Australasian Journal of Philosophy
Vol. 71, No. 1; March 1993

ANOTHER DISGUISE OF THE SAME FUNDAMENTAL PROBLEMS: BARWISE AND ETCHEMENDY ON THE LIAR

Graham Priest

I. Introduction

Trying to find a solution to the semantic paradoxes has been a perennial theme in logic this century. A comprehensive review of all the solutions suggested would fill many volumes. Nor has the stream of proposed solutions dried up of late; the last twenty years have seen a number of solutions of elegance and technical virtuosity that few previous proposals can match.

However, as I have argued elsewhere[1] the programme of trying to find a solution, that is, of showing that, and why, the paradoxical arguments are unsound is, in the words of Imré Lakatos, a degenerating research programme. The proliferation of solutions is, indeed, eloquent testament to this fact. Another mark of such a programme is that the proposed solutions, whilst solving many interesting technical problems, are dogged by the same fundamental problems. This is often hidden by the fact that the technical machinery of the proposed solution allows the problems to appear in a different guise each time; but in the end it becomes clear that they are the same problems — just in disguise.

A recent proposed solution to the semantic paradoxes is due to Barwise and Etchemendy[2] — hereafter, BE. It is technically sophisticated and very ingenious; but in the end it fits the familiar pattern, and so provides one more epicycle on the degenerating research programme. To establish this is the point of this paper.

II. Preliminaries

First, some preliminaries. In evaluating and comparing solutions of the liar paradox it helps to have a template against which to fit them. The following will serve here.[3] The liar concerns a sentence, α, 'p is false', where 'p' appears to refer to the proposition expressed by α. This is substituted in the T-schema 'q is true iff β', where β is any sentence and 'q' refers to the proposition that β to expresses. The result is a sentence 'p is true iff p is false' (τ). A contradiction is then inferred by *consequentia mirabilis* ($\alpha \equiv -\alpha / \alpha \wedge -\alpha$). Solutions are one (or both) of two kinds. The first objects to the inference *consequentia mirabilis*; the second, and by far the more common of the two, is the kind we are concerned with here. Solutions of this kind fault the T-schema, and in particular, claim that its instance τ does not express a true

[1] Priest [6, ch. 1], [7].
[2] See [2]. Subsequent page references are to this.
[3] See Priest [6, ch. 1].

proposition. A minimum condition for any formal solution of this kind must be to give a semantical interpretation of the language which delivers this result. This, in itself, is not terribly difficult, and can be done in numerous different ways. The crucial question is how philosophically satisfactory the interpretation and its ramifications are.

Just because the paradoxes *are* paradoxes, proposed solutions must fix on some intuitively acceptable principle and deny it. But any proposed solution can, after setting up some appropriate machinery, say: if it were not thus and so, contradiction would arise; hence it is thus and so. Thus, a solution that does not provide an independent rationale for locating the failure where it does, will appear arbitrary and unsatisfactory.[4]

BE are in agreement on this point. As they put it (p. 7), without an independent rationale:

> . . . the paradox remains paradoxical, despite the treatment. For the liar has forced us to abandon intuitively plausible semantic principles without giving us a reason beyond the paradox itself, to suspect their falsity. We see *that* they are false, without understanding *why*.

They go on to contrast such unsatisfactory proposed solutions — the majority, in fact — with those that do provide an independent reason, diagnoses as they call them; and claim that their solution to the liar is a diagnosis. I will argue that this is not so; in the end their own rationale for faulting certain intuitive principles is, exactly, the avoidance of inconsistency.

That BE's own proposed solution to the liar paradox falls foul of this point, is the first piece of evidence that it fits the familiar picture. But there is another, and much more important, reason. Solutions to the semantic paradoxes aim to show, appearances notwithstanding, that our semantic discourse is consistent. Almost invariably, proposed solutions introduce semantic concepts of a novel kind which allow for reformulations of the liar paradox (often called extended or strengthened paradoxes). In fact, these paradoxes are the genuine liar paradox in the novel contexts; that is, they are the boundary-violating constructions produced by the diagonal heuristic.[5] An heroic stance might be to say that the novel concepts are senseless, illegitimate, or whatever. But this would be quite self-refuting, since the concepts are a part of the very solutions proposed. The only other possibility is to relegate these concepts to a 'metalanguage', a more expressive language used for talking about the original discourse. But this move is equally self-defeating. For the original aim was to show that our semantic discourse is consistent; a solution which produces semantic discourse not within its own scope therefore fails. The problem is merely shifted to the discourse of the metalanguage.[6] As we will see, BE's solution fits the familiar pattern here too.

[4] And beg the question against dialetheism; see Priest [6, ch. 0].

[5] See Priest [7, section 4i].

[6] See Priest [8], and [6, ch. 1]. The latter shows why proposed solutions to the semantic paradoxes must fall foul of this situation.

III. Russellian Models

In fact, BE offer two solutions to the liar paradox. Both solutions apply the techniques of situational semantics;[7] and to give a formal account of the notion of proposition they require, they employ Aczel's beautiful machinery of non-well-founded sets (AFA).[8] The two solutions differ in the precise account of proposition they employ. The first of these is derived from Russell; the second from Austin. In the end, as we will see, BE argue that the Austinian solution is better. I will look at both solutions in turn.

It is not my intention here to give a comprehensive account of the wealth of material in BE's book; but to make this paper intelligible to those that have not read it I will summarise those parts of it that are relevant to the discussion, cutting a few simplifying corners where this does no harm. I should say, by way of preliminary remark, that the distinction between sets and proper classes plays some rôle in the account. However, it does not play a major role until the end; and there the only important thing to know is that only some collections are sets, namely, those that can be members of other collections. Thus, if X is a proper class, there can be no collections of the form <X,a>, etc.

The first notion of proposition BE employ is the Russellian one. Atomic propositions (or at least their set-theoretic representations) are of the form <F, b>. Here, F is a 'fact' of the form <P, $a_1, \ldots, a_n$>, where P is an n-place property and $a_1, \ldots, a_n$ are objects. b is either 0 or 1, and is a 'sign bit', indicating the polarity of the fact. I will use '-' to denote a monadic operator toggling 1 and 0. It should be observed that one kind of atomic proposition is of the form <<T, p>, b> where T is the one place property of truth, and p is a proposition. p may be this very proposition itself or some other proposition of which it is a constituent. (This is where Aczel's theory of non-well-founded sets is employed.)

Molecular propositions are generated by the recursion:

> If p and q are propositions then <∧, {p, q}> and <∨, {p, q}> are propositions (which we will write as p∧q and p∨q respectively).

Negated propositions can be introduced by definition, using the sign bit. (The book itself uses overlining for negation. For typographical purposes I will use -.) If p is an atomic proposition, <F, b>, – – p is <F, -b>; if p is q∧r, -p is -q∨-r; if p is q∨r, -p is -q∧-r. Since -p is the dual of p, it is clear that – – p=p. As the book does, I will write [Tp] for the proposition <<T, p>, 1>, and [Fp] for -[Tp]. As usual, p⊃q is just -p∨q.

Of course, since propositions are not well-founded, it is not immediate that the above is all legitimate. By applying Aczel's theory, it can, however, be shown to be so. Similar remarks apply at several points in what follows, and I shall take them as read.

We can now define the notion of model appropriate for Russellian propositions, and, crucially, the notion of being true in such a model. (Warning: what I call

[7] Barwise and Perry [3].
[8] Aczel [1].

being true, BE call *being made true*. They call a proposition, p, true in a model if [Tp] is made true. Despite the risk of confusion, I shall use my terminology since it is more standard.)

A *situation* is just, in effect, a collection (set or class) of Russellian atomic propositions. We define what it is for a proposition, p, to be *true* in a situation, s, ($s \models p$) by the following recursion:

If p is atomic, $s \models p$ iff $p \in s$
If p is of the form $q \wedge r$, $s \models p$ iff $s \models q$ and $s \models r$
If p is or the form $q \vee r$, $s \models p$ iff $s \models q$ or $s \models r$

(Strictly speaking, and for technical reasons, the situations in the above definition must be sets. However, once $\models$ is defined for sets, we can define it generally as follows: $s \models p$ iff for some sub*set* s' of s, $s \models p$. Then all collections, s, satisfy the above conditions.)

If p is not true in s it is *false* in s. A situation, s, is *coherent* if there is no p such that $p \in s$ and $-p \in s$. (As may easily be checked, if s is coherent then for no proposition, p, $s \models p$ and $s \models -p$.) A situation, s, is *almost semantically closed* iff for every proposition, p we have both of:

$s \models [Tp]$ iff $s \models p$
$s \models [Fp]$ iff $s \models -p$.

A *model* is any coherent semantically closed situation; and a *maximal model* is any model not properly contained in any model. It is maximal models that are, for BE, the set-theoretic representations of the real world. It should be observed that a maximal model may well not be complete (in the sense that for every proposition, p, either p or -p is in it). The liar itself provides a counter-example, as we shall see. Of course, it has to be shown that there are maximal models, but that is not too difficult.

Although the construction is somewhat unfamiliar, the result it produces is essentially a familiar one, as BE observe. Given any coherent situation, suppose that we assign one of the three values {t, u, f} to all propositions, p, by the following conditions:

p is t if $s \models p$
p is f if $s \models -p$
p is u otherwise

Then, as may easily be verified, the assignments are exactly those of Kleene's strong 3-valued logic, which is, therefore, the 'internal logic' of the model. Thus a model is essentially a Kripke fixed-point under the Kleene logic; and a maximal model is essentially a maximal fixed-point.

IV. The Liar: A Diagnosis?

We can now discuss the first of BE's solutions to the liar. Using non-well-founded sets, we can construct the liar proposition p=[Fp]. The instance of the T-schema, τ of section II, expresses the proposition [Tp]≡[Fp]. Given any (maximal) model m:

$m \nvDash [Tp] \equiv [Fp]$

since m is consistent (and *consequentia mirabilis* is valid in the internal logic of the model). It follows that $m \nvDash [Fp]$, for otherwise $m \vDash p$, so $m \vDash [Tp]$, and the instance of the T-schema would be valid. (For similar reasons, $m \models [Tp]$.) Thus, p (=[Fp]) is false in m, though [Fp] is not itself true in m. As BE put it: p is false, though its falsity is not a 'fact of the world'.

Now, as BE are the first to point out, there is something highly counter-intuitive about the claim that there is a false proposition whose falsity is not a fact. Why should we suppose that there could be such a thing? The only reason there is for the existence of such a thing is the stipulation that m be consistent. If this is not clear, merely consider what happens when we drop the requirement of consistency (coherence). The notion of a model still makes perfectly good sense (though the maximal model is rather uninteresting). But now, as may easily be checked, the 'internal logic' of the model is Dunn's 4-valued semantics for first degree entailment, rather than the strong Kleene 3-valued logic (and models are essentially the fixed points of Woodruff[9]). There is nothing to prevent the instance of the T-schema for the liar proposition being true in m. Moreover, if we now add a requirement that a model, m, be complete, in the sense that for every atomic proposition (and so every proposition) $p \in m$ or $-p \in m$, the internal logic of models is the logic LP[10] with corresponding fixed points. Moreover, it is now not difficult to find fixed points where [T-p] has the same value as -[Tp].[11] In these there are no similar ineffable semantic facts; for if p is false in m, -p is true in m, and so is [Fp].

Thus, that there are ineffable facts is purely an artifact of BE *imposing by fiat* the coherence condition (which is really a consistency condition since there is nothing incoherent about models not satisfying it, as we have just seen). Hence the 'explanation' of how it is that the liar argument does not lead to inconsistency is entirely circular, and no explanation at all. Thus it suffers from the first problem we noted in section II.

V. Semantic Ascent

But now let us turn to the more crucial question of whether the existence of ineffable facts is merely counter-intuitive. Is it? No; it is a good deal worse than that. According to BE the liar proposition is false; but this fact itself cannot be truly expressed. As they put it (p. 101): 'the Liar proposition is indeed not true, but there is no true proposition which expresses this fact'. In the mouth of anyone who endorses the Russellian solution, this is a classic case of self refutation! — and of just the kind we noted in section II. BE do indicate a way that one could go if pressed. This is to conceptualise the theorist's own discourse as 'metatheoretic' in

[9] See [9].
[10] Priest [5], [6, ch. 5].
[11] As in Dowden [4].

nature (p. 88f) and so, presumably, accessing a different set of propositions. But in this case we are back to the faithful device of a metalanguage again.

To hammer this home, note the following. Given a model m, there can be no propositional function, f, which defines in m the collection of propositions false in m, in the sense that:

$$m \models f(p) \text{ iff } m \not\models p$$

If there were, by the standard constructions, we could construct a proposition q=f(q), and substitution would immediately give us a (metatheoretic) contradiction:

$$m \models f(q) \text{ iff } m \not\models f(q),$$

which is impossible if the construction is consistent (which it is with respect to ZFC.) This is, after all, merely a version of Tarski's Theorem. Yet BE appear to have such a propositional function, $m \not\models p$! Thus, the only way that BE can avoid contradiction is by locating their own semantic discourse outside the discourse for which they are giving the semantics.

The situation here, then, is exactly the one we flagged in section II. The new notions provide for an 'extended' paradox, which forces metalinguistic ascent, on pain of inconsistency.

VI. Austinian Models

BE do not note the full seriousness of the problem here. Yet they do note that the situation is 'embarrassing', and so opt for a different notion of proposition. Let us now see whether the solution this provides is any better.

We can think of an Austinian proposition as essentially a pair $\langle s, p\rangle$ where s is a situation and p has the structure of a Russellian proposition. (Beware though: both s and p may have components that are propositions, and the propositions involved are Austinian, not Russellian. And since s is now required to be a member of something else, we must take it to be a set, not a proper class.) The second component is supposed to express the 'propositional content' of an utterance, whilst the first represents the situation which the utterance is *about*.

The next job is to define the appropriate notion of model for Austinian propositions. We can define a notion of truth by clauses similar to those which define truth in a situation in the Russellian case. Let Q be the proposition $\langle s, p\rangle$; then:

If p is atomic, Q is true iff $p \in s$.
If p is of the form $q \wedge r$, Q is true iff $\langle s,q\rangle$ is true and $\langle s,r\rangle$ is true
If p is of the form $q \vee r$, Q is true iff $\langle s,q\rangle$ is true or $\langle s,r\rangle$ is true

A proposition is false iff it is not true.

It should be noted that truth, as just defined, is internal to an Austinian proposition in a certain sense. For this reason the notion may be thought somewhat

Pickwickian. For example, if p is any atomic (Russellian) proposition and Q = <{p, -p}, p> then both Q and -Q (i.e. <{p, -p}, -p>) are true! Thus real truth must be truth where the situation involved is real in some sense. We must therefore define which situations those are. A situation (set or class), s, is a *partial model* iff it is coherent and for every proposition, Q:

if [TQ]∈s then Q is true (*1)
if [FQ]∈s then Q is false (*2)

A *maximal model* is any model not properly contained in any model. It is important to note that with the new notion of proposition, maximal models are complete, as can quickly be demonstrated. Moreover, it follows from this that the converses of *1 and *2 also hold. It is maximal models which are BE's candidates for reality. If m is some maximal model and sOm, they call s *actual* (with respect to m). Thus, we can say that <s,p> is *really* true (with respect to m) iff <s,p> is true and s is actual.

VII. The Liar: Another Diagnosis?

Given some maximal model, m, and situation s⊑m, we can construct a liar proposition Q = <s, [FQ]> by applying the machinery of non-well-founded sets. The instance of the T-schema expressed by τ of section II is now <s, [TQ]≡[FQ]>. Can this be true? Calculating, we get:

<s, [TQ]≡[FQ]> is true iff <s, [TQ]⊃[FQ]> and <s, [FQ]⊃[TQ]> are true
iff <s, -[TQ]∨[FQ]> and <s, -[FQ]∨[TQ]> are true
iff <s, [FQ]> and <s, [TQ]> are true
iff [FQ]∈s and [TQ]∈s.

And this cannot be so.

But does this provide any independent understanding of why this instance of the T-schema fails? Why cannot [TQ] and [FQ] both be in s? In fact, this seems to be guaranteed twice over. The first reason is that BE *stipulate* that m must be consistent. But this can no more provide a decent explanation of why inconsistency is avoided than in the Russellian case. The second reason is that *1 and *2 appear to entail that this situation cannot arise. But a little thought shows that they do so only if something cannot be both true and false. And the something in question in this case is, of course, Q, the liar proposition. Again, the explanation as to why τ does not express a true proposition is entirely circular.

Of course, since both propositions and truth are represented within set-theory, the contradiction of the liar being true and false is represented as a set theoretic contradiction: [FQ]∈s and [FQ]∉s. Thus someone might argue that the consistency of set-theory provides an independent rationale for the failure of this instance of the T-schema. This would not be a terribly happy argument, however; first, because the consistency of set-theory is threatened by exactly the same phenomenon as threatens the consistency of semantics: the paradoxes of self-reference. Thus, to assert the consistency of set theory is already to presuppose that diagonal paradoxes have been defused. Secondly, and more importantly, the consistency of set-theory can be used as an argument for the consistency of the liar only if the representation of semantic notions in set theory is adequate. But if set theory is consistent and the liar is not, this just shows that the representation is inadequate, since pertinent properties are not mirrored in the representation. Hence it is the adequacy of the representation which now begs the question, and prevents the explanation being non-circular.

VIII. Semantic Ascent, Again

Let us now turn to the second and more important question of whether this solution requires, essentially, 'metalinguistic ascent'. The reader may have noticed that *1, promoted to an equivalence — or *2 promoted to an equivalence, which is the same thing, given that m is maximal and consistent — is itself a version of the T-schema; and may have wondered how this avoids contradiction. To see this, take a liar sentence Q = <s, [FQ]>, and substitute; we then get the following reasoning:

[FQ]∈m iff Q is false (*2 and its converse)

iff <s, [FQ]> is false
iff [FQ]∉s.

Let us call this *Diag*. By *Diag*, either [FQ]∈m-s or [FQ]∈s-m. In the first case, Q is false, and its falsity is 'a fact in m'. In the second case, Q is true, but not really true (since it is not about an actual situation).

BE's discussion focuses on the first case. (Indeed, they say virtually nothing about the other case, though it seems to me to be just as interesting: why can one not make a statement about a non-actual situation?) They describe this by saying that the liar 'diagonalises out' of any actual situation. Q is not in s, but is in some more generous actual situation. This suggests that paradox will arise if we consider the total actual situation; the liar cannot, by definition, diagonalise out of this. So let m be any maximal model, and consider its liar proposition, $\underline{Q}$ = <m, [FQ]>. Using this in *Diag*, we now generate a contradiction [FQ]∈m and [FQ]∉m. Q, the global liar, is the form of the extended liar that does the damage in this construction.

Set theory to the rescue! We cannot, in fact, construct the global liar since m is, in general, a proper class. (This is where the distinction becomes crucial.) Thus this evasion of the semantic contradiction piggy-backs upon a purported solution to the set theoretic contradictions. The set/class distinction is a highly counter-intuitive one, and as a solution to the set-theoretic paradoxes it will not work.[12] Still, leave this aside. If we apply this distinction conscientiously it blocks the above argument. But that is not all it blocks. For it means that there are no propositions about the global situation at all. Nor is there any proposition attributing a property to the global situation, since this would require m to be a member of something, too. And

[12] See Priest [6, ch. 2].

it won't do to say that this distinction is merely an artifact of the modelling process which carries no real weight; for then the supposed solution to the global liar is equally an artifact, and carries no real weight. So we must take the distinction seriously.

But then what are we to make of the fact that BE make numerous claims about the global situation? The book is replete with statements about maximal models. Thus, the solution is self-refuting, in just the way we observed in section II. Indeed, BE's own description of the situation is as clean a one-line self-refutation as one can get: '. . . while the world is as total as one could want, we cannot, in general, make statements about the world as a whole' (p 154)!

BE are aware of this further embarrassing situation, and make some suggestions as to how to overcome it in the appendix. These involve making the world, m, a set, and thus a fit subject of a proposition (without at the same time reinstating contradiction). One way to do this is just to limit the class of propositions, and so the world, w, to some subset of propositions — say to those of some fixed rank (when AFA is modelled in the cumulative hierarchy). *Diag* then just shows that the 'global liar' Q = <w, [FQ]> is not a proposition. But this is disingenuous. What this means, of course, is that Q is not a proposition of that rank (or whatever). Structurally, it is still a proposition, just one larger than the arbitrary cut-off point.

A more elegant way is suggested by BE themselves (p. 188f). They show what they call the Reflection Theorem: for any maximal Russellian model, m, there is a set, m (a mirror), such that for any sentence of a reasonably generous language, the (Russellian) proposition it expresses is true in m iff the (Austinian) proposition it expresses about m is true. We may thus take m to 'represent the whole world'. Again, this suggestion does not seem to get very far: for it is propositions about maximal Austinian, not Russellian, models for which we were supposed to be finding an *ersatz*. If we are going back to Russellian models then the problems lie elsewhere, as we have seen. (And as far as I can see, there in no hope of proving an analogue of the Reflection Theorem for Austinian models.)

In any case, neither of these approaches solves the problem; they just cloak it. For the problem was to make sense of the talk of (Austinian) maximal models. And neither construction gets rid of this, but merely adds *further* talk (of propositions of rank κ, mirrors, or whatever). Indeed, the very statement of the Reflection Theorem, which underlies the supposed adequacy of BE's proposal, makes sense only if *bona fide* talk of maximal models makes sense.[13]

Thus, BE have little option but to locate their own discourse in a stronger metatheory. Here, this amounts to a two set-theory policy: one formal set-theory to provide the semantic machinery of propositions, and a stronger and informal one to talk about this.[14] At any rate, they can avoid paradox and self-refutation only by semantic ascent.

[13] The situation here is quite typical of what happens when set theorists suggest some *ersatz* for classes. See Priest [6, ch. 2].

[14] This ultimately unsatisfactory device is often used unconsciously in this sort of context. See Priest, [6, ch. 2].

IX. Conclusion

We have now looked at both of BE's solutions to the paradoxes. As we have seen, they fit the familiar pattern. Not only do their faultings of certain principles fail to have independent rationale, but if their solutions are to avoid inconsistency they are impaled on the dilemma of self-refutation or metalinguistic ascent; the latter being ultimately just as self-undercutting. Their solutions are therefore in the same boat as all the others — a boat that appears to be sinking.[15]

University of Queensland

Received October 1991
Revised December 1991

REFERENCES

1. P. Aczel, *Non-well-founded Sets* (CSLI Lecture Notes, #14) (Stanford: Centre for the Study of Language and Information, 1988).
2. J. Barwise and J. Etchemendy, *The Liar* (Oxford: Oxford University Press, 1987).
3. J. Barwise and J. Perry, *Situations and Attitudes* (Cambridge, MA: Bradford Books, 1983).
4. B. Dowden, 'Accepting Inconsistencies from the Paradoxes', *Journal of Philosophical Logic* 13 (1984) pp. 123-130.
5. G. Priest, 'Logic of Paradox', *Journal of Philosophical Logic* 8 (1979) pp. 219-241.
6. G. Priest, *In Contradiction* (The Hague: Martinus Nijhoff, 1987).
7. G. Priest, 'Unstable Solutions to the Liar Paradox' in J. Bartlett and P. Suber (eds), *Self-Reference* (The Hague: Martinus Nijhoff, 1987) pp. 145-175.
8. G. Priest, 'Boolean Negation, and All That', *Journal of Philosophical Logic* 19 (1990) pp. 201-215.
9. P. Woodruff, 'Paradox, Truth and Logic', *Journal of Philosophical Logic* 13 (1984) pp. 213-232.

[15] I would like to thank the anonymous referees of the Journal for helpful comments.