GRAHAM PRIEST

THE LOGIC OF PARADOX

'Indeed, even at this stage, I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from consistency.'

WITTGENSTEIN 1930, *Philosophical Remarks*, p. 332.

0. INTRODUCTION

The purpose of the present paper is to suggest a new way of handling the logical paradoxes. Instead of trying to dissolve them, or explain what has gone wrong, we should accept them and learn to come to live with them. This is argued in Sections I and II. For obvious reasons this will require the abandonment, or at least modification, of 'classical' logic. A way to do this is suggested in Section III. Sections IV and V discuss some implications of this approach to paradoxes.

I. PARADOXES

I.1. The logical paradoxes (which are normally subdivided into the set theoretic ones such as Russell's and the semantic ones such as the liar) have been around for a long time now. Yet no solution has been found. Admittedly, the liar paradox, which has been known for over 2000 years, has often been ignored as a triviality unworthy of serious consideration. (Though medieval logicians such as Buridan by no means regarded it as such.) However, in the last three quarters of a century, there has probably been more intensive work done in trying to find a solution for the logical paradoxes than on any other topic in the history of logic. The roll call of those who have tried to find a solution reads like a logician's honours list. Yet no widely accepted solution has been found.

I.2. Of course, we know how to avoid the paradoxes formally. We can

Journal of Philosophical Logic 8 (1979) 219–241. 0022–3611/79/0082–0219$02.3

avoid the semantic paradoxes, e.g., by a hierarchy of Tarski meta-languages, and the set theorectic ones, e.g., by the class/set distinction of von Neumann. But these are not solutions. A paradox is an argument with premises which appear to be true and steps which appear to be valid, which nevertheless ends in a conclusion which is false. A solution would tell us which premise is false or which step invalid; but moreover it would give us an *independent reason* for believing the premise or the step to be wrong. If we have no reason for rejecting the premise or the step other than that it blocks the conclusions, then the 'solution' is *ad hoc* and unilluminating. Virtually all known 'solutions' to the paradoxes fail this test and this is why I say that no solution has yet been found.

I.3. This massive failure on the part of the logical community suggests that trying to solve the paradoxes may be the wrong thing to do. Suppose we stop banging our heads against a brick wall trying to find a solution, and accept the paradoxes as brute facts. That is, some sentences are true (and true only), some false (and false only), and some both true and false! ('This sentence is false' and 'The Russell set is a member of itself' are paradigm examples of such paradoxical sentences). Of course, this requires giving up Aristotle's dictum 'The firmest of all principles is that it is impossible for the same thing to belong and not to belong to the same thing at the same time in the same respect' (*Metaphysics* Γ, 3(1005^{b} 19–23). Still, all progress depends on parting with tradition in some way or other.

II. SEMANTIC CLOSURE

II.1. If my argument for accepting the paradoxes were merely that no one had yet solved them, then my position, though plausible, would not be very tempting. However, I believe that there are theoretical reasons why paradoxes must be accepted; must, that is, if we are to get to grips with the notion of mathematical provability (in the naïve sense).

II.2. Proof, as understood by mathematicians (not logicians) is that process by which we establish certain mathematical claims to be true. In other words, suppose we have a claim (for the sake of definiteness say, a claim of arithmetic) whose truth we wish to settle. We look for a proof or a refutation (i.e., a proof of its negation). But a proof from what? Presumably from other claims we already know to be true. We can, of course, ask how we knew these claims to be true, and it may be because we have proofs of them. However, on pain of infinite regress, we cannot go on doing this indefinitely. Sooner or later we must come to claims we know to be true without a proof, where the question of proof does not, as it were, arise. These are axioms in the old-fashioned sense: self evident truths. (Why they are self evident, I need not go into. It suffices that there must be some such things.) In our present case these are presumably facts about numbers, such as that every integer has a successor different from anything gone before or the basic facts about addition. These are the sorts of things that children become familiar with when they learn to count and to do arithmetic. (We can, of course, look for 'proofs' of these axioms in some foundational system such as *Principia Mathematica*. However, these 'proofs' are not proofs in the sense we are concerned with – means of coming to know that the things proved are true.)

Thus we see that we establish claims of mathematics, if they are not axioms, by proving them (in the naïve sense) from those axioms. This all seems obvious to the point of banality. However, it runs us straight into a major problem. For there seems to be no doubt that this procedure could be formalized. The axioms could, in principle, be written in a formal language and the proofs set out as formal proofs. The formal system that resulted would encode our naïve proof methods. Moreover, there seems no reason to doubt that all recursive functions would be representable in the system. For certainly all recursive functions are naïvely definable. However, according to Gödel's incompleteness theorem in any such formal system there will be sentences that are neither provable nor refutable – at least if this set of axioms is decidable. If this were all there is to Gödel's theorem, the result might be surprising but not particularly worrying. The incompleteness of the formal system would merely show that there were mathematical problems beyond the powers of our proof procedures to settle. But this is not all there is. For some of these unprovable sentences can be *shown* to be true, i.e., proved in the naïve sense. But the formal system was constructed in such a way that it encoded our informal proof methods. So there can be no such proof. Gödel's theorem presents an epistemological problem that has never been squarely faced. How is it to be resolved?

One way out of the problem is to accept that the set of axioms is not recursive. Thus, assuming Church's thesis there would be no effective way of deciding whether something was self evident! This is obviously not on. Another possibility is that our naïve proof procedures are not formalizable. This seems implausible, though a case can be made out. Gödel (1947) has suggested that we may from time to time increase our stock of axioms by adding one which has inductive support (in the sense that it implies many things known to be true and nothing known to be false). On this picture we may suppose that our naïve proof methods at a certain time, say t, can be formalized by a system S. However, it may happen at a later time that to encode our naïve proof procedures we require another system S' formed from S by adding to its axioms a sentence A from which we can prove (in S) a number of true things and nothing false. However plausible this account is, it does not solve the problem. For consider one of the sentences that is independent of S but informally provable. Then this is already provable at t. We do not have to collect inductive evidence for it to see whether it should be added to the axioms of S. It can be established deductively by means already to hand. Hence the problem now appears with respect to our naïve proof methods at time t.

II.3. We have then, found no way out of the problem. Yet Gödel's suggestion does raise the question of how exactly it is that we are able to prove these independent sentences. This we will now investigate. We will see that it provides an answer to the problem.

Suppose that P is the formal system which encodes our naïve proof ability, (possibly at a certain time) and let g be the (code number of the) Gödel-sentence

$$\neg \exists x \, \mathrm{Prov}(xg)$$

where Prov (xy) is the primitive recursive predicate 'x is the code of a proof (in P) of sentence y'. Now, g is not provable in P, and this can be proved purely syntactically. However, the question is not how g is shown to be independent but how g is shown to be true. For g *can* be shown to be true. Usually it is said (with a wave of hands) that g 'expresses its own unprovability' and hence is true. When this rather vague hint is spelt out precisely, we obtain the following argument.

(1) $\exists x \, \mathrm{Prov}(xg) \Rightarrow \ulcorner \exists x \, \mathrm{Prov}(xg) \urcorner$ is true

(2) $\Rightarrow g$ is provable

(3) $\Rightarrow g$ is true

(4) $\Rightarrow \neg\exists x \text{ Prov}(xg)$

Hence $\neg\exists x \text{ Prov}(xg)$.

Step (2) depends upon the fact that 'Prov' really does represent the proof relation. Step (3) depends on the fact that whatever is provable in P is true. Steps (1) and (4) follow from the Tarski bi-conditional $T\ulcorner\psi\urcorner \longleftrightarrow \psi$ (where $\ulcorner\psi\urcorner$ is the code of ψ).

Now the important point to note about this reasoning is that it involves an essential *detour* through P's meta-language. The use of the notion of truth, its properties and its relation to provability are essential to the proof. We can not prove g in P, but we can prove it in P's meta-language.

II.4. Thus, we see that our naïve notion of proof appears to outstrip the axiomatic notion of proof precisely because it can deal with semantic notions. Of course, we can formalize the semantics axiomatically but then naïvely we can reason about the semantics of *that* system. As long as a theory can not formulate its own semantics it will be Gödel incomplete, i.e., there will be sentences independent of the theory which we can establish to be true by naïve semantic reasoning.

II.5. What happens, however, if we take a theory which is semantically closed? (A theory is semantically closed if it can formulate its own semantics.) If it is semantically closed, there is nothing to stop us carrying out the reasoning of Section II.3 in the system itself. For example, consider the following theory PS. The language of PS is that of Peano arithmetic with the additional two place predicate 'Sat (x, y)'. (Intuitively, this is going to be read 'the (finite) sequence x satisfies the formula with code number y'.) We can code up the notions of formula, ith variable, finite sequence of numbers, ith member of a sequence, length of a sequence, in the usual way. The axioms of PS are then those of Peano arithmetic plus all instances of the Tarski satisfaction scheme:

$$(\forall x)\,(x \text{ is a sufficiently long finite sequence} \rightarrow (\text{Sat}\,(x\ulcorner\psi\urcorner) \longleftrightarrow \psi'))$$

where ψ' is ψ with every variable 'v_i' free in ψ replaced by x_i (the ith member of the sequence x) and 'x is sufficiently long' is '$(\forall i)$ if the ith variable occurs free in ψ then the length of x is greater than i'. (Alternatively, this scheme could be replaced by a finite number of axioms: one for each atomic predicate of PS – including 'Sat' – and one for each truth function and quantifier. For details of this sort of construction see Hilbert and Bernays (1939, pp. 334–5). All instances of the satisfaction scheme would then be provable.) Now PS is clearly semantically closed and we can perform the sort of reasoning we used in Section II.3 to prove $\neg \exists x \text{Prov}(xg$ *in PS itself* (where 'Prov' is now the proof predicate of PS).

II.6. We are now in a position to solve the problem posed in Section II.2. We saw that the problem arises because we can establish sentences as true that are not provable in some particular system. We now see that this problem is avoided by using a semantically closed formal theory. For the reasoning that is used to show that this sort of sentence is true can be represented *within* a semantically closed theory. Of course, this also shows that Gödel's proof of the existence of independent sentences breaks down. For the reasoning gives a proof of the 'independent' sentence in the theory. (Whether there are independent sentences is a topic I will return to in Section IV.13.)

We might well ask where exactly Gödel's proof goes wrong. The place is not difficult to locate. For, of course, his proof works only if the theory under consideration is *consistent*. It is well known that semantically closed theories are *inconsistent*. (See, e.g., Fraenkel Bar-Hillel and Levy (1973, p. 312). It is in this way that Gödel's result is avoided.

We have seen then that the way to avoid the problem – in fact, the only reasonable way – is to accept that the correct formalization of our naïve methods of proof must be a semantically closed and inconsistent theory. In fact, any formal analysis of our naïve methods of proof must use a semantically closed and inconsistent theory or be inadequate.

At this point it might be objected that an inconsistent theory can not possibly codify our naïve proof methods. For in an inconsistent theory everything is provable. However, this objection presupposes that the underlying logic of the theory is 'classical'. This will be rejected in Section III.1. The fact that there are pathological 'classical' proofs (using, e.g., the rule $(A \wedge \neg A)/B)$) of anything in an inconsistent theory just shows that the logic of naïve proofs is not classical.

II.7. However, the fact that a semantically closed theory is inconsistent is important. For, in fact, the inconsistencies generated by semantical closure are precisely the semantical antinomies. Since Tarski (1936) it has been clear how semantic closure generates the semantic paradoxes. This then is the theoretical argument I referred to in Section II.1 for having to come to accept certain contradictions. For any acceptable analysis of our naïve notion of proof requires the use of a semantically closed and therefore inconsistent theory. Hence any adequate analysis of the naïve notion of proof will require us to accept the semantical antinomies as facts of life.

II.8. Although much of what I have said in this section is technical, the spirit is easy to grasp. The formal logician is essentially an applied mathematician. It is his job to construct mathematical systems which model (in the physicist's sense, not the logician's) some natural phenomenon. The phenomenon the logician is particularly interested in, is normal (naïve) reasoning carried out in a natural language. (After all, mathematicians use ordinary (not formal) languages – even if they are appendixed by certain technical terms – and ordinary reasoning). And the mathematical systems he uses are formal languages, mathematical semantics, etc.

Now it is a standard view due to Tarski that natural languages are both semantically closed and contain paradoxes. As he puts it:

> If we are to maintain the universality of everyday language in connection with semantical investigations, we must, to be consistent [*sic*!] admit into the language in addition to its sentences and other expressions, also the names of these sentences and expressions, and sentences containing the names, as well as such semantic expressions as 'true sentence', 'name', 'denote', etc. But it is presumably just this universality of everyday language which is the primary source of all semantic antinomies, like the antinomies of the liar and of heterological words.
>
> TARSKI (1944)

This is a view which I endorse. (In fact, I have claimed elsewhere, Priest (1974, Chapter 5, Section 1), that all the logical antinomies are due to semantic closure. But that is another story.) Hence, although a semantically open formal theory may be an adequate model for certain limited purposes, a semantically closed formal theory – with paradoxes – is required in general.

III. THE LOGIC OF PARADOX (LP)

III.1. In the two previous sections I have argued that we will have to learn to handle systems with contradictions in them. If we are to do this we will have to relinquish 'classical' logic. For contradictions have a horrible way of infecting classically formalized theories. The whole population of mathematical sentences becomes stricken with contradiction, rendering them unfit for work. However, if we can isolate the paradoxes and prevent them from contaminating everything else, we will be alright. In this section I will construct such a logic. However, I wish to emphasize that the arguments of the previous two sections are independent of the fate of the logic constructed. Even if it turns out to be unsatisfactory for some reason, the foregoing considerations still hold good.

III.2. In fact, systems that are not wrecked by contradictions have already appeared in the literature. Both Routley (1977) and da Costa (1974) have argued that such logics would allow us to investigate inconsistent but non-trivial theories. However, none of the systems proposed have the simple and intuitively plausible semantics of LP. For as well as providing the motivation for constructing such a logic, the preceding sections actually suggest a way to do it. This ensures that the logic has very intuitive semantics.

III.3. Classical logic errs in assuming that no sentence can be both true and false. We wish to correct this assumption. If a sentence is both true and false, let us call it 'paradoxical' (p). If it is true and not false, we will call it 'true only' (t) and similarly for 'false only' (f). However, having made this assumption, we shall continue to reason normally.

III.4. A sentence is true iff its negation is false. Hence the negation of a true and false sentence is false and true, i.e., paradoxical. The negation of a true only sentence is false only. (If its negation were true, it would have to be false.) Similarly the negation of a false only sentence is true only. We could record these results in the following table.

$\neg$	
t	f
p	p
f	t

III.5. Similar reasoning gives the following table for conjunction:

$\wedge$	t	p	f
t	t	p	f
p	p	p	f
f	f	f	f

I will do just a couple of examples. If A is t and B is p, then both A and B are true. Hence $A \wedge B$ is true. However, since B is false, $A \wedge B$ is false. Thus $A \wedge B$ is paradoxical.

If A is f and B is p, then both A and B are false. Hence $A \wedge B$ is false. If $A \wedge B$ were true as well, then both A and B would be true, but A is false only. Hence, $A \wedge B$ is f.

III.6. Reasoning in a similar way we can justify the table for disjunction.

$\vee$	t	p	f
t	t	t	t
p	t	p	p
f	t	p	f

Alternatively, '$A \vee B$' can be defined as '$\neg(\neg A \wedge \neg B)$'.

We can define '$A \rightarrow B$' as '$\neg A \vee B$' and '$A \longleftrightarrow B$' as '$A \rightarrow B \wedge B \rightarrow A$'. This gives the following tables:

$\rightarrow$	t	p	f
t	t	p	f
p	t	p	p
f	t	t	t

$\longleftrightarrow$	t	p	f
t	t	p	f
p	p	p	p
f	f	p	t

To what extent '$\rightarrow$' can be held to be some form of implication, we will return in Section IV.4.

III.7. In fact, these are the matrices of Kleene (1952, p. 332 ff). However, his interpretation is very different to ours. The matrices for '$\vee$', '$\wedge$', '$\neg$', are also those of Lukasiewicz (1920). However, only 't' is designated in his system. We will designate both t and p since both are the values of true sentences. Formally let L be a propositional language whose set of propositional variables is P. Let v: $P \rightarrow \{t, p, f\}$ (i.e., v is an evaluation of the propositional variables). Let ν^+ be the natural extension of ν to all the sentences of L using the above truth tables. If Σ is a set of sentences of L, we define:

$$\Sigma \models A \text{ iff there is no } \nu \text{ such that } \nu^+(A) = f \text{ but for all } B \in \Sigma, \nu^+(B) = t \text{ or } p.$$

$$\models A \text{ iff } \phi \models A \quad (\text{i.e., for all } \nu, \nu^+(A) = t \text{ or } p).$$

III.8. THEOREM: A is a two valued tautology iff $\models A$.

Proof. The proof from right to left is immediate since every two valued evaluation is a three valued evaluation.

Conversely, if ν is a three valued evaluation, let ν_1 be the two valued evaluation formed by changing all p's to t's. By checking the truth tables, we can see that if $\nu^+(A) = f$, $\nu_1^+(A) = f$. The result follows.

The LP matrices together with an interpretation similar to the one I have given occur in Asenjo (1966). Also, although, he does not state it explicitly, he seems to take both t and p (his 0 and 2) to be designated values. However he does not take his semantics very seriously. For immediately after giveing the semantics he states that $\neg(A \wedge \neg A)$ (which he calls the law of contradiction) should not be a theorem of any system of logic suitable for formalizing inconsistent theories. However, from this theorem we can see that according to these semantics $\neg(A \wedge \neg A)$ is logically valid.

III.9. We have seen that the property of being a tautology is preserved in LP. However, the deducibility relationship is changed. One can check the following:

$A \models A \vee B$	$A, B \models A \wedge B$	$A \to B \models \neg B \to \neg A$
$A \to (B \to C) \models B \to (A \to C)$	$A \models B \to A$	$\neg A, \neg B \models \neg(A \vee B)$
$\neg A \to \neg B \models B \to A$	$\neg(A \vee B) \models \neg A$	$A \models \neg\neg A$
$\neg\neg A \models A$	$\neg A \models \neg(A \wedge B)$	$\neg(A \to B) \models A$
$A \wedge B \models A$	$A, \neg B \models \neg(A \to B)$	$A \to B \models A \wedge C \to B \wedge C$
$\neg A \models A \to B$	$A \to (A \to B) \models A \to B$	$A \to \neg A \models \neg A$.

However, it is easy to find evaluations which show that the following do not hold.

$A \wedge \neg A \models B$	$A \to B, B \to C \models A \to C$	
$A, \neg A \vee B \models B$	$A, A \to B \models B$	$A \to B, \neg B \models \neg A$
$A \to B \wedge \neg B \models \neg A$		

III.10. THEOREM: If A and B have no propositional variables in common and if B can take the value f then $A \not\models B$.

Proof. Let ν be an evaluation such that $\nu^+(B) = f$.

Let ν_1 be like ν except that for all the propositional variables q occurring in A, $\nu_1(q) = p$. It is easily checked that $\nu_1^+(A) = p\ \nu_1^+(B) = f$. Hence the result.

III.11. THEOREM: If $A_1 \ldots A_n \models B$, then $A_1 \ldots A_{n-1} \models A_n \rightarrow B$.

Proof. Trivial.

III.12. It is easy to extend LP to a quantificational logic LPQ. Let L be a first order language. For simplicity we assume it has no constants or function symbols.

Let $\mathbf{A} = \langle DI \rangle$.

D is a domain of objects. If P_n is an n-place predicate of L and $\bar{x}$ an n-tuple of members of D, I maps $\langle P_n \bar{x} \rangle$ into $\{t, p, f\}$.

Let S be an evaluation of the variables of L (i.e., a map from the variables into D. We define the 'truth value' of a sentence under S as follows:

If A is of the form $P_n(v_i \ldots v_j)$, A is t, p, or f, according to whether $I\langle P_n \langle Sv_i \ldots Sv_j \rangle\rangle$ is t, p or f.

If A is of the form $\neg B$ or $B \wedge C$, the truth conditions of A are given by the matrices of Sections III 6–7.

If A is of the form $(\forall v)B$, then

$$A \text{ is } t \text{ iff for all } d \in D \quad B \text{ is } t \text{ under } S(v/d)$$

(where $S(v/d)$ is $S - \langle v\, S(v)\rangle \cup \langle v\, d \rangle$).

A is f iff for some $d \in D$ B is f under $S(v/d)$

A is p otherwise (i.e., for all $d \in D$, B is t or p under $S(v/d)$, and for some $d \in D$. B is p under $S(v/d)$).

This last definition can be justified by an obvious extension of the arguments used to justify the truth tables of Sections III 4–6. I will do an example. If B is t or p for all $S(v/d)$ then B is true for all $S(v/d)$. Hence $(\forall x)B$ is true. If B is p for some $S(v/d)$, then it is false for some $S(v/d)$. Thus $(\forall x)B$ is false. Hence A is paradoxical.

The 'truth conditions' of $(\exists x)B$ can be given as follows. (Alternatively, '$(\exists x)B$' can be defined as '$\neg(\forall x)\neg B$').

$(\exists x)B$ is t iff for some $d \in D$ B is t under $S(v/d)$
$(\exists x)B$ is f iff for all $d \in D$ B is f under $S(v/d)$
$(\exists x)B$ is p otherwise (i.e., for all $d \in D, B$ is f or p under $S(v/d)$ but for some $d \in D$ B is p under $S(v/d)$).

As usual one can prove that if A is a closed sentence, then the truth value of A is independent of the S one chooses.

If Σ is a set of sentences of L, we define:

$\Sigma \models A$ iff there is no $\mathbf{A}$ and no S such that A is f under S and for all $B \in \Sigma, B$ is t or p under S.

$\models A$ iff $\phi \models A$ (i.e., for all $\mathbf{A}$ and S, A is t or p under S).

III.13. THEOREM: A is a two valued logical truth iff $\models A$.

Proof. The proof from right to left is immediate since every two valued model is a three valued model.

Conversely if $\mathbf{A}, S$ are a three valued model and evaluation respectively, let $\mathbf{A}_\perp$ be the two valued model obtained by changing all the p's of I to t's. One easily checks that if A is f under S, in $\mathbf{A}, A$ is f under S in $\mathbf{A}_\perp$. The result follows.

III.14. As in III.9 the deducibility relation is changed in LPQ. For although we have

$(\forall x)A, (\forall x)B \models (\forall x)(A \wedge B)$
$(\forall x)A \models (\forall x)(A \vee B)$
$(\forall x)A \models A(x/y)$
$(\forall x)(A \to B) \models (\forall x)(\neg B \to \neg A)$
$(\forall x)(A \to B) \models (\exists x)A \to (\exists x)B$
$(\forall x)A \models \neg(\exists x)\neg A$
$(\forall x)(A \to B) \models (\forall x)A \to (\forall x)B$
$(\forall x)A \models (\forall x)(B \to A)$
$(\forall x)A \models (\forall x)(\neg A \to B)$

it is easy to find counter-examples which show that the following fail to hold.

$(\forall x)A, (\forall x)(A \to B) \models (\forall x)B$
$(\forall x)(A \to B), (\forall x)\neg B \models (\forall x)\neg A$
$(\forall x)(A \to B), (\forall x)(B \to C) \models (\forall x)(A \to C)$.

III.15. THEOREM (Substitutivity of Equivalents): If A, A^1 have variables amongst $y_1 \ldots y_n$ and B^1 is like B except that it contains A^1 where B contains A, then

$$(\forall y_1 \ldots y_n)(A \longleftrightarrow A^1) \models B \longleftrightarrow B^1.$$

Proof. The proof is by induction over the formation of B.

If A is B then the result follows since $(\forall y)C \models C$.

If a truth function is used in the construction, then the following rules are used to establish the result

$$A \longleftrightarrow B \models \neg A \longleftrightarrow \neg B$$

$$A \longleftrightarrow B \models A \wedge C \longleftrightarrow B \wedge C.$$

If a quantifier is used in the construction, then the result follows from the two rules.

If $A \models B$ and x does not occur free in A then

$$A \models (\forall x)B.$$

and

$$(\forall x)(A \longleftrightarrow B) \models (\forall x)A \longleftrightarrow (\forall x)B.$$

III.16. It is not difficult to give an axiom or rule system for LP and LPQ. However I will leave these technical problems to another paper. Also, LP bears an interesting relation to the semantics for first degree entailment presented by Dunn (1976) and the Routleys (1972). Again, I will consider this matter in another paper.

IV. CONSEQUENCES

IV.1. The most obvious thing about the logic of paradox is that it forces us to give up as invalid certain principles of deduction that one would not normally suspect. For example the rules

$$\frac{A \quad \neg A \vee B}{B} \qquad \frac{A \rightarrow \neg B \quad B}{\neg A} \qquad \frac{\forall x A \;\; \forall x(A \rightarrow B)}{\forall x B}$$

are validity preserving but not truth preserving. We can soften the blow to the intuitions a little by pointing out that although these inferences are not generally valid, they are valid provided all the truth values involved are classical (i.e., true only or false only). Let us call a rule that is truth preserving only under such conditions 'quasi-valid'. Then quasi-valid inferences are perfectly O.K. provided we steer clear of funny sentences such as 'This sentence is false' and 'The Russell set is a member of itself'. Obviously all classically valid rules of inference are either valid or quasi-valid.

IV.2. However, quasi-valid rules are generally invalid. So it might reasonably be said that they should not be used. This in turn would defeat our purpose. For the aim of the exercise was to construct a logic which could be used (in connection with a semantically closed theory) to capture naïve mathematical reasoning. However, eschewing quasi-valid rules would obviously have a crippling effect on mathematical reasoning. How can one reason without *modus ponens*? There are two ways one can cope with this problem. I will consider these in turn.

IV.3. First let us stick to the fragment of language which contains only '$\neg$', '$\wedge$' and '$\vee$'. The main loss here is the disjunctive syllogism. However, giving this up is no option. For C. I. Lewis' well known proof shows that this leads straight to $(A \wedge \neg A)/B$ as follows.

$A \wedge \neg A$	(1) Assumption
A	(2) from (1)
$A \vee B$	(3) from (2)
$\neg A$	(4) from (1)
B	(5) from (3), (4) by disjunctive syllogism.

So this has got to go. However, if this were the sum total of our losses, we might well feel that we could get by.

IV.4. The next domain of loss is connected with '$\rightarrow$'.

As we saw, the rules of *modus ponens*, *modus tollens* and *reductio ad absurdum* have to be given up. However, expressing '$\rightarrow$' in terms of '$\neg$' and '$\vee$' we see that these are really variants on the disjunctive syllogism. This suggests that it may be our identification of '$A \rightarrow B$' with '$\neg A \vee B$' that is causing the problem.

Now there are many and well known arguments against reading material implication as any form of implication. See Anderson and Belnap (1975). I will not rehearse these again. I will just say that our semantics may well be considered adequate for '$\neg$', '$\vee$', '$\wedge$' but need somehow to be extended to deal with '$\rightarrow$'. In virtue of all the work in this area, this seems quite plausible.

IV.5. However the problem of how to extend LP is not easy. In particular, it is well known that any system containing the '$\rightarrow$' of H, S4, R, E or T (see Anderson and Belnap (1975, Chapter 1)) is collapsed into triviality by the addition of the naïve comprehension scheme of set theory. (See, e.g., Routley (1977, Section 6.) A similar point can be made for semantically closed theories. To be precise, in any semantically closed theory in a language which contains a connective '$\rightarrow$' satisfying the rules

$$\frac{A \quad A \rightarrow B}{B} \qquad \frac{A \rightarrow (A \rightarrow B)}{A \rightarrow B} \qquad \text{(Absorption)}$$

anything can be proved. The proof is as follows. Consider the sentence 'If this sentence is true then A is'. Symbolically let (1) be

(1) $\quad T(1) \rightarrow A$

where 'Tx' is 'x is true'.

Then by the Tarski truth scheme, we have:

$$T(1) \longleftrightarrow (T(1) \rightarrow A).$$

Using absorption from left to right we get

$$T(1) \rightarrow A$$

and *modus ponens* from right to left

$$T(1).$$

Hence again by *modus ponens* we have

$$A.$$

Thus either *modus ponens* or absorption must be given up. Plausibly one might get by without absorption, but all the systems we mentioned contain it.

IV.6. The next area of loss is at the quantifier level. As we have defined '$\rightarrow$',

$$\frac{(\forall x)(A \rightarrow B) \quad (\forall x)A}{(\forall x)B}$$

is invalid. Thus, if we identify 'All A's are B's' with '$(\forall x)(A \rightarrow B)$', we have to give up the following.

$$\frac{\text{All } A\text{'s are } B\text{'s} \quad \text{Everything is } A}{\text{Everything is } B}.$$

If we found another way to obtain '→' then this, of course, might change. However, it is interesting to observe that however we define 'All A's are B's' (let us write this neutrally as $[A, B]_x$) one of the following must be given up

$$\text{A1} \quad \frac{[A,B]_x}{[\neg B, \neg A]_x} \qquad \text{A2} \quad \frac{(\forall x)A}{[B,A]_x} \qquad \text{A3} \quad \frac{(\forall x)A \quad [A,B]_x}{(\forall x)B}.$$

For if we hold all these we can obtain $(A \wedge \neg A)/B$ as follows. Let x be some variable not in A or B.

1)	$A \wedge \neg A$	Assumption
2)	$(\forall x)(x = x)$	Self evident truth (!).
3)	A	1)
4)	$A \wedge (\forall x)(x = x)$	2) & 3)
5)	$(\forall x)(A \wedge x = x)$	4)
6)	$[\neg B \vee x \neq x, A \wedge x = x]_x$	5) & A2
7)	$\neg A$	1)
8)	$(\forall x)\neg(A \wedge x = x)$	7)
9)	$[\neg(A \wedge x = x), \neg(\neg B \vee x \neq x)]_x$	6) & A1
10)	$(\forall x)\neg(\neg B \vee x \neq x)$	8) & 9) & A3
11)	$(\forall x)(B \wedge x = x)$	10)
12)	$B \wedge (\forall x)(x = x)$	11)
13)	B.	12)

All the inferences with the exception of A1, A2 and A3 are LPQ valid. Hence one of these has to go. Perhaps the most plausible is A2, in virtue of its connection with the paradox of implication $A \rightarrow (B \rightarrow A)$. But this is a guess. Still, if we could hold our losses to the disjunctive syllogism, absorption and A2, we might still hope (optimistically) to get by.

IV.7. The above approach to the problem (cutting our losses) is possible but obviously requires a great deal more work before we can reasonably assess its viability.

There is a much more radical approach. However, this is so radical one is tempted to dismiss it out of hand. Still let us consider it. Let us assume that *modus ponens*, *reductio ad absurdum*, etc. really are only quasi-valid. We can not give them up without crippling classical reasoning. Why should we not go on using them anyway?

IV.8. The proposal is that we allow ourselves quasi-valid inferences even though they are not generally valid. We do know that quasi-valid inferences are truth preserving provided that there are no paradoxical sentences involved (see Section IV.1). Hence, if we were certain that we were not dealing with paradoxical sentences, we could use quasi-valid rules with a clear conscience.

IV.9. Of course, paradoxical sentences do not bear their mark on their sleeve and there seems no reason to suppose that the class of paradoxical sentences is decidable (i.e., we have no effective way of telling, in general, when a sentence is paradoxical). However, paradoxical sentences seem to be a fairly small proportion of the sentences we reason with. (I would claim that they occur only in very specific circumstances: when there is some kind of semantically closed self referentiality, see Section II.8.) In view of this, it seems reasonable to formulate the following methodological maxim.

MM Unless we have specific grounds for believing that paradoxical sentences are occurring in our argument, we can allow ourselves to use both valid and quasi-valid inferences.

It would seem plausible to claim that in our day-to-day reasoning we (quite correctly) presuppose that we are not dealing with paradoxical claims. (This would explain why non-logicians are normally at such a loss when presented with logical paradoxes. For it is then clear that a presupposition of ordinary reasoning is being violated). Hence MM has the effect of legitimising the status quo. However, where the man in the street refuses to go on reasoning with paradoxical sentences, we know now that it is perfectly correct to continue provided we restrict ourselves to valid (not quasi-valid) inferences. Since these are very weak, this has precisely the effect of cordoning off the dangerous singularities which are the paradoxes.

IV.10. Perhaps the immediate reaction is that if we were to allow ourselves invalid inferences, we would no longer be sure of our conclusions. This is correct, but is not as impressive as it at first sounds. Even if one is cast-iron certain that ones rules of inference are valid, one can be no more certain of ones conclusions than one is of ones premises. In fact the new situation is little different. If our premises are not paradoxical, quasi-valid rules of inference are truth preserving. Hence we can be as sure of the truth of our conclusions as we are of the truth – only of our premises. This will normally be only a little less than (if not the same as) the degree to which we are sure of the truth of our premises.

IV.11. Although this is not a very strong argument against our present proposal, it does point to an important epistemological consequence of the proposal. In fact we find ourselves with a new argument for fallibilism in general, but in mathematics in particular. Suppose that we have an argument for a certain mathematical statement. Suppose also that the argument employs some quasi-valid inferences (as most arguments do). Then providing we have no specific grounds for believing there are any paradoxical sentences in the proof, we may invoke MM and claim to have proved the statement. But what would constitute specific grounds for believing we have a paradoxical sentence in the proof? Obviously if we can show that if a certain sentence is true, it is false and vice versa (as in the case of the liar) then we have a paradoxical sentence on our hands. If the sentence contains semantic terms and is involved with self reference, self applicability, etc., we may not have found a demonstration of paradoxicality, yet we have excellent grounds for suspicion. We should tread warily. However, it is always possible that semantic terms may be smuggled implicitly into a sentence without our knowledge or that paradoxicality arises for some other reason of which we are not aware. We may then invoke the maxim but find that at a later time evidence turns up to the effect that there are paradoxical sentences in the proof and that they occur in such a way as to invalidate a quasi-valid rule of inference. We will then have to reject the proof that we previously accepted. Although this may sound a little unrealistic, there is in fact nothing essentially new in this situation to those who are familiar with a little of the history of mathematics.

IV.12. Thus, this approach may not be as implausible as it seemed at first sight. In fact, of the alternatives considered, it seems, if anything, the more plausible. It allows us to have our cake and eat it, as it were. Whereas the first alternative suffers from the definite suspicion that there may not be enough cake.

IV.13. Finally in this section I wish to reconsider the matter of incompleteness. In Section II.6 we saw that provided we use a semantically closed theory we can avoid there being sentences independent of the theory, whose truth we can establish. However this leaves open the question of whether for every axiomatic arithmetic there are independent sentences.

We saw in Section II.6 that Gödel's theorem can not be applied to semantically closed theories to show that there are independent sentences, since Gödel's theorem applies only to consistent theories. There is however another theorem due to Tarski which, if correct, would show that there must be such sentences. Tarski's theorem states that the set of true sentence sentences of arithmetic is not arithmetic. If this is true then the set of true arithmetic statements is certainly no axiomatic. However the proof of Tarski's theorem breaks down if there are paradoxical sentences. The standard proof of Tarski's theorem is as follows:

Suppose the set of true sentences were arithmetic, i.e., there were an arithmetic sentence of one free variable Tx such that every instance of $Tk \longleftrightarrow A$ were true, where k is the code number of A. Then by the usual diagonal argument we can find a formula $\neg Tj$ whose code number is j. Substituting this for 'A' we get

(1) $$Tj \longleftrightarrow \neg Tj.$$

This cannot be true. Hence the assumption is incorrect.

This proof ignores the possibility of paradoxical sentences. If Tj is t or f then (1) is indeed false. But if Tj is paradoxical (which it obviously is, since it is the arithmetic version of 'This sentence is false') then (1) is true! Hence the proof is invalid.

Thus the standard theorems showing that no axiomatic arithmetic can be complete fail. There is no reason to believe that the set of true arithmetic sentences is not axiomatic. I conjecture that it is.

V. CONCLUDING SELF-REFERENTIAL POSTSCRIPT

V.1. It is always difficult to admit that something you have written is false. But this is the position I must now admit to being in. For what I have been saying is not without significance for what I have been saying. In particular, if what I have been saying is true, then some of the things I have been saying are false (as well). In particular, I have made certain claims about the truth conditions of sentences and this is precisely the kind of semantic self reference that leads to paradox.

V.2. To see this, consider the truth conditions of $\ulcorner A$ is true$\urcorner$. The Tarski biconditional gives us that

$$A \text{ is true iff } A.$$

Hence if A is true, $\ulcorner A$ is true$\urcorner$ is true. If A is false, $\ulcorner A$ is true$\urcorner$ is false.

It follows that A is true only (true and not false) iff $\ulcorner A$ is true$\urcorner$ is true only. A is paradoxical iff $\ulcorner A$ is true$\urcorner$ is paradoxical. A is false only iff $\ulcorner A$ is true$\urcorner$ is false only. Let us summarize the information as follows.

A	A is true
t	t
p	p
f	f

Symmetrical considerations give us the following table for $\ulcorner A$ is false$\urcorner$.

A	A is false
t	f
p	p
f	t

Now consider the metalinguistic statement

(1) Some sentences are true and false

(i.e. '$\exists s$(s is true and s is false)' where the quantifier ranges over all true or false sentences – which of course includes paradoxical ones).

Then using the above tables and the truth conditions for quantifiers given in Section III.12, (1) can be seen to be true, in fact paradoxical. Thus its negation

No sentence is true and false

is true too. Both my claim that there are paradoxical sentences and Aristotles' claim (reported in Section I.3) that there are none are true!

Perhaps even more surprisingly both of the claims

All sentences are either true or false

Some sentences are neither true nor false

are true! One would expect the former to be true but not the latter since no provision has been made for truth-valueless sentences. This should serve as a warning that we cannot read off metalinguistic facts about LP from its matrices in a cavalier way.

V.3. The point of course, is that once we have given up demanding that the object theory be consistent, there is no reason to demand that the metatheory be consistent. Indeed this is forced on us if we wish to give a coherent account of paradoxicality. Any object theory inconsistency $A \wedge \neg A$ is transformed simply by the *T*-scheme into a metatheoretical contradiction ⌜*A* is true and *A* is not true⌝.

Perhaps one of the most interesting inconsistencies in the metalanguage is provided by the following:

(2) This sentence is not true only.

If (2) is true only, then it is certainly true and hence not true only. Thus (2) is not true only. But if (2) is not true only it is false. Hence it is true only. Thus (2) is true only. We see that (2) is true only and not true only.

This reasoning is quite sound and underlines the difference between this approach and the superficially similar approach taken by van Fraassen (1968) and others of calling paradoxical sentences neither true nor false. It might be thought that our p could be equated with *neither true nor false*. But this would be a mistake. Calling paradoxical sentences neither true nor false may get one out of the liar paradox. However it does not avoid the extended liar paradox. (For such is (2).) Since the aim of this sort of approach is to *avoid* paradoxes, something new (and usually *ad hoc*) has to be done to avoid it. By contrast the aim of this paper has not been to avoid paradoxes but to show how they can be accepted without coming to grief. Once we have accepted (as I have argued we must) certain paradoxes as facts of life, then the paradoxical properties of (2) appear as just another fact.

V.6. This final section illustrates the fact that the subject of paradoxical assertions is one full of surprises. However *that* it should be so is not particularly surprising. After all, as we discussed in Section IV.9, we all normally assume that we are not reasoning about a paradoxical situation: when we meet a contradiction we take it as a sign that something has gone wrong and refuse to go further. (And let me add again, before I am accused of accepting any old contradiction, most contradictions *are* a sign that something has gone wrong: that we have an untrue premise.) It is precisely when we do go further that our familiar world disappears and we find ourselves in strange new surroundings. The new terrain clearly needs to be explored. Where it will lead is not yet clear. Yet one consequence for the history of mathematics already stands out. The discovery by Russell of a set which was both a member of itself and not a member of itself, is the greatest mathematical discovery since $\sqrt{2}$.

University of Western Australia

BIBLIOGRAPHY

Anderson, A. R. and Belnap, N. D., *Entailment*, Princeton U.P. (1975).
Asenjo, F. G., 'A Calculus of Antinomies', *Notre Dame Journal of Formal Logic* **XV** (1966), 497–509, **VII**, 103–105.
Curry, H. B., 'The Inconsistency of Certain Formal Logics', *Journal of Symbolic Logic* 7 (1942), 115–117.
da Costa, N., 'On the Theory of Inconsistent Formal Systems', *Notre Dame Journal of Formal Logic* **XV** (1974), 497–509.
Dunn, J. M., 'Intuitive Semantics for First-Degree Entailments, and "Coupled Trees" ', *Philosophical Studies* **29** (1976), 149–168.
van Fraassen, B., 'Presupposition Implication and Self Reference', *Journal of Philosophy* **65** (1968), 136–52.
Fraenkel, A., Bar Hillel, Y. and Levy, A., *Foundations of Set Theory*, North Holland, Amsterdam, 1973.
Gödel, K., 'What is Cantor's Continuum Problem', *American Mathematical Monthly* **54** (1947), 515–25. Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics*, Blackwell, 1964.
Hilbert, D. and Bernays, P., *Grundlagen der Mathematik* Vol. II, Springer Verlag, 1939.
Kleene, S. C., *Metamathematics*, North Holland, Amsterdam, 1952.
Lukasiewicz, J., 'On 3-valued Logic' (1920) in S. McCall *Polish Logic*, Oxford U.P., 1967.

Priest, G., 'Type Theory in which Variables Range Over Predicates', Ph.D. thesis, University of London, 1974.

Routley, R., 'Ultralogic as Universal', *Relevance Logic Newsletter* 2 (1977).

Routley, R. and Routley, V., 'The Semantics of First Degree Entailment', *Nous* VI (1972), 335–359.

Tarski, A., 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* 1 (1936), 266–405. Reprinted in English in *Logic Semantics and Metamathematics*, Oxford U.P., 1956.

Tarski, A., 'The Semantic Conception of Truth and the Foundations of Semantics', *Philosophy and Phenomological Research* 4 (1944), 341–376. Reprinted in L. Linsky (ed.), *Semantics and the Philosophy of Language*, University of Illinois, 1952.

Wittgenstein, L., *Philosophical Remarks*, Blackwell, 1964.