is false would belong to s. Since s is assumed actual, λ_s would be false.

There are two points which must be made here. First, the fact that λ_s is false can never belong to s. (By the same reasoning.) Intuitively, once we recognize that the Liar proposition concerning a given situation is false, *we are in a new situation*. This *diagonalization* connects the Barwise-Etchemendy treatment of the Liar with ideas found throughout logic. The connection makes the work quite attractive. We can make a new situation s' by adding to s the fact that λ_s is false. And we can form a proposition μ which says about s' that λ_s is false. Now this μ is true. But $\lambda_{s'}$, the Liar proposition about s', is different from μ. Note that this semiformal argument is a way to make precise the less formal treatment of the paradox that I outlined above.

Second, our whole discussion up until now concerned the case when s is actual. It turns out that there are nonactual situations s in which λ_s is true.

I should mention a suggestion on how to read the book. Although the book is shorter than most of the ones reviewed in the *Bulletin*, one need not read the whole book to get the flavor. One could concentrate on Chapters 1–3, 8–10, 12, and 13. These constitute a full development of the topics of this review. I presented most of this material in four hours of lectures for the Michigan-Ohio-Ontario Logic Seminar. I think that in the setting of a seminar, this is the appropriate selection.

Of course, if you skip half the book you will miss many interesting things. In particular, you will omit the Russellian treatment of propositions. This is easier to understand than the Austinian account. You would also not see a proof-system based on the ideas of the book or the interesting development of a brand of model theory based on situations and propositions.

Incidentally, the writing in this book is very good. One probably should have studied the basics of model theory to read it, but it is not necessary to have a background, in, say, linguistics or the philosophy of language. There are many helpful exercises, too.

One of the most interesting aspects of the book is the influence of ideas from speech-act theory and pragmatics. This is clearest from Chapter 12 where the authors model the difference between denying that a proposition is true and affirming that its negation is true. Mathematically, this difference essentially calls for the lattice of types to be equipped with a complement operation. Philosophically, it means that the different uses of language are playing a role in determining the truth and meaning of sentences.

The distinction between negation and denial is one of a whole host of related facts concerning negation that could be brought to bear on paradoxes like the Liar. One succinct statement of some of the pragmatic component of negation might be found in Givón [**5**]:

> ... Negatives are uttered in a context where corresponding affirmatives have already been discussed, or else where the speaker assumes the hearer's belief in—and thus familiarity with—the corresponding affirmative.

Indeed, one can well imagine a purely pragmatic treatment of the Liar paradox, and that some other purely pragmatic generalizations might be modeled in a mathematically interesting way.

The remainder of this review will be devoted to a discussion of the mathematical foundations of the book. I would like to point out why the authors do not work with the conventional axioms of set theory, and then to present the new axiom that they use. I then consider the question of what it means to change the axioms of set theory. Finally, I have a few ideas on whether the change suggested by the book is warranted.

Let us see what happens when we try to set up the mathematical machinery needed in this treatment of the paradox. It is natural that each situation be some set, namely the set of infons it contains. An infon might be some ordered tuple. For example, given a situation s, the infon that λ_s, the Liar proposition about s, is true might wind up as an ordered quadruple $\langle infon, tr, \lambda_s, 1\rangle$. (The first *infon* might be some natural number which tells us that we are looking at an infon, *tr* a number standing for the truth predicate, and the number 1 for truth.) In the same way, λ_s itself might be a set such as $\langle prop, s, \langle infon, tr, \lambda_s, 0\rangle\rangle$. To repeat, λ_s would be equal to this set. The problem is that we need to prove that there is a unique such λ_s for each s. However, the usual axioms of set theory do not imply this. Indeed, it is a consequence of the Foundation Axiom (FA) that λ_s *never* exists. This is because by FA, each set has an ordinal rank. And by the way tuples are construed as sets, the rank of any tuple is strictly greater than the rank of each of its components. So no tuple can be a component of a component of itself.

At this point, one might want to reconsider the goal of modeling *about* and *part of* using the membership relation. (A desperate move might be to change the usual definition of ordered tuples, but this would not work either.) Instead, one might want to merely axiomatize the properties of those primitives and then build models from scratch. (This is the route taken in *situation theory*, a study which includes the examination of primitives used in situation based applications.) A second alternative, taken in **The Liar**, is to drop FA in favor of an axiom AFA which I describe below. In effect, *one changes the meaning of* "*set.*"

I am hesitant about this approach. This is not because FA is now a standard axiom of set theory and AFA is unfamiliar. It is rather because FA embodies the *iterative conception of set*. Although troublesome at times, the iterative conception is clearer and more compelling than any conception I now know of which suggests AFA. I stress that this is not to say that no new understanding is possible. I think that if one does emerge in the next few years, it will be due in part to the influence of the book under review.

Sixty years after being banished from set theory, non-wellfounded sets have returned. Set theorists had from the start wondered whether a set could be a member of itself, but the now-standard axiomatizations of set theory such as ZFC all rule such sets out by including FA. As a result, only a very small number of people since the 20s have ever entertained the possibility that non-wellfounded sets could exist. The current resurgence is mainly due to Peter Aczel **[1]**, who not only developed an elegant and unified theory but also showed how non-wellfounded sets can be applied to the study of communication and computation. He formulated the Anti-Foundation Axiom (AFA), which contradicts FA, and then he used AFA to build a model of Milner's Synchronous Calculus of Communicating Systems (SCCS). (Actually AFA had been previously studied by Forti and Honsell, and other anti-foundation axioms had been proposed by Boffa, Finsler, Gordeev, Scott, and others.) Aczel's course on non-wellfounded sets given at Stanford in 1985 prompted Barwise and Etchemendy to think about the uses of AFA in connection with the Liar. It turns out (serendipitously) that AFA is a very good tool for building models of all types of circular phenomena.

In order to state the AFA, we will need a few general definitions concerning sets and graphs. Let x be an arbitrary set. We form a graph G_x as follows. First we describe the vertex set of G_x. This is: the set x itself, all elements of x, all elements of elements of x, all elements of elements of .... Now given two vertices v and w, we put an edge $v \to w$ into G_x if $w \in v$.

Suppose we want to characterize the class of graphs G_x that arise in this way. First, they are all *accessible pointed graphs* (*apg's*). That is, they are all directed graphs, and they have a top vertex which can reach all other points by some finite path. Second, they have *decorations*. A decoration of an apg G is a function d defined on the nodes of G such that for all v, $d(v) = \{d(w) \colon v \to w \text{ in } G\}$. In the case of G_x, the identity function is a decoration.

What more can be said about the graphs G_x and about their decorations? FA tells us that each G_x is well-founded—it has no infinite paths. That implies that there are no loops $v \to v$, and more generally that there are no finite cycles $v_0 \to v_1 \to \cdots \to v_n \to v_0$. More generally, FA implies that if an apg has a decoration, then it is well-founded. Other axioms of set theory, namely Replacement and Extensionality, imply the Mostowski Collapsing Lemma: Every well founded apg has a unique decoration.

Now the Anti-Foundation Axiom may be stated:

(AFA) Every apg has a unique decoration.

For example, consider the apg consisting of a single vertex v and a single edge $v \to v$. Then AFA implies that there is a unique set with this structure. This set is traditionally called Ω. So $\Omega = \{\Omega\}$, and Ω is the unique solution of $x = \{x\}$.

We should mention before going any further that AFA is consistent. Every model of ZFC extends to a model of ZFC with FA replaced by AFA. The canonical extension is quite natural; cf. [**1**, Chapter 3]. There is nothing special about ZFC here. Indeed the same results hold for weaker theories such as KP.

We should also note that FA is not needed in the study of infinity. Neither is it required to carry out the foundational task of set theory. All of the sets that come up in modeling the natural numbers or real numbers are well-founded. So there is no mathematical need to rule out what Mirimanoff, who first defined well-foundedness, called "exceptional sets." If the AFA were to be adopted as an axiom instead of FA, none of your theorems would suddenly become false. None of your conjectures would be any easier to prove, either.

Before turning to the question of why anyone would want to assume AFA, we should first ask why it is so popular to assume FA. In the presence of the other axioms, FA gives us a clear picture of the universe of sets. We iterate the power set function along the ordinals to form a hierarchy of sets V_α in the usual way. Then FA is equivalent to the statement that an object is a set iff it belongs to some V_α. That is, every set belongs to the hierarchy obtained by iterating the power set operation along the class of ordinals. This axiom leads to a picture of the universe of sets as "created step-by-step from below."

This picture of the mathematical universe as generated in stages from the empty set (or even from atoms) is related to the view that the physical world is built from indivisible particles, or that the social world is composed primarily of independent individuals. This connection is the real cultural significance of FA in mathematics. It connects us with a deeply rooted atomistic paradigm that occurs throughout science. Conversely, to deny the iterative conception is to challenge "common sense." There is nothing wrong with this—indeed, the challenge to this paradigm seems to be one of the most important intellectual stirrings of modern times.

The authors of **The Liar** refer to Aczel's new "appealing alternative conception of sets" (p. 35). But Aczel's book contains no explicit discussion of the kind. Barwise and Etchemendy themselves come closest when they say that

> Aczel's conception of a set arises directly out of the intuition that a set is a collection of things whose (hereditary) membership relation can be depicted, unambiguously, by graphs of this sort [apg's]. The liberating aspect is that we allow arbitrary graphs, including graphs that contain proper cycles (p. 37).

It is not clear to me that this is a conception of a set, that it could really be motivated by someone who had not first considered the AFA.

What is needed most is a clear and persuasive conception of a set (or something else) under which AFA would be not only true but obviously true. The best possible genesis of such a conception would be some compelling new understanding of the physical and social worlds. However, we need not wait for this; there are other possibilities.

One place to look for a new conception is in the metaphors we use when speaking of sets and the relation of membership. The fact that we use *in* so often instead of *is a member of* is important because it suggests a spatial metaphor. We also speak of a set *containing* its elements, and we ask whether a set is *empty* or not. The iterative conception is consistent with the view that sets are containers. Under it, the relation of a set to its elements is that of a box to its contents, except that we identify all the empty boxes. Note that one box might be inside a second box, but no box can be inside itself. The point of all this is that a metaphor, detectable from language, can support a conception. It seems hard to find any metaphor having to do with sets, spatial or otherwise, under which the AFA is true. Without one, it will be hard to find an alternative to the iterative conception that is as appealing.

In the following passage, Barwise and Etchemendy make a point about modeling informal concepts with set-theoretic objects. In it, they disclose that perhaps they are interested not in set membership as such but rather in some other relation, one suggested by words like *about, constituent of,* or *involves.*

> Similarly, we have seen many features of our set-theoretic models that do not reflect theoretical commitments about the nature of the semantic objects modeled. For example, the fact that our models of propositions can have themselves as set-theoretic constituents (due to our use of AFA) is not meant to involve a similar claim about real propositions. What it reflects is our belief that propositions can sometimes be *about* themselves. (From the postscript to the second printing, emphasis in the original.)

(Presumably, though, some features of the models do reflect critical commitments. I have tried to isolate those commitments in my discussion of their overall view of the world. For example, there is an implicit commitment to a claim that for some or perhaps even every situation s, there is a unique Liar proposition λ_s about s.) However, since these concepts are abstract our intuitions are not clear. If we consider the world of all abstract objects under the relation of involvement, for example, then the uniqueness half of AFA is suspicious. For that matter, the existence part is also questionable.

AFA is an axiom whose motivation might come after its successful use, and certainly **The Liar** ought to stimulate a good deal of work. In the coming years we should expect to see technical results concerning AFA, further applications in computer science and linguistics, and deeper philosophical probing. Perhaps then will emerge a new conception of some sort of object (sets, perhaps) under which an anti-foundation axiom is obviously true.

References

1. P. Aczel, *Non-well-founded sets*, CSLI Lecture Notes no. 14, 1988.

2. J. Barwise, *Three views of common knowledge*, Theoretical Aspects of Reasoning about Knowledge. II (M. Vardi, ed.), Morgan Kauffmann, Publishers, 1988, pp. 365–380.

3. J. Barwise and J. Perry, *Situations and attitudes*, Bradford Books, MIT Press, 1983.

4. D. A. Bochvar, *On a three-valued logical calculus and its application to the analysis of contradictories*, Math. Sb. **4** (1939).

5. T. Givón, *Negation in language: pragmatics, function, ontology*, Cole, P. Syntax and Semantics, vol. 9, pp. 69–112, Academic Press, New York, 1978.

6. S. Haack, *Philosophy of logics*, Cambridge University Press, 1985.

7. S. Kripke, *Outline of a theory of truth*, J. Philos. **72** (1975), 690–716. Reprinted in Martin [**9**] 53–81.

8. G. Lakoff, *Women, fire, and dangerous things: What categories tell us about the mind*, Univ. of Chicago, 1987.

9. R. L. Martin, *Recent essays on truth and the liar paradox*, Oxford Univ. Press, New York, 1984.

10. A. Tarski, *The concept of truth in formalized languages*, Logic, semantics, metamathematics, Clarendon Press, 1956.

LAWRENCE S. MOSS
IBM THOMAS J. WATSON RESEARCH CENTER