WHAT CAN WE LEARN FROM THE PARADOXES?

PART II

J. L. MACKIE
University College,
Oxford

In Part I of this article[1] I claimed that the method there used to resolve the liar, heterological, and class paradoxes would also resolve those of Richard and Berry; but this claim has still to be defended. In this Part I shall defend it, and also discuss three other developments or difficulties: a group of paradoxes studied by Arthur Prior, the varieties of self-reference, and a particularly obstinate kind of variant of the paradoxes.

There is an initially attractive but spurious solution of Richard's paradox which consists essentially in pointing out that if the proposed 'diagonal' definition of a non-terminating decimal would be, say, the rth definition in the alphabetically ordered list of definitions to which it refers, it is indeterminate at the rth decimal place: the directive for constructing the diagonal decimal there becomes: 'The rth figure is to be 1 if it is 0, and 0 if it is not 0'. So, we are inclined to say, this alleged definition just fails to *be* a definition of a non-terminating decimal, and we van happily accept the appropriate application of Thomson's barber theorem, which says that there is no such definition.[2]

But this solution is spurious, for there is still a contradiction. If the alleged definition is, as we now want to say, not a definition, it drops out of its place in the ordered list of definitions, and then it does successfully and unambiguously

[1] (Reference.)
[2] Something like this is said in T. J. Richards, 'Self-Referential Paradoxes', *Mind* LXXVI (1967), pp. 387-403, but criticized by Richards himself in a final footnote and by F. Jackson in *Mind* LXXX (1971), pp. 284-285.

define a new 'diagonal' number. So it ought to be on the list. Can we not show, then, not merely that if this is a definition, it isn't (because of the difficulty about the rth place) and therefore that it isn't, but also that if it isn't a definition, then it is, and therefore that it is?

What is at this deeper level paradoxical in Richard's paradox and evades our spurious solution is exactly mirrored in Berry's, which is much simpler, but lacks the mathematical interest given the Richard's by its resemblance to Cantor's diagonal proof. I shall consider a variant of Berry's paradox which is due to Max Black: 'The least integer not named in this book'. Supose that the largest integer 'named' (that is, mentioned or used or referred to or defined) elsewhere in the book is 256. Then this phrase appears to 'name' 257. But then 257 is 'named' in the book after all. And then the phrase doesn't 'name' 257 (but, presumably, 258). But if it doesn't 'name' 257 it doesn't 'name' 258 either; so 256 is after all the largest integer 'named' in the book, so this phrase does 'name' 257 after all. So if it 'names' 257, it doesn't so it doesn't; but if it doesn't, it does, so it does.

Both these paradoxes seem to illustrate Russell's vicious circle principle, that if we try to base the definition of something on a totality to which the something would belong, we may get into trouble.[3] But it is not enough simply to *ban* such definitions: we must find out more about what is wrong with them.

It is worth noting some non-paradoxical possibilities in this neighbourhood.

First, we could read the Berry-type reference as referring only to *other* 'namings' of integers: then it unambiguously 'names' 257. Secondly, we could read it so that it merely generates a sequence of numbers. The first time you read it, the largest integer so far 'named' is 256, so this phrase 'names' 257. But now 257 has been named, so the second time you read it is 'names' 258. And so on. Similarly Ri-

[3] Whitehead and Russell, *Principia Mathematica* I, pp. 37-38.

chard's procedure could generate a sequence of 'diagonal' decimals, each one being incorporated in the table from which the next is diagonally constructed. Thirdly, we could read the Berry-type reference as explicitly self-referring, as giving the instruction 'Consider the least integer which this remark, among others, does not invite you to consider'. If this is a paradox it is easily resolved; we can without hesitation set this aside as giving no substantial direction.

But a deeper level of paradox lies in a reading that is different from all of these. This phrase is obstinately paradoxical not when it refers only to other references, nor when it refers at each reading only to references other than itself-at-that-reading, nor when it refers explicitly to itself (as well as others), but when it refers generally to all references in the book, so that it is potentially within its own field of reference, but is actually so if and only if it 'names' an integer.

At this point we simply have to recognize that, and how, we can construct such obstinate paradoxes, using the linguistic resources that we have. To construct them, we need at least implicit quantification as well as the possibility of self-reference. Also, we need to refer to a property, such as being a successful definition, which is a derivative feature in the same way that being true, being true of itself, being self-membered, and their opposites are. We can then arrange things so that what something's being a successful definition should be derivative from includes, crucially, whether it is a successful definition or not; the feature of being, or not being, a successful definition then becomes non-derivable in this case.

The philosophical solution of this puzzle consists essentially in seeing how such obstinate paradoxes are constructed, e.g. how something's status as a definition is made (inversely) dependent on itself. Once we have understood this, we are no longer committed to endorsing contradictions ourselves. We do not have to say that the Berry-type formula both does and does not 'name' 257, or that the Richard formula both

is and is not a definition of a non-terminating decimal. What is by now the plain truth is that the status of each formula is indeterminate. The former, taken as it is intended to be taken, in indeterminate as to whether it 'names' 257 or not; and the latter is similarly indeterminate as to whether it defines or fails to define whatever decimal would result from its application to the ordered list of all other definitions of non-terminating decimals. But these indeterminacies are not due to vagueness: they result from following in a precise manner the standard meaning rules for the terms they use.

This same kind of complication, of deeper paradox, can also be found in e.g. the class paradox. For if we said simply that there is no class of all and only the non-self-membered classes, and equally none of all and only the self-membered classes, the non-self-membered classes would become a fully determinate group, and they would then have to form a class (from the point of view of general philosophy, even if not of this or that set theory). Here too we can argue that if the supposed paradoxical class doesn't exist, then it does. So our final judgement here is not after all an application of the barber theorem; it is not just that there is no such class as Russell's; but rather that whether there is or not is indeterminate. It is not only, as I said in Part I, that the derivative feature of being non-self-membered is here non-derivable, the existence or non-existence of Russell's class is a derivative feature which, being (inversely) self-dependent, is also non-derivable.

I turn now to Prior's family of paradoxes.[4] Its founder is Epimenides the Cretan. To get anything like a paradox out of his dictum that all Cretans are liars, we must take this as meaning that everything said by any Cretan is false. Since it is itself something said by a Cretan, it cannot then be true. But it seems easy to get out of any paradox here by conclud-

[4] See A. N. Prior, 'On a family of paradoxes', *Notre Dame Journal of Formal Logic*, Vol. 2 (1961), pp. 16-32, also 'Some Problems of Self-Reference in John Boridan', British Academy Lecture (1962), reprinted in *Studies in Philosophy* ed. J. N. Findlay, and *Objects of Thought*, Ch. 6.

ing that what Epimenides said was simply false. But this is where the hidden subtlety comes in. As Prior, following Church, points out, if this dictum is false, as apparently it must be, there must be some other, true, Cretan statement. But now we seem to have discovered a logically necessary connection between two distinct occurrences: A Cretan's merely *saying* one thing guarantees that some Cretan should say something else. This would violate Hume's principle, which in itself is utterly convincing, that there cannot be logically necessary connections between distinct events. There would, of course, be nothing puzzling or anti-Humean in a logical connection between a certain saying's *being true* and some other occurrence; or between a saying's *having reference* and some other occurrence; or between one event and the aptness of some description given to another. But is would be very different from all of these, and very surprising, if Epimenides's merely saying that all Cretan statements are false required the making, and the truth, of some other Cretan statement.

Prior's development of this point makes it all the more surprising. He gives a formal proof of a general thesis in predicate logic of which this is just one interpretation among many, another, for example, being 'If it is feared by a schizophrenic that nothing feared by a schizophrenic is the case, then at least two things are feared by schizophrenics'. More surprising again are the results of contraposing this thesis, for example: 'Unless something else is said by a Cretan, it is not said by a Cretan that nothing said by a Cretan is the case'. Prior also constructs other examples of the same sort. One, based on a puzzle of Buridan's, is that it simply cannot be the case that two people, *A* and *B*, each say something true, a third, *C*, says something false, and a fourth, *D* says that exactly as many truths as falsehoods are uttered on this occasion. And he concludes, though reluctantly, that 'we must just accept the fact that thinking, fearing, etc., because they are attitudes in which we put ourselves in relation to the real

world, must from time to time be oddly blocked by factors in that world, and we must just let Logic teach us where these blockages will be encountered'.[5] Prior also says that 'D's saying what is attributed to him is not more blocked, as far as this logic goes, by the saying of *A*, *B*, and *C* than their sayings are blocked by what *D* is supposed to say; and if you hear all these four people together and then ask yourself "Which of them is it who hasn't really said anything?", there is no more reason for answering "*D*" than there is for answering "*A*", "*B*" or "*C*". Prior admits that he would like some favouritism here, but doesn't see where it could come from.[6] (He rejects, for good reasons, a language hierarchy that would yield such favouritism).

These conclusions, as Prior himself makes clear, are highly improbable; and yet they are supported by what seems to be impeccable formal logic. Logic seems to bar, in certain circumstances, the occurrence of certain sayings, thinkings, and so on which are distinct from those circumstances and which in themselves, apart from those circumstances, are logically possible and can be coherently described. Here, if anywhere, we have an apparent conflict between formally valid reasoning on the one hand and common sense, backed by a Humean philosophical principle, on the other.

To resolve this puzzle, let us first ask the blunt question, 'Well, what if the event which "Logic" tells us cannot occur did occur none the less?'. What if neither Epimenides himself nor any other Cretan ever said anything else, and yet one day Epimenides remarked (though heaven knows why he should) that nothing said by a Cretan is the case. In these peculiar circumstances, his remark in fact refers only to itself. And what then would it be for it to be true? By the rules, it would be for it not to be true. And *vice versa*. In these circumstances, both the derivative properties, truth, and the absence of truth, are non-derivable. In these circumstances,

[5] *Op. cit.*, p. 32.
[6] *Op. cit.*, p. 21.

there is nothing for the dictum's being true to be and nothing for its not being true to be. Contingently, in these special circumstances, Epimenides's remark (though still fully equipped with *meaning*) acquires the lack of content, the failure to raise a substantial issue, as well as the formal contradictoriness-by-the-standard-rules, which the simple liar utterance always has. It becomes, as we may say, liar-paradoxical.

Now since this is what *would* happen if what 'Logic' tells us cannot occur did occur, it looks as if 'Logic' is assuming that such liar-paradoxicality cannot happen. And this is so. The formal proofs on which Prior relies appeal, naturally enough, to the laws of non-contradiction and excluded middle *as applied to such formulae as would, when interpreted, become e. g. 'No Cretan statements are true'*. It is being implicitly assumed in the formal procedures that each such item must itself be ither true or not true, and not both. But, as we have seen, in special circumstances such an item becomes indeterminate, the ordinary logical laws fail to apply to it because it raises no substantial issue. To rely on the ordinary logical laws, then, or on proofs which incorporate them, is in effect to deny that liar-paradoxicality can occur, and hence to infer that the combinations of circumstances which would produce it are logically impossible. But of course it can occur, both in the simple liar case and, with the help of these contingent combinations of circumstances, in these more complex ones. The conclusions which can properly be drawn, then, are only of the forms:

If it is *not-liar-paradoxically* said by a Cretan that nothing said by a Cretan is the case, then at least two things are said by Cretans.

Unless something else is said by a Cretan, it is not *not-liar-paradoxically* said by a Cretan that nothing said by a Cretan is the case.

And so on.

In particular, *A, B, C,* and *D* can all say what Buridan's puzzle reports them as saying, and the penalty is merely that

D's remark (but not *A*'s, or *B*'s, or *C*'s) *then* becomes liar-paradoxical, which yields the favouritism which Prior would have preferred, but which 'Logic' did not provide.

Note that if we change the puzzle so that *B*'s remark is false instead of true, *D*'s utterance becomes *truth-teller-paradoxical*: contingently, in these circumstances, it works like 'What I am now saying is true'. 'Logic' will not now forbid *D*'s saying what he is supposed to say, and yet it is as contingently empty in these circumstances as in those in which 'Logic' did forbid this.

The solution of these anti-Humean paradoxes, then, lies first in recognizing that liar-paradoxicality can occur, and can indirectly and contingently, in special circumstances, befall utterances which in other circumstances are non-paradoxically true or false, and secondly in understanding such liar-paradoxicality, in seeing that in these circumstances there is nothing for the being true, or not being true, of these items to be.[7]

I have said, in Part I, that there is nothing wrong in general and in principle with self-reference. But special kinds and cases of self-reference (as well as items referring to one another in ways that might be said to involve indirect self-reference) produce paradox and emptiness, of either the liar or the truth-teller variety. Let us try to sort out the various kinds of self-reference.

As many writers have pointed out, there is a harmless sort of thing which Ross[8] calls spurious self-reference. 'The sounds I am now producing are faint', 'This is an English

[7] Arthur Prior died in 1969. It is only fair to record that I put forward essentially this solution to him both privately and in an article, 'Conditionally-Restricted Operations', in *Notre Dame Journal of Formal Logic,* Vol. II (1961), pp. 236-243, but his reply (cf. p. 31 of his own article) is that this leaves *us* too, the commentators, and not merely such characters as *D*, thinking paradoxically 'in the only too straightforward sense of contradicting ourselves; and the job of being rigorously rational even about irrationality . . . is just not done'. For discussion of this point, see the last three paragraphs of this article.

[8] A. Ross 'On Self-Reference and a Puzzle in Constitutional Law', *Mind* LXXVIII (1969), pp. 1-24.

sentence', and the like are harmless because the subject can be separately identified as an utterance, sentence, or some such item, and the predicate introduces either a non-derivative property or one which, while it may be derivative from the use of a certain verbal predicate —that 'This is an English sentence' is an English sentence is derivative from the use of the predicate-phrase 'is an English sentence' together with the place of the phrase in the English language —is not derivative from *itself* or its own absence. By contrast, what Ross calls genuine, and potentially harmful, selfreference occurs when a derivative property would have to be derived, if at all, from itself or from its own negation. We must take 'property' in a wide sense here, since we want to include 'Obey this order' and 'Don't obey this order' as cases of genuine and harmful self-reference; here it is what would count as obeying the order that has to be derived in the former from itself, in the latter from its own negation.

Restricting ourselves to self-reference which is in this sense genuine, we have still to allow for the distinction between partial and total self-reference. 'All Cretan statements are true' (or 'are false'), said by a Cretan, is only partially self-referring. Partially self-referring items, no less than totally self-referring ones, lend themselves to an endless expansion, but the expansion throws off branches which are not endless. In some, though not all, cases these branches are conjunctive components of what is said. For example, if we use '*C*' for Cretan statements, '*T*' for true, '*S*' for 'All *C* are *T*' as said by a Cretan, and '*C*' for Cretan statements other than *S*, we have:

$$S = \begin{cases} \text{All } C' \text{ are } T \\ \quad \& \\ \text{It is } T \text{ that} \end{cases} \begin{cases} \text{All } C' \text{ are } T \\ \quad \& \\ \text{It is } T \text{ that} \end{cases} \begin{cases} \text{All } C' \text{ are } T \\ \quad \& \\ \text{It is } T \text{ that} \end{cases} \text{etc.}$$

This yields an endless series of terminating statements, 'All *C'* are *T*', 'It is *T* that all *C'* are *T*', 'It is *T* that it is *T* that all *C'* are *T*', and so on, and one non-terminating ex-

pression 'It is T that it is T that . . .' (for ever). Here all the item is truth-teller-paradoxical but empty.

Despite this unavoidable incompleteness of even a partially self-referring expression, we should not say that it is necessarily vicious or improper. A partially self-referring expression could be used, and could be recognized as valid and authoritative in that all its finite branches were acceptable. These branches need not be tediously repetitive as in the example given. It makes good sense, for example, to say that a person has the right to know the rights he has, including this one: this is indeed a principle well worth asserting, particularly against illiberal regimes. If we call this principle R, the person in question A, and sum up A's other rights as the right to X, we have:

$$R = \left\{\begin{array}{c} A \text{ has the right to } X \\ \& \\ A \text{ has the right to} \\ \text{know that} \end{array}\right. \left\{\begin{array}{c} A \text{ has the right to } X \\ \& \\ A \text{ has the right to} \\ \text{know that} \end{array}\right. \left\{\text{etc.}\right.$$

The terminating branches now give A the right to X, the right to know that he has the right to X, the right to know that he has the right to know that he has the right to X, and so on. After the first two or three these are not likely to be of practical importance, but there is no need to question their validity.

Partially self-referring expressions containing negations or the word 'false' applied to universal quantifications, or unnegated existential quantifications, are more awkward to handle in this way because some or all of the branches are disjoined rather than conjoined. Thus if we now use 'F' for false, 'S' for 'All C are F' as said by Epimenides, and 'C'' for Cretan statements other than this S, we have

$$S = \left\{\begin{array}{c} \text{All } C' \text{ are } F \\ \& \\ \text{It is } F \text{ that} \end{array}\right. \left\{\begin{array}{c} \text{All } C' \text{ are } F \\ \& \\ \text{It is } F \text{ that} \end{array}\right. \left\{\begin{array}{c} \text{All } C' \text{ are } F \\ \& \\ \text{It is } F \text{ that} \end{array}\right. \text{etc.}$$

which (if we use '*Q*' for 'All *C*′ are *F*') has the form *Q* & {It is *F* that *Q OR* It is *T* that {*Q* & It is *F* that . . . which reduces to *Q* & (— *Q OR* (*Q* & (— *Q OR* . . .

In this case logical calculation reduces the statement simply to the conjunction of *Q* with a non-terminating item 'It is *F* that it is *F* that . . .' (for ever). But other negative but partially self-referring expression would be less accommodating. 'A does not have the right to know what rights he has', for example, yields an awkward sequence of &s and *OR*s which cannot be reduced by cancellation.

Can we ascribe truth or falsity (and the like) to such incomplete because partially self-referring expressions?

A Cretan remark that all *C* (including this one) are *T* will be contingently and non-paradoxically false if there is some other, false, Cretan statement. But if all other Cretan statements are true, its partial self-reference becomes crucial: what ought to be the decisive element in its analysis is the non-terminating item in its expansion, 'It is *T* that it is *T* that . . .'. But this cannot be decisive; there is a little bit of emptiness here, and it matters. For this remark to be true would be for every conjunct in its expansion to be so; but there is nothing that it would be for this non-terminating conjunct to be true, so there is nothing that it would be for the remark as a whole to be true either. So this remark is either contingently false or, contingently, truth-teller-paradoxical. But we can say that when it is truth-teller-paradoxical it is true in so far as it is determinate: everything in its expansion that could be true is so. And if someone insisted that our principle *R* about rights could not be, say, morally or legally valid because it too, as we have seen, has an incomplete and therefore empty item in its expansion, we might withdraw to the claim that it is valid in so far as it is determinate.

Similarly Epimenides's remark that all *C* (including this one) are *F* is contingently and non-paradoxically false if there is some other, true, Cretan statement. But if all other Cretan statements are false, again its partial self-reference becomes

crucial. We are inclined to say that if it is false, it must be true, and *vice versa*. But we should rather say that in so far as it asserts its own falsehood, it, like its truth-teller counterpart, contains a little bit of emptiness, that the contradiction is only apparent, since it disappears in the expansion, where we have only the sequence 'It is *F* that it is *F* that . . .' which there is little temptation to call either true or false. So Epimenides's remark is either contingently false or, contingently, liar-paradoxical; but having seen how this paradoxicality arises, we need no longer be either surprised or disturbed by it. What is on the face of it rather more surprising is that when this remark is liar-paradoxical, it is also true in so far as it is determinate; but again once we have seen how this comes about, we shall cease to be surprised.

So far I have been assuming that the partial self-reference is conscious and explicit, so that our expansions are fair representations of what is said in every sense, including 'what the speaker intends to convey'. But suppose that Epimenides does not realise that his, too, is a Cretan statement —he has, perhaps, forgotten his nationality— and does not intend his dictum to be even partially self-referring: he would not say 'including this one'. The expansion does not now represent what Epimenides intends to convey. But it still gives the formal conditions which (in the contingent circumstance that the speaker is a Cretan) would have to be satisfied if his remark were to count as true. So we can still say that while there is something that it would be for Epimenides's remark to be false, there is nothing that it would be for it to be true. The complication here is that what Epimenides intends to convey may well be true. But it is still not surprising if what someone intends to convey is true, while the remark by which he tries to convey this, taken literally, is unable to be true.

These considerations make it even clearer than before that self-reference, even then it is harmful, does not result in lack of meaning, and that there can be no question of a

general logical or linguistic ban on even genuine self-reference. We cannot, from the point of view of general philosophy, get rid of the troublesome constructions. All we can do is to understand how they are, or can in special circumstances become, crucially self-dependent, how their little bits of emptiness may come to the surface, and how *then* there may be no substantial issue about, say, their truth.

It is a fascinating peculiarity of this group of paradoxes that we seem to be able to take any proposed solution and build out of it a new paradox. Suppose we have decided to characterize some utterances as indeterminate on the ground that they are liar-paradoxical or truth-teller-paradoxical in the ways I have explained. Then what are we to say about each of the following utterances.

P: 'This sentence, standardly construed, is indeterminate'.
Q: 'This sentence, standardly construed, either makes a false statement or is indeterminate'.

For *P*, the reasoning within the paradox runs that (assuming throughout that it is standardly construed) if it were true, it would be false, and if it were indeterminate it would be true, but that it can be simply and non-paradoxically false. Can we conclude happily, then, that it is false? Surely not. This sentence is totally self-referring, and being indeterminate is a derivative property which is here made self-dependent in just the sort of way that we have found in general to produce indeterminacy. If we say that this sentence is false, we can be asked in what its falsity consists, and we can only reply, in its not being indeterminate because it is false. It's false because it's false because it's false because... If self-dependence ensures emptiness, *this* is empty. We have here a conflict between the apparently neat answer given by the internal logic and the result of an informal, externally critical reflection. But the latter must prevail. We cannot let the reasoning within the paradox here force us to call *P* false any more than we can let the reasoning within one of Prior's

paradoxes tell us that certain thinkings, fearings, etc. just cannot occur in certain circumstances.

But if *P is* indeterminate, how do we evade the argument that it is therefore true, since that is just what it says? Turning to *Q*, however, we see that whatever we had said about *P*, we should somehow have had to evade such an argument about *Q*. For the reasoning within the paradox says that if *Q* is true, it is false or indeterminate, but if it is false it is true and also if it is indeterminate it is true. The internal logic here leaves us no resting place. But we surely want to say that *Q* is indeterminate, so we must block the argument from indeterminancy back to truth.

This seems at first sight to be a difficulty for other approaches as well. Consider

P′: 'This sentence, standardly construed, violates type restrictions'.

Q′: 'This sentence, standardly construed, either makes a false statements or violates type restrictions'.

If *P′* were true (on any level) it would have to violate type restrictions; and if it violates type restrictions, is it not true? Since a type theorist will deny that a sentence can *both* violate type restrictions and be true, he may be tempted, by this reasoning within the paradox, to say that *P′* can neither be true nor violate type restrictions, but must be false. But *P′ obviously* violates type restrictions. Whatever level it belongs to, since it speaks about a sentence on that level (itself) it must belong to a higher level. It is not enough for the type theorist merely to say that he does not allow such formulae to be constructed. Within his system of formal languages they do not occur, but we want to know what comment he will make on these ordinary language constructions which patently have occurred, and whether he can comment on them without contradicting himself.

He can and will claim that he does not himself use any unrestricted concept of truth as, say, things being as they are

said to be, but only an appropriate set of recursively-defined truth-predicates, one for each language or language-level, and he can then say that while P' violates type restrictions there is no truth-predicate which he is on that account committed to applying to it: its type violation will prevent it from satisfying any properly constructed truth-definition. And if we ask whether it none the less satisfies the informal concept of things being as they are said to be, he will say that this unrestricted concept is incoherent, and that there may well be cases where no-one can tell whether it applies or not.

For the moment this gets the type theorist off the hook. But can *we* not deal with the difficulty in much the same way? If P is, as our informal argument seems to show, indeterminate, then it does not succeed in saying that it is itself indeterminate, though it seems to. The argument back from the hypothesis that P is indeterminate to the conclusion that it is true is incoherent: it is convincing only because we do not hold fast to that hypothesis, but take it along with the incompatible assumption that P determinately asserts its own indeterminacy. We inconsistently combine a literal, mechanical, interpretation of its content with the critically reflective judgement that it is indeterminate.[9]

In saying this, am I after all conceding point (6) of those listed at the beginning of Part I, and in particular giving up the simple unrestricted concept of truth? I think not. I am still saying that truth consists in things being as they are stated to be; the sophistication enters in our handling of the question, 'How, in this utterance, *are* things stated to be?'. And while I temporarily let the type theorist off the hook a moment ago, I think that his claim that the unrestricted concept of truth is incoherent can only be interpreted in an equivalent way. It is the mechanical handling of this concept that is incoherent; if you want a concept that can be handled

[9] An essentially equivalent solution of this sort of puzzle was outlined in J. J. C. Smart and J. L. Mackie, 'A Variant of the Heterological Paradox', *Analysis*, Vol. 13 (1953), pp. 6-66, and Vol. 14 (1954), pp. 146-149.

mechanically you need a restricted one; but any difficulty in the notion of things being as they are stated to be must lie in some problem about how they are stated to be. And what in the Tarskian tradition is called a definition of *truth* is mainly a clarification of what is stated, and perhaps a restriction of what can be stated.

To say, as I have, that an utterance may fail to state what it appears to state, e.g. its own indeterminacy, is reminiscent of Prior's remark (to which I objected) that thinkings, fearings, etc. may be oddly blocked. But I think there is a difference. The failure of *P*, for example, to say that it is itself indeterminate in such a way that its indeterminacy would make it true is not shown simply by 'Logic' — that is, by the need to avoid the formal contradiction that would result if it were taken as saying this. It is explained by *P*'s own indeterminacy, and this in turn is explained by its total and genuine self-reference and the self-dependence that results therefrom.

P and *Q*, however, are relatively easy to deal with just because their self-reference is explicit and total. But what if we combine the present complication with the previous one? Let us consider a remark by Epimenides, still forgetful of his own nationality, 'Every sentence uttered by a Cretan, standardly construed, either makes a false statement or is indeterminate', or, what comes to the same thing, 'No sentence uttered by a Cretan, standardly construed, makes a true statement.' Let us assume, what might be the case, that no *other* sentence uttered by a Cretan, standardly construed, makes a true statement. We cannot without contradicting ourselves allow that Epimenides's remark makes a true statement. And yet if it fails *for whatever reason* to make a true statement, we must ourselves say exactly what Epimenides has said; how then can we deny that this is a sentence uttered by a Cretan which, standardly construed, makes a true statement? How can we now avoid contradicting ourselves? This example, which brings together several of my recent themes, seems to

me to be by far the thoughest version of the liar.

It is tempting to look for help from a distinction between what Epimenides intends to convey and what his sentence, as uttered by a Cretan, has as its contingent formal truth-conditions. I have tried this without success. But whether this resort helps or not with this example, it can, as so often, be sidestepped by using a variant to which the distinction does not apply. Consider a sheet of paper on which are written just two sentences, 'Two and two make five' and 'There is no sentence written on this sheet which, standardly construed, would make a true statement'. Again we cannot say that the second sentence would make a true statement, and yet if we say, for whatever reason, that it would not, then, since the only other sentence on the sheet would make a false statement, we are committed to using or endorsing a sentence which is another token of the same type and with the same reference as this one. How can we avoid contradicting ourselves here? Uncritical reliance on formal logic would result in a proof that two and two make five.

Suppose that we expand 'true' here, replacing 'would make a true statement' with 'would state that things are as they in fact are'. And remember that the 'things' in question include the success or failure of this sentence in this respect. I think we can and must say that because of the very tricky kind of self-reference and consequent self-dependence in this case, there just is no *how things are* in the key respect. Consequently, we cannot either endorse or deny a sentence-token of the same type and with the same reference as the second sentence on that sheet. We cannot now drive a wedge between what we say about that sentence and what we allow it to say about itself: standardly construed, it would state just what our own use of the same type sentence would. We must just admit that the issue that it appears to raise is indeterminate, and hope that our investigations into self-reference have explained why this is so. This sentence's indeterminacy with respect to truth is of a kind which prevents our saying even that it is

not true, and therefore from arguing, by a further step, that it is true. Awkward as this conclusion is, it has the merit of being analogous to what we found it necessary to say about the 'deeper paradoxes' that we identified in those of Richard and Berry.

As Prior says, anyone who writes about the logical paradoxes has the job of being rigorously rational even about irrationaly — that is, about irrational thinking, and about states of affairs that seem in themselves to defy reason. I am not sure that I have succeeded, but the clear inadequacies of the rival approaches show that for a philosophical solution, and not merely an exclusion device, something along the lines I have indicated is required.

RESUMEN

En la primera parte de este artículo se sostuvo que el método que se usó para resolver las paradojas del mentiroso, la de la heterología y la de conjuntos, resolvería también las de Richard y la de Berry. En esta segunda parte se defenderá esta afirmación y además se discutirán tres problemas: un grupo de paradojas que estudió A. N. Prior, las variantes de las paradojas de autorreferencia y un tipo de variante particularmente difícil de paradojas.

Aquello en lo que reside lo paradójico de la paradoja de Richard está exactamente reflejado en la paradoja de Berry. A continuación se pasa a considerar una variante de la paradoja de Berry debida a Max Black: "El último entero sin nombrar en este libro." Ambas paradojas parecen ilustrar el principio del círculo vicioso de Russell. Ahora, no basta con prohibir las definiciones que poseen esta falacia: debemos buscar aún más sobre aquello que está erróneo en ellas.

Un nivel más profundo de paradoja se encuentra al considerar que la frase anterior se refiere generalmente a todas las referencias en el libro y la solución filosófica de este misterio consiste en ver cómo se construyen estas paradojas, esto es, cómo el estatus de una definición se hace depender de sí mismo. Una vez que se entiende esto, no se está comprometido a apoyar contradicciones. Lo que es claro es que el estatus de cada formulación de paradoja es indeterminado.

A continuación se analiza la familia de paradojas tratadas por A. N. Prior. El fundador de ellas se puede decir que es Epiménides el Cretense. Prior señaló que si el enunciado de Epiménides es falso, como aparentemente lo es, debe haber otro enunciado dicho por un cretense que sea verdadero. Así, toda la apariencia es la de haber descubierto una relación lógicamente necesaria entre eventos diferentes, y esto violaría el principio de Hume que estipula justamente lo contrario.

Prior llevó a cabo una prueba formal de una tesis general en lógica de predicados, la cual es susceptible de tener interpretaciones sorpresivas del tipo: "A menos que algo más sea dicho por un cretense, no es dicho por un cretense el que nada dicho por un cretense es el caso." Estas conclusiones, como Prior mismo lo reconoce, son altamente improbables, pero son fundadas en lo que parece ser una lógica formal impecable. Aquí tenemos un conflicto aparente entre el razonamiento formalmente válido, por una parte, y el sentido co-

mún, apoyado por el principio filosófico de Hume, por otra. Para resolver este enigma debe considerarse que en circunstancias especiales de anunciado de Epiménides deja de tener contenido. Llega a ser paradójicamente-mentiroso; y la solución de estas paradojas antihumeanas reside en primer lugar, en reconocer que puede originarse una situación paradójicamente mentirosa y, en segundo lugar, en entender que en estas circunstancias no hay nada que sea verdadero o no verdadero.

Comúnmente no hay nada equivocado en principio con la autoreferencia, pero en casos especiales se producen paradojas de la variedad del mentiroso o la del sincero. Pero en general, en estas variedades de la auto-referencia no se presenta una falta de significado y no hay algún problema sobre prohibiciones lógicas o lingüísticas. No podemos, desde el punto de vista filosófico, deshacernos de estas construcciones molestas. Todo lo que podemos hacer es entender cómo son y cómo entonces puede no haber una disputa importante sobre su verdad.

Por último, se considera una peculiaridad de este grupo de paradojas que consiste en que para cualquier solución que se proponga es posible generar una nueva paradoja.