ON A FAMILY OF PARADOXES

A. N. PRIOR

1. Some paradoxical statements are, on the face of it, awkward for the propounder only, while some are also awkward for the looker-on. The Eubulidean version of the Liar paradox is of the second sort—if a man says 'What I am now saying is false', not only he himself but we who look on seem forced to say contradictory things (that his statement must be true because even if it were false it would be true, *and* that it must be false because even if it were true it would be false). On the other hand, if Epimenides the Cretan says that nothing said by a Cretan is the case, it appears that he has landed *himself* in a hole, but the beholder can contemplate his position without unease, simply saying that what Epimenides says must be false because even if it were true it would be false, and so concluding that it is false without further ado.

2. Church, however, has pointed out that there is a *little* further ado for the beholder nevertheless. For if what Epimenides says is false, then its contradictory, i.e. that *something* said by a Cretan is the case, must be true, and as the only Cretan statement we have been told about is false, this true Cretan statement which there must be, must be some other one than this. In other words, this one Cretan statement cannot even be made unless some other Cretan statement is made also.

3. Let us try formalising this proof in the propositional calculus enriched by (*a*) variables standing for monadic proposition-forming 'functors' of propositions (we shall use the one variable d for this purpose), and (*b*) quantifiers binding variables of any categories. We shall use [∀] for the universal quantifier and [∃] for the existential; for the rest, Łukasiewicz's symbols [*look at the original if you want to see what these are!*], with [↔] for material equivalence as in *Aristotle's Syllogistic*. For postulates: substitution for variables (with the usual restrictions in the presence of quantifiers), detachment, Łukasiewicz's rules for the quantifiers, definitions of the various truth functions in terms of → and ∀ (¬p = (p→∀pp)), and the one axiom ((p→q)→r)→((r→p)→(s→p)). This gives the full ordinary propositional calculus, but does not give any laws like dp→(d¬p→dq), p↔q→(dp→dq), dd∀pp→dp, which in effect restrict the values of d to

truth-functors. d can thus be used to stand for, among other things, the functor 'It is said by a Cretan that—', and where it occurs in the proofs below as a free variable it will be helpful to assign this value to it illustratively. Note that allowing this as a value of d involves the view that 'It is said by a Cretan that *p*' is not a sentence about the sentence '*p*' but a new sentence which, like 'Not *p*', is about whatever '*p*' is about; e.g. 'It is said by a Cretan that Socrates is ill' is not about the sentence 'Socrates is ill' but is another sentence which, like that one, is about Socrates.

4. We can now give the following sketch proof of Church's conclusion as informally derived in 1 and 2:—

T1. $\forall p(dp\to\neg p) \to (d\forall p(dp\to\neg p)\to\neg\forall p(dp\to\sim p))$ — from $\forall p(dp)\to dq$, by substitution.
T2. $d\forall p(dp\to\neg p) \to (\forall p(dp\to\neg p)\to\neg\forall p(dp\to\neg p))$ — from T1 and $(p\to(q\to r))\to(q\to(p\to r))$.
T3. $d\forall p(dp\to\neg p) \to \neg\forall p(dp\to\neg p)$ — from T2 and $(p\to(q\to\neg q))\to(p\to\neg q)$.
T4. $d\forall p(dp\to\neg p) \to \exists p(dp\wedge p)$ — from T3 and equivalence of 'not-none' and 'some', i.e. of 'not-all-not' and 'some'.
T5. $d\forall p(dp\to\neg p) \to (d\forall p(dp\to\neg p)\wedge\neg\forall p(dp\to\neg p))$ — from T3 and $(p\to q)\to(p\to(p\wedge q))$.
T6. $(d\forall p(dp\to\neg p)\wedge\neg\forall p(dp\to\neg p)) \to \exists p(dp\wedge\neg p)$ — substitution in $dq\to\exists pdp$.
T7. $d\forall p(dp\to\neg p) \to \exists p(dp\wedge\neg p)$ — syllogistically from T5 and T6.
T8. $d\forall p(dp\to\neg p) \to (\exists p(dp\wedge p)\wedge\exists p(dp\wedge\neg p))$ — from T4, T7 and $(p\to q)\to((p\to r)\to(p\to(q\wedge r))$.

5. What T8 asserts, with our illustrative value for d, is that if it is said by a Cretan that whatever is said by a Cretan is not the case, then something said by a Cretan is the case, and something said by a Cretan is not the case. In order to pass from here to 'There are at least two statements said by a Cretan (or Cretans)' we need to introduce a functor [p=q] for 'That p is the same thing as that q', either undefined with the two special axioms $p=p$ and $p=q\to(dp\to dq)$, or by definition as $\forall d(dp\to dq)$, which will make these axioms theorems. We can then define 'dp for at least two p's' (put $\exists_{2+}pdp$ for this) as short for $\exists p\exists q(dp\wedge dq\wedge\neg p=q)$, and proceed thus: —

T9. $p=q \to ((dp\wedge p)\to(dq\wedge q))$ — $p=q\to(dp\to dq)$, subst.
T10. $p=q\to((dp\wedge p) \to \neg(dq\wedge\neg q)$ — T9, $(p\to(q\to(r\wedge s)))\to(p\to(q\to\neg(r\wedge\neg s)))$.
T11. $(dp\wedge p\wedge dq\wedge\neg q) \to \neg p=q$ — T10, $(p\to(q\to\neg r))\to((q\wedge r)\to\neg p)$.
T12. $\exists p\exists q(dp\wedge p\wedge dq\wedge\neg q) \to \exists p\exists q(dp\wedge p\wedge dq\wedge\neg q\wedge\neg p=q)$ — T11, $(p\to q)\to(p\to(p\wedge q))$, quantification theory.
T13. $\exists p\exists q(dp\wedge p\wedge dq\wedge\neg q) \to \exists_{2+}pdp$ — T12, Df. $\exists_{2+}$
T14. $(\exists p(dp\wedge p)\wedge\exists q(dq\wedge\neg q)) \to \exists p\exists q(dp\wedge p\wedge dq\wedge\neg q)$ — subst. in $(\exists pdp\wedge\exists qgq)\to\exists p\exists q(dp\wedge gq)$.
T15. $(\exists p(dp\wedge p)\wedge\exists q(dq\wedge\neg q)) \to \exists_{2+}pdp$ — T14, T13, syll.
T16. $d\forall p(dp\to\neg p) \to \exists_{2+}pdp$ — T8, T15, syll.

And this is what we want—'If it is said by a Cretan that whatever is said by a Cretan is not the case, then at least two things are said-by-a-Cretan'.

6. If d is confined to truth-functors there is a shorter proof of T8, viz. this: There are only four monadic truth-functors, V ($Vp=(p\to p)$), S ($Sp=p$), $\neg$ and F($=\neg V$). $S\forall p(Sp\to\neg p)$ ($=\forall p(p\to\neg p)=\forall p\neg p$), $\neg\forall p(\neg p\to\neg p)$ and $F\forall p(Fp\to\neg p)$ are all clearly false, so for these substitutions the antecedent of T8 is false and the whole true. $V\forall p(Vp\to\neg p)$ is true, but so are both $\exists p(Vp\wedge p)$ (provable from $VVp\wedge Vp$), and $\exists p(Vp\wedge\neg p)$ (provable from $VFp\wedge\neg Fp$) so for this value of d both antecedent and consequent of T8 are true. For the rest, if d is confined to truth-functors, $p=q$ or $\forall d(dp\to dq)$ is just $p\leftrightarrow q$ and $\exists_{2+}pdp$ is not only logically implied by but logically equivalent to $\exists p\exists q(dp\wedge p\wedge dq\wedge\neg q)$, and if we used this form to define $\exists_{2+}pdp$ the step from T8 to T16 would be immediate.

7. In fact, however, in proving T8 and T16, I have not made use of any of those methods which would be available if d were confined to truth-functors. In this respect the proofs in 4 and 5 are like those used by Tarski to establish the equivalence of $p\wedge q$ to the *first* and more complicated of his formulae with no constants but $\leftrightarrow$ and $\forall$, and unlike the proofs he uses to establish the equivalence of $p\wedge q$ to his second and simpler formula. (See his *Logic, Semantics and Metamathematics*, Paper I, and my critical notice of this work in *Mind*, July 1957, pp. 401–3.) In 4 and 5 nothing whatever is assumed about the possible values of d except that they must be functors constructing a statement out of a statement; if there are any other such functors beside truth-functors, TT1–16 hold for them; and in

particular if 'It is said by a Cretan that—' is such a functor, TT–16 hold for that functor too. At the same time, nothing is assumed in 4 and 5 that is peculiar to *non*-extensional functors or, e.g. to ones involving the notion of assertion; TT1–16 apply to truth-functors too, also to modal functors (if these construct statements out of statements), and to ones involving not only the notion of assertion but also those of believing, hoping, fearing, etc. (under the same proviso). The following nice example of these other possible interpretations of T16 is due to P. T. Geach: If it is feared by a schizophrenic that nothing feared by a schizophrenic is the case, then there must be at least one other schizophrenic fear beside this one. And the possibility of transposing our whole discussion into such terms as these has at least this importance: There is some temptation to argue that the functor 'It is said that—' takes as its argument not a sentence but the name of one, so that 'It is said that Peter is ill' is about the sentence 'Peter is ill' rather than about Peter; but there is surely not even a superficial plausibility in saying that 'It is *feared* that Peter is ill' is about the sentence 'Peter is ill' rather than about Peter, i.e. no plausibility in saying in this case that the subordinate sentence is being *mentioned* rather than being used (in the way that subordinate sentences are).

8. Geach has also pointed out that similar consequences follow from supposing that it is said by a Cretan (feared by a schizophrenic, possible, not the case, etc.), not that *nothing*, but just that *not everything* that is said by a Cretan (feared by a schizophrenic, etc.) is the case. This more modest assertion by a Cretan that not everything said by Cretans is the case, i.e. that at least *something* said by a Cretan is not the case—this is not, like the more sweeping Cretan assertion considered earlier, something that one cannot consistently suppose true. It is, however, something that one cannot consistently suppose false; for if it were false that some Cretan assertions are false, the truth would then be that no Cretan assertions are false, and so not this one either. But what this true assertion says is that at least one Cretan assertion is false; this cannot be the Cretan assertion we know about, for that one is *not* false, so if this Cretan assertion is so much as *made* (not only 'if it is true'—if it is made, it *is* true), there must be some other Cretan assertion beside it.

9. I want now to emphasise the *limited* character of what has been demonstrated so far. It has not been shown to be categorically impossible that a Cretan should ever say that nothing (or that not everything) said by

a Cretan is true. What has been proved is not the categorical impossibility of anything in the nature of self-reference in assertions, fears, etc. All that has been proved is a *hypothetical* impossibility—what we have, if we apply the law of transposition to the theorems as stated above, is that *unless something else* is said by a Cretan (feared by a schizophrenic, possible, false, etc.) it cannot be said by a Cretan (feared by a schizophrenic, etc.) that nothing (or not everything) that is said by a Cretan (feared by a schizophrenic, etc.) is the case. If nothing else is said by any Cretan (even Epimenides) then indeed it is impossible for Epimenides the Cretan to say that nothing said by a Cretan is the case; whatever noises he makes, he will not under those circumstances be able to say *that* by them; though oddly enough the thing itself—that nothing said by a Cretan is the case—will under those circumstances be true, simply because there will under those circumstances be no Cretan assertions at all. If there *are* other assertions by Cretans, but all of them false, it will still be true, but still not sayable by a Cretan, that nothing said by a Cretan is true. That is , there will still be nothing wrong with *what we suppose the Cretan to say*, but only with *the supposition* that he says it. If on the other hand, there is at least one true assertion by a Cretan, it *will* be possible, at least as far as the above reasoning goes, for Epimenides to say that nothing said by a Cretan is the case, though of course this statement will then be a false one.

10. In fact we have proved in 4–5 nothing more than some simple corollaries of the obvious truth that *if it is a fact that no fact is asserted by a Cretan, then THIS fact (that no fact is asserted by a Cretan) is not asserted by a Cretan either*. Symbolically we might prove this thus:—

T17. $\neg\exists p(dp\wedge p)\to(d\neg\exists p(dp\wedge p)\to(d\neg\exists p(dp\wedge p)\wedge\neg\exists p(dp\wedge p)))$ — $p\to(q\to(q\wedge p))$, subst.

T18. $(d\neg\exists p(dp\wedge p)\wedge\neg\exists p(dp\wedge p))\to\exists p(dp\wedge p)$ — $dq\to\exists p\,dp$, subst.

T19. $\neg\exists p(dp\wedge p))\to(d\neg\exists p(dp\wedge p)\to\exists p(dp\wedge p))$ — T17, T18, $(p\to(q\to r))\to((r\to s)\to(p\to(q\to s)))$.

T20. $\neg\exists p(dp\wedge p))\to(\neg\exists p(dp\wedge p)\to\neg d\neg\exists p(dp\wedge p))$ — T19, $(p\to(q\to r))\to(p\to(\neg r\to\neg q))$.

T21. $\neg\exists p(dp\wedge p)\to\neg d\neg\exists p(dp\wedge p)$ — T20, $(p\to(p\to q))\to(p\to q)$.

Similarly (corresponding to 8), if no falsehood is asserted by a Cretan, then the falsehood (as it will then be) that some falsehood is asserted by a Cretan, cannot be asserted by a Cretan either.

11. These limitations to what can be proved by the methods of 4 and 5 are, I think, to be welcomed rather than regretted. That nothing said by a Cretan is the case, is something that is in fact false; it is, however, logically conceivable that it should be true; and in either case, it is something that can be said. And to say that it could not under any circumstances, even the actual ones, be said by a Cretan, would surely be to put Cretans at an excessive disadvantage beside the rest of mankind. That our theorems stop short of this extreme seems therefore a recommendation of our logic. There are other points, however, at which the limitations of our methods appear as odd and unpleasant gaps which cry out for filling up, if need be in some other way.

12. Let me turn at this point to a slightly more complicated case than any we have so far considered. L. J Cohen, in the *Journal of Symbolic Logic* for September 1957, invites us to consider a policeman who testifies that nothing the prisoner says is true, while the prisoner says that something the policeman says is true. It is clear in the first place that the policeman cannot be right, for if (as the policeman avers) nothing the prisoner says is true, then the prisoner must speak falsely in saying that something the policeman says is true, and the truth must be that *nothing* the policeman says is true, and so not that thing either. But since what the policeman says—that nothing the prisoner says is true—thus implies its own falsehood, and is false, the truth must be that *something* the prisoner says is true. Now either this true thing the prisoner says is the statement we know about—that something the policeman says is true—or it is something else. If it is something else, the prisoner says something else. If not—if the prisoner's true statement is his statement that something the policeman says is true—then the *policeman* must say something else, for the only statement of the policeman that we know about *isn't* true. So we have this now proved: If the policeman and the prisoner make the two statements mentioned by Cohen, then at least one of them must say something else besides. Once again, there is no question here of a proof that the policeman and the prisoner categorically cannot make the pair of statements mentioned; all that is proved is that *if* neither of them says anything else, then necessarily *either* the policeman does not say that nothing that the

prisoner says is true or the prisoner does not say that something that the policeman says is true. And I want to draw attention now not to the condition but to the disjunctive character of what comes after it. Our logic does not provide any means of deciding *which* of the two statements is precluded, or of proving that *both* are. And some may feel that this is an undesirable lacuna; and may feel this still more strongly after considering some allied cases.

13. In the Middle Ages the following puzzle was propounded by Jean Buridan (I modify slightly his example, which was passed on to me by P. T. Geach): Suppose there are four people who on a certain occasion say one thing each. A says that 1 and 1 are 2—a truth. B says that 2 and 2 are 4—a truth. C says that 2 and 2 are 5—a falsehood. And D says that exactly as many truths as falsehoods are uttered on this occasion. But if what D says is true, that makes 3 truths to 1 falsehood, so that it is false; while if it is false, that makes two truths and two falsehoods, and it is true. This reasoning can, I am sure, be formalised by the method of 4 and 5 into a proof of the following theorem: If not more than one thing is said by each of A, B, C and D on a certain occasion, and no one else says anything, then if A says that if p then p and B says that if p then p and C says that both p and not p, then D cannot say that exactly as many people speak truly as speak falsely on this occasion.

14. Note again that there is no question of proving that D categorically cannot say the thing ascribed to him. The impossibility only arises *if* A, B and C say certain things. If, for example, all three of them say that $p \rightarrow p$ (or $\forall p(p \rightarrow p)$), there is no reason at all why D should not say that exactly as many people speak truly as falsely on this occasion, though under these circumstances such a statement would be clearly false. However, since what we have here is a thesis of the form $p \rightarrow (q \rightarrow (r \rightarrow (s \rightarrow \neg t)))$, this says no more and no less than $p \rightarrow (t \rightarrow (r \rightarrow (s \rightarrow \neg q)))$, i.e., if not more than one thing is said by A, B, C and D on a certain occasion, and no one else says anything, then if D says that exactly as many people speak truly as falsely on this occasion, C that (for some p) both p and not p, and B that (for all p) if p then p, then A cannot say that if p then p. In other words, D's saying what is attributed to him is not more blocked, as far as this logic goes, by the sayings 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’ . For all that this fragment of logic has to tell us, it might just be a matter of who gets his say in first—if B, C and D really have said the things attributed to them, and it may be that they really can do this if they are quick enough, then A cannot say on that occasion that 2 and 2 are 4, or that if p then p. I must confess to a feeling that I would like to see a little more favouritism here; but I am not at all clear as to where it is going to come from, and I am not sure that the feeling isn’t anyhow just a prejudice born of too much reading of *Principia Mathematica*.

15. It has been suggested (by B. Sobociński) that D’s utterance must be separated from the rest as being in a ‘different language’ from them. On this view, D’s ‘truly’ and ‘falsely’ are ambiguous expressions, and if L is the language of A, B and C, and D means ‘truly-in-L’ and ‘falsely-in-L’, then his own language cannot be L but must be some other. Hence he cannot himself be counted either among those who speak truly in L or among those who speak falsely in L, and his statement, though false in his own language, is not false-in-L and is not and cannot be among the statements which it is itself about. To this I would reply that in the story as given, nothing whatever is said about the *language in which* A, B, C and D say or do not say the things attributed to them; nor is D depicted in the story as making any reference either to his own language or to that of the others. And as for ‘truly’ and ‘falsely’, ‘x says truly that p’ is to be understood throughout as simply short for ‘x says that p, and p’; ‘x says falsely that p’ as simply short for ‘x says that p, and not p’; ‘y says that x says truly that p’ for ‘y says that both x-says-that-p and p’; ‘y says that x says something true’ for ‘y says that for some p, both x-says-that-p and p’; and analogously for ‘y says that y says something true’, and so on.

16. A language or languages *could*, however, be mentioned if one wished. In other words, the story of A, B, C and D *could* be re-told, and the associated theorem proved, with the simple ‘says that’ replaced by ‘says-in-L that’, where L is some specific language. That is, we can prove that if no more than these four things are said-in-L on this occasion, then if A and B say-in-L that for all p if p then p, and C says-in-L that for some p both p and not p, then D cannot say-in-L that exactly as many people have spoken-in-L truly as have spoken-in-L falsely. And one could say, albeit a little loosely, that those who insist on a rigid hierarchy of ‘language-levels’ provide the ‘favouritism’ requested in 14, by asserting the final consequent of the preceding theorem categorically, i.e., in their metatheory of L they would say that whether A, B, and C speak as narrated or not, neither D nor anyone else can speak in L itself of what is said in L (whereas—this is where the favouritism comes in—they would not say that regardless of what D, C and B say-in-L, A cannot say-in-L that for all p, if p then p; at least, they would not say this if they supposed L rich enough to contain the propositional calculus with quantifiers). To this I would answer that (i) there are certainly languages L of which this last would be true, but (ii) it does not follow from this that there is no consistent language in which we can ever speak of what is said-in-any-language. This very thing, in fact, could not be said of the language in which it is said, if the advocates of rigid language-levels are right; and apart from that, there is no reason to suppose that the language used in this paper is inconsistent. What is true—and can be said in this language—is that there is no consistent language, and indeed no language at all, in which we can *always* speak of what is said-in-any-language.

17. Continuing the discussion of what cannot be proved by the methods of 3-5, it must be further mentioned that there is nothing in this part of logic to prevent Epimenides the Cretan from saying, regardless of what other Cretans say, that everything said by Cretans is *true*. And there is in fact no reason why he should not be supposed to say this, in the case in which there is some other Cretan statement which is false; for then this one would be obviously false also. But suppose there are no other Cretan statements but true ones. Would this one be true then? All but itself being favourably accounted for, whether *all* Cretan statements are true will depend on whether this one is. But whether this one is true depends on whether *all* are true, for that all are true is what it says. So we have an impassethere is not and cannot be any reason for judging this assertion true rather than false, or false rather than true. And this seems to me sufficient reason for denying that this *could be said* by Epimenides under such circumstances. But there is no law in the system of 4 and 5 which would be instanced by ‘If it is asserted by a Cretan that whatever is asserted by a Cretan is the case, then something asserted by a Cretan is not the case’. With the symbols available, the only such law could be $d\forall p(dp \to p) \to \exists p(dp \wedge \neg p)$ — a principle which was suggested to me on these grounds by J. L. Mackie—but this is easily falsified by letting d be ‘It is the case that—’, making the whole equivalent to $\forall p(p \to p) \to \exists p(p \wedge \neg p)$, which has

a logically true antecedent and a logically false consequent. (We could also falsify it by letting d be the modal functor L, 'Necessarily'.)

18. It seems similarly desirable to lay it down that if it is asserted by a Cretan that *something* asserted by a Cretan is the case, then something asserted by a Cretan must *be* the case. For if either nothing else were asserted-by-a-Cretan at all, or there were other things asserted by Cretans but all of them false, then the truth or falsehood of this Cretan assertion that some Cretan assertion is true would depend on whether it was *itself* true (all other cases being non-existent or unfavourable), and this in turn (since *what is said* is that some Cretan assertion is true) would depend on whether *some* Cretan assertion is true, and we are infinitely see-sawing again. Yet we cannot derive the principle mentioned at the beginning of this section as an instantiation of $d\exists p(dp \wedge p) \to \exists p(dp \wedge p)$; for with $\neg$ for d this is a plain falsehood (likewise with the modal functor $\Box\neg$, 'It is impossible that—').

19. In the last two sections, although it is not possible to replace 'It is asserted by a Cretan that—' by certain truth-functors and modal functors, it *is* possible to replace it by fIt is feared by a schizophrenic that—' and other functors involving the notion of mental attitudes. And it may be that just as there are special laws (like the law of extensionality $(p \leftrightarrow q) \to (dp \to dq)$) which fit truth-functors only, others which fit modal functors only, or only modal functors and truth-functors (all these being over and above what can be laid down or proved for all d-functors whatever), so there are special laws which only fit 'mental-attitude-functors', these special laws possibly including the pair mentioned in the last two sections.

20. It may also be observed that the counter-examples given to the formulae mentioned in 17 and 18 (call them 'Mackie's formulae') are ones in which the antecedent is a *necessary* truth (and the consequent *necessarily* false); and this may be of importance. Take the counter-example in 18, with $\forall p(p \to p)$ for its antecedent. Why can we not proceed with $\forall p(p \to p)$, 'Every proposition implies itself, as we did in 14 with the supposed Cretan assertion that every Cretan assertion is true? Why, that is, can we not say something like this: Every *other* proposition implies itself because in all the implications involved we have either antecedent and consequent both true or antecedent and consequent both false; so that leaves this proposition itself to consider; but how can we decide whether $\forall p(p \to p)$

implies itself without *first* assigning some truth-value to $\forall p(p \to p)$? Does not this land us in a circle as in the other case? No, because we know both that $\forall p(p \to p)$ implies itself and that it is true because any proposition *must* imply itself—being a proposition necessitates self-implication. This solution is suggested by an early comment of McTaggart's on Wittgenstein (*Mind*, October 1923; *Philosophical Studies*, VIII); maybe it has a superstitious ring to contemporary ears, but I must say I would rather be suckled in this particular outworn creed than go back still further to the Ramified Theory of Types, which would at this point deny that $\forall p(p \to p)$ itself was among the propositions substitutable for q in $\forall p(p \to p) \to (q \to q)$. Further, if one confined Mackie's formulae to cases in which the antecedent is contingent, this might turn out to tie up, in a contingent way, with the restriction suggested in 19; that is, it might turn out that the only contingent antecedents of the forms given ($d\forall p(dp \to p)$ and $d\exists p(dp \wedge p)$) are ones in which the d is a functor involving 'attitudes' like saying, thinking, hoping, etc.

21. What makes it a little odd that we cannot get what we want here by the methods of 4 and 5, is that we can get things so very *like* what we want by those methods. For what our T16 amounts to is that if a Cretan asserts that all Cretan assertions are false, what he says is not really falsifiable by purely 'internal' considerations alone. The self-refuting character of such an assertion can be offered as a *ratio cognoscendi* for its falsehood, but not as a *ratio essendi*. That its truth would entail its falsehood sufficiently proves that such an assertion must be false if it is made, but it cannot even be made unless there is some *other* reason for its falsehood than this one, namely a Cretan assertion distinct from itself which, being true, falsifies it by the straightforward method of being a counter-instance. Its self-refutation is only a *sign* that there must be a more straightforward refutation somewhere, if the thing is to be really asserted at all. And similarly with the self-confirmation of the weaker Cretan assertion that *some* Cretan assertion is false. It is interesting and perhaps even surprising that so much can be proved by so pure a logic as that used in 4–5; but what also seems surprising is that when this much can be thus proved, we cannot thus prove what is required for the cases considered in 17 and 18.

22. A further limitation to the logic of 4 and 5 may be noted in the following context: There can be a very great difference between the two forms $d\exists pdp$ and $\exists pddp$. For it is quite certain that if anyone *says that*

there is something he is saying (d∃pdp) he cannot but be right; while it is strongly arguable that if there is anything *that a person says that he is saying* (∃pddp), he cannot but be wrong. For the first: the theorem d∃pdp→∃pdp, 'If X says that there is something that X says, then there is something that X says' (namely *that*—that there is something that he says) is a simple substitution in dq→∃pdp. For the second, we might begin from Geach's adaptation of a paradox of Buridan's: Suppose Simple Simon says 'I say that the earth is flat'; we reply 'It isn't'; and Simple Simon retorts 'I didn't say it was—I said that I said that it was'. If the 'I say' of S.S.'s first remark is performative rather than informative, his retort is false; but, Geach has pointed out (Buridan himself oddly failed to see this), if the retort is correct (as it is if the original 'I say' is informative) then the original statement is not, for in the original statement he says that he is saying that the earth is flat when in fact (as he himself points out in his retort) he is not saying that the earth is flat, but saying that he is saying that it is. This solution presupposes rather more than the apparatus of 3, 4 and 5, namely (a) that the proposition that someone says that p is always a different proposition from the proposition p itself (Simple Simon's saying that the earth is flat is a different thing—a different thing to assert, think, fear, etc.—from the earth's being flat) and (b) that anyone can only say one thing at a time. And by the methods of 3-5 we can prove the principle

∀p¬(dp=p)→(∀p∀q((dp∧dq)→p=q)→∀p(ddp→¬dp)),

which with the above (a) and (b)—which amount to the affirmation of the two hypotheses in the case in which d is 'X says at t that—'—will yield by detachment the conclusion that whatever anyone says at t that he says at t, he does *not* say at t, i.e. whatever anyone says at t that he says at t, he says falsely that he says at t.

23. But the main thing to be noticed with these theorems is not the need for the special hypotheses (a) and (b) in the case of the second one, but something which some people feel requires to be laid down even with the first one. Mackie has raised this point particularly in connection with the case of the theorem d∃pdp→∃pdp in which we let our d be 'Descartes thinks that—'. Descartes himself can be regarded as having argued in his *Cogito*, or in the patter which accompanies his *Cogito*, that if he thinks there is something that he thinks, then there *is* something that he thinks (namely, that there is something that he thinks), so that he cannot possibly

be wrong about this. And I cannot see that this reasoning, so far as it goes, can be gainsaid. But Mackie suggests that a man cannot think *at all* (whether truly or falsely) that there is something that he is thinking, unless there is some other thing that he thinks besides. And I take it that anyone who agreed with this would also say that no one can say that he is saying that p unless he is also saying something besides this about what he is saying (a principle which with the postulate (b) of the last section would imply that no one can ever *say what he is saying* at all). I am myself inclined to think that the sort of self-confirmation and self-refutation involved in these cases is harmless. We are no doubt concerned here, as in 15 and 16, with talk about our talk (thinking about our thinking, etc.), but not, as in the earlier cases, with talk about the truth of our talk (though we draw conclusions about that). So it is not the truth of Descartes' thought that would give it its truth (as in the case in 15), or even its falsehood that would do so (as in the case in 8), but its very existence, i.e. its being thought; and 'self-confirmation' in this sense seems to me perfectly in order. But whether principles of the type suggested by Mackie are desirable or not, they are certainly not obtainable in the system 3–5.

24. As a formula embodying the principle to which Mackie is appealing here, Geach has suggested

(1) ∃pdp→∃p(dp∧¬(p=∃pdp)),

i.e., dp holds for some p, only if there is a p other than the assertion itself that dp holds for some p, for which it holds; e.g., something is being thought, only if something other than that something is being thought, is being thought; and again, something true is being said by a Cretan, only if something true and other than that something true is being said by a Cretan, is being said by a Cretan. For its 'dual', Geach gives

(2) ∀p(¬p=∀pdp→dp)→∀pdp,

i.e., if dp holds for every p other than the assertion itself that it holds for every p, then it holds for every p absolutely. For example, if every Cretan assertion is true apart from a Cretan's assertion that every Cretan assertion is true, then every Cretan assertion, *simpliciter*, is true.

25. Are these formulae of Geach's open to objections similar to those which beset Mackie's formulae of 17 and 18? In the first place, we may note that the results of substituting F, V and ¬ (or logically equivalent

functors, i.e., d's such that either $dp \leftrightarrow Fp$, $dp \leftrightarrow Vp$ or $dp \leftrightarrow \neg p$ is a law of the system) are provable even in our system of 3–5. (1) d/F and (2) d/V are settled by the mere fact that $\exists pFp$ is false and $\forall pVp$ true. Of the others, we may take as an example (1) d/¬, i.e.

$$\exists p \neg p \rightarrow \exists p(\neg p \wedge p \neq \exists p \neg p).$$

For this we simply prove in succession $\exists p \neg p$, $(p = \exists p \neg p) \rightarrow p$, $\neg p \rightarrow (p \neq \exists p \neg p)$, $\neg p \rightarrow (\neg p \wedge (p \neq \exists p \neg p))$, and then our formula. The crucial case is therefore that in which we so substitute for d as to turn dp into the simple p (or into a formula logically equivalent to the simple p, e.g., $\neg\neg p$), i.e., we must consider particularly these two formulae:—

(3) $\exists pp \rightarrow \exists p(p \wedge (p \neq \exists pp))$

(4) $(\forall p((p \neq \forall pp) \rightarrow p) \rightarrow \forall pp.$

These are in fact independent of the basis in 3–5.

26. That they do not follow from this basis is clear from the fact that they are inconsistent with laws of extensionality, e.g., $(p \leftrightarrow q) \rightarrow (p = q)$, which are known to be consistent with that basis. For if we put $\leftrightarrow$ for = in (3) and (4), as we would be entitled to do in a purely extensional system, we would obtain contradictions (since $p \leftrightarrow \exists pp$ is logically equivalent to p and $p \leftrightarrow \forall pp$ to $\neg p$). On the other hand, equally consistent with our basis in 3–5 (as may be shown by a simple four-valued matrix) are the two formulae

(5) $\exists_{2+}pp$

and

(6) $\exists_{2+}p \neg p$,

asserting that there are at least two distinct truths (i.e., p's such that p) and at least two distinct falsehoods (i.e., p's such that not-p), in the non-metalinguistic sense sketched in 5. And given these, it is not difficult to prove (3) and (4). For by (5), $\exists pp$ is not the only true proposition, which (again interpreted non-metalinguistically) is essentially what the consequent of (3) asserts. And by (6), $\forall pp$ is not the only false proposition, so that the antecedent of (4), which in effect denies this, is false.

27. There are thus no truth-functional counter-examples to Geach's formulae, as there were to Mackie's of 17 and 18. But might there not be others? Intuitively, the following case seems possible:— Let us suppose that all his life Mr. X was a great talker, and it became his ambition to talk his way right into the 21st century. Now picture him old and dying on the night of December 31st, in the year 1999. The clock is nearly pointing to midnight, and with his last breath the man says despairingly, "Everything said by me was-or-is said in the 20th century." But unknown to Mr. X, his clock was slow, and in fact the New Year had come in just before he spoke. It seems to me as obvious as anything of this sort can be, that this man's dying utterance was an honest error. But putting the functor "If it is said by Mr. X that—then it is said in the 20th century that—" for d in Geach's (2), it will assert that if everything *except* the assertion that everything said by Mr. X is said in the 20th century, is said in the 20th century if said by Mr. X (and this is *ex hypothesi* the case), then *absolutely* everything said by Mr. X is said in the 20th century. From this it would seem to follow that what Mr. X said was actually *true*, which seems monstrous. It is not, indeed, quite as bad as that—what follows is rather that what would normally be expressed by Mr. X's last utterance is true, so that either he said something true by those words, or he said nothing at all by them, or something other than what would normally be expressed by them. This is bad enough; still, if the present paper shows anything it is that our intuitions in this area are not to be trusted, so Geach's formulae may well be laws nevertheless.

28. Summing up where we have got to so far: I have admitted that there are certain limits to the possibility of self-referring assertions, beliefs, fears, etc., some of these limits being established within a very general logic and some apparently requiring postulates of a more special sort. I have felt compelled, for example, to deny that a Cretan can assert either that all Cretan assertions are false or that all Cretan assertions are true (or that some are false or that some are true) if there are no other Cretan assertions of any kind. On the other hand I have insisted that even a Cretan can make these assertions if the conditions are favourable—if there is some other Cretan assertion and it is a true one, then even a Cretan may assert truly that some Cretan assertion is true or falsely that none is; and if there is some other Cretan assertion and it is a false one, then even a Cretan may assert truly that some Cretan assertion is false or falsely that none is. And there are some logicians who would say that here I am being less restrictive than I ought to be. What more do they want? and why?

29. The 'residual unease' which there may be at what I have said so far, has been expressed by J. L. Mackie in the following way: I have asserted in 17 that if a Cretan says that all Cretan statements are true, then we can look at other Cretan statements and if we find any of them false then the Cretan statement that all Cretan statements are true can be written down as *another* false Cretan statement, and that finishes the matter (though if all other Cretan statements were *true* we would perhaps be in a fix). But this, Mackie says, is as if we had a conjunction $p \wedge q$ in which we first found q false, then on the strength of that found $p \wedge q$ false, and could only give p any truth-value at all *after* we had taken these steps and assigned 'false' to the conjunction as a whole. And it seems incredible that the truth-value of a component of a conjunction should be thus determined by the truth-value of the conjunction as a whole. This argument, Mackie insists, does not depend on identifying the *meaning* of a universal proposition with the meaning of a conjunction of singulars (or of a singular and an exceptive). All he says is this: A Cretan's assertion that all Cretan assertions are true is true if and only if it is itself true and all other Cretan assertions are true. This looks as if its truth at least partly depends on its truth, and we know that this kind of 'dependence' does not admit of such reciprocation. When I argue that nevertheless its untruth need not thus depend on itself, for we could establish that solely on the grounds of the untruth of some *other* Cretan assertion, I am still making use of the above if-and-only-if proposition, and moreover I am using it in a queer way—I get the falsehood of the original assertion from the falsehood of one member of the equivalent conjunction, and thereby and only thereby get the falsehood of the other member. For myself, I can only say that it seems that things do go this way sometimes.

30. Anyone sufficiently moved by the preceding argument may go beyond anything I would myself contend for, and hold that it is categorically impossible for a Cretan to make assertions about the truth-value of all or some Cretan assertions (in the sense which we who are not Cretans are able to give to 'all' and 'some'). But there are at least as good grounds for complaining that the system I have developed is *too* restrictive as there are for complaining that it is not restrictive enough; and it is to this new sort of complaint that I shall confine my attention from now on.

31. In 9, for example, and in 11–14, I have spoken freely about certain things being 'unsayable' in certain circumstances. Yet it seems quite obviously empirically possible that on a certain occasion (to take the example of 13) four persons A, B, C and D should respectively utter the sentence 'I and 1 are 2', '2 and 2 are 4', '2 and 2 are 5' and 'Exactly as many true things as false ones are being said on this occasion', and none of them utter anything further. (Cf. my review of Lewis Carroll, *Journal of Symbolic Logic*, September 1957, p. 310.) I am consequently committed to a distinction between the mere utterance of such sentences and actually saying what would normally be said by them. And independently of these puzzles there does seem to me to be everything to be said for making much a distinction. If Plato really says that Socrates is wise, then what we have here is not a relation between Plato and a sentence but one between Plato and Socrates; and whether Plato succeeds in thus relating himself to Socrates by relating himself in another way to a certain sentence, may well depend on all sorts of circumstances that we may take a while to notice. What turns out to be less straightforward than one might expect is the relation between the sort of thing done with functors like 'X says that—' in 3-5, and ordinary Semantics; I mean the kind of thing you get in Tarski's paper on Truth. In the system of 3–5, 'X says truly that p' can be defined very simply as 'X says that p, and p', so that we have it as a law that if X says that p, then he says so truly if and only if p. This, as far as it goes, is very like Tarski's 'Convention T' (*Logic, Semantics, Metamathematics*, pp. 187–8), though much simpler. But whereas the above rule is instantiated by

(a) *Whoever says that snow is white says so truly if and only if snow is white,*

the corresponding instantiation of Tarski's convention would be

(b) *The English sentence 'Snow is white' is true if and only if snow is white.*

Tarski's convention is not concerned with *saying truly that* something-or-other, but with the kind of truth that can be predicated of a form of words.

29 We might try to relate the two conceptions by equating (b) with

(c) *Whoever, speaking English, utters the sentence 'Snow is white', says something true thereby if and only if snow is white.*

But this, I would contend, is false, and could only be derived from the true principle (a) by means of

(d) *Whoever, speaking English, utters the sentence 'Snow is white', thereby says that snow is white,*

which I should say is also false. For example, since snow *is* white, Tarski cannot say that snow is white by uttering either that sentence or any other, if he says nothing else immediately after I have said that either what I am then saying or what he will say immediately after, but not both of these things, is false (cf. 14). At least, (d) could fail for the assertion ascribed to myself. Though one does not need to drop (c) outright—one only needs to tack on to it 'provided that he does say *something* by it'. (d) also holds with the same proviso.

32. It is of some interest that, as Geach has pointed out, Jean Buridan was led by some of his paradoxes to a 'non-Tarskian' view of the language he was considering; in fact Buridan went much further in this direction than I would, abandoning even (a) above (or the principle that (a) illustrates). He argues, e.g. that there are circumstances in which the sentence 'What Plato says is false', uttered by Socrates, would be false even though what Plato says *is* false, and even though the same sentence, uttered by 'Robertus' on the same occasion, would not be false but true. He thinks this is what would happen if 'What Plato says is false' were the sole utterance of Socrates and 'What Socrates says is true' the sole utterance of Plato. I need not reproduce his argument; the case is clearly the same as Cohen's court-case, my own view of it being that under these circumstances at least one of the two philosophers would not succeed in saying anything at all, true or false, by his sentence. I would agree with Buridan that Robert could say something by uttering the same sentence on the same occasion, and this although Robert and Socrates utter it for the same reason, both falsely believing that what Plato is saying is 'God doesn't exist'. (This last subtlety is in the original.)

33 But do I in fact gain anything by the small pinch of non-Tarskianism that I have allowed myself? We must not forget how widely the variable d of our TT1-21 may range, and M. Dummett has pointed out that one of its possible values is Ėpimenides utters words which conventionally signify in his language that—'. Then we get, analogously to Geach's modification of the Epimenides in 8, the conclusion that unless someone (himself or another) has uttered *other* words which conventionally signify in his language something that isn't so, Epimenides cannot even *utter the words* which conventionally signify in his language that

someone has uttered words which conventionally signify in his language something that isn't so. The answer to this, I suspect, is that *signifying that* something or other is not something that can be infallibly effected by our 'conventions'.

34. Even apart from this point of Dummett's, however, the distinction drawn in 25 between uttering such-and-such and *saying that* so-and-so, is only relevant in that very limited number of cases in which we let our d be 'It is said that—', 'It is denied that—', or some function of these. For example, what are we really supposing when we think we are supposing that some schizophrenic fears at t that nothing feared by any schizophrenic at t is the case, and that nothing else is feared by a schizophrenic at t? We are certainly not supposing that he utters words. Or again, take the following case: Mr. X, who thinks Mr. Y a complete idiot, walks along a corridor with Mr. Y just before 6 p.m. on a certain evening, and they separate into two adjacent rooms. Mr. X thinks that Mr. Y has gone into Room 7 and himself into Room 8, but owing to some piece of absent-mindedness Mr. Y has in fact entered Room 6 and Mr. X Room 7. Alone in Room 7 just before 6, Mr. X thinks of Mr. Y in Room 7 and of Mr. Y's idiocy, and at precisely 6 o'clock reflects that nothing that is thought by anyone in Room 7 at 6 o'clock is actually the case. Now in 4–5 it has been rigorously proved, using only the most general and certain principles of logic, that under the circumstances supposed Mr. X just *cannot* be thinking anything of the sort. What, then, *are* we in our muddle supposing him to be doing? Certainly something which to himself looking back on it a moment later would be quite indistinguishable from thinking that nobody in Room 7, etc. (and he might go home without ever learning of his error). How, we all want to cry out, can what a man is thinking and even what a man can be thinking on a given occasion, depend on what number is written on the other side of a door? That what a man can *truly* think should depend on things like this, is reasonable enough; but that what he can think *at all* should depend on such factors—can we swallow that? These cases are surely in a way *worse* than the simple Liar; for one might well agree that no man ever does just sit down and say (or think or fear) that whatever he says (or thinks or fears) is false; even the most stupid person must see that this is self-defeating and not do it without inserting or intending the obvious provisoes. But in the cases we are now considering, the things that we are supposing to be thought (feared, etc.) are things that

quite easily could be thought (feared, etc.) by an intelligent and logically-instructed person, and that could even be thought (feared, etc.) by the very person we are puzzling about, if it were not for some quite contingent circumstance of which that person might well be for ever unaware.

35. It will not *quite* do to say that what is vexing us here is the idea that Mr. X could *think that he thinks* something when in fact he is not and cannot be thinking this thing. Indeed, if we suppose him to think that he thinks this thing at the same time as he thinks it, the situation (for the supposer) is vastly eased. For we can now suppose him to think falsely that no thoughts in Room 7 are true ones, this being false for the straightforward reason that *the thought that he is thinking that no thoughts in Room 7 are true ones* (which we now suppose him to be thinking as well as the other) is a true thought in Room 7. However, this is a somewhat special case, and if in our puzzle we replace thinking by fearing throughout, then part of the puzzle would be that Mr. X (even supposing him the best introspector in the world) could think that he is afraid that nothing feared in Room 7 is the case, when in fact he is not and logically cannot be afraid of anything of the sort, and this not because of a logical but because of an empirical fact of which he happens to be ignorant. But with thinking, as we have stated the case in the last section, the difficulty is simply that to us something should appear to be quite obviously empirically possible which is in fact not even logically so.

36. It is rather tempting to say about the man in Room 7 that what we have misdescribed as a thought of his about Room 7 is in fact a thought of his about Room 6. But we would not consider ourselves justified in saying this in other cases which closely resemble the present one but do not happen to issue in paradoxes. Let us suppose, for example, that Mr. X has gone not into Room 7 but into Room 9, and knows perfectly well that that is where he is, but still thinks mistakenly that Mr. Y is in Room 7 when in fact he is in Room 6. We may again suppose him to think at 6 o'clock that nothing thought at 6 o'clock in Room 7 is the case; but now there is nothing at all contradictory in this supposing; we may even suppose him to think *rightly* that nothing thought in Room 7 is the case, this being true because although Mr. Y is not in Room 7 someone equally idiotic is (or perhaps because no one is). For such a case we would surely say that Mr. X was right about Room 7, though for wrong reasons. And if in fact the occ∀pant of Room 7 was a perfectly sensible person whose thoughts at 6

o'clock were true ones, we would say that Mr. X had thought something about Room 7 that was wrong, rather than that he had not been thinking about Room 7 at all but about Room 6 (and so was actually *right* in what he thought). So I don't think this way out will do.

37. J. L. Mackie suggests that while we cannot deny the empirical possibility of Mr. X's thinking that nothing thought at 6 in Room 7 is the case, even under the circumstances envisaged in 34, the reasoning in that section shows that he cannot think this *non-paradoxically*, paradoxicality and its absence being features of thinking which are not always introspectible. But either Mackie's phrase 'paradoxical thinking' refers to some species of thinking or it does not (it would not if 'paradoxical' were an *alienans* adjective like '*soi-disant*'). If it does not, i.e. if paradoxical thinking is no more a kind of thinking than imaginary money is a kind of money, then the conclusion of the argument of 34 is admitted. If, on the other hand, paradoxical thinking *is* thinking, then that argument shows that under the circumstances described it cannot occur, i.e. Mr. X cannot think either paradoxically or non-paradoxically, in Room 7 at 6, that nothing thought in Room 7 at 6 is the case, if this is all that is thought (paradoxically or non-paradoxically) in Room 7 at 6; for if he did, it both would and would not be the case that nothing thought in Room 7 at 6 was the case. The trouble here is that if we suppose Mr. X to have this thought it is not merely Mr. X but we who 'think paradoxically', in the only too straightforward sense of contradicting ourselves; and the job of being rigorously rational even about irrationality (which is surely what all this consideration of paradoxes is in aid of) is just not done.

38. Further, even if we take the line that Mackie's 'paradoxical thinking' is not thinking, while this provides at least a verbal solution to the case of Mr. X (more than verbal if we can see what, positively, this paradoxical-thinking *is*), it gives rise to analogous problems of its own; at least it does so if it makes sense to say that someone paradoxically thinks that p. For this then becomes a possible value of dp in 3–5, and we can show that no one can paradoxically-think that nothing that he paradoxically-thinks is the case, unless there is something else, and something that is the case, that he paradoxically-thinks as well.

39. At this point I must confess that all I can say to allay the misgivings expressed in the past four sections is that so far as I have been able to find out, my terms are the best at present offering. I have been driven to my

conclusion very unwillingly, and have as it were wrested from Logic the very most that I can for myself and others who feel as I do. So far as I can see, 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.

40. Look back again at the grand simplicities of 10, and apply them here. If it is a fact that *no* fact is being assented to in Room 7 at 6, then *this* fact (that no fact is being assented to, etc.) cannot be being assented to in Room 7 at 6. There just isn't any way round this, is there? Not, anyhow, unless one says with the Ramifiers that there is no such thing as a plain fact, but only first-order facts, second-order facts, and so on; that the fact that no first-order fact is being assented to in Room 7 is itself not a first-order but a second-order fact; and that the fact that no fact of any order is being assented to in Room 7 is not and cannot be assented to by anyone at all, even in Room 9, because there is not and cannot be any such fact. This would be to dispose of an argument for certain restrictions on what is allowed to be sayable, thinkable, etc., by admitting both these and countless other restrictions by another door; not, it seems to me, the shrewdest of bargains. One can admit, however, that it is when he is 'order-jumping' (or at least when someone in his neighbourhood is doing so) that the world's best introspector is liable to find himself deceived.

Manchester University
Manchester, England