Can Contradictions Be True?
Author(s): Timothy Smiley and Graham Priest
Source: *Proceedings of the Aristotelian Society, Supplementary Volumes*, Vol. 67 (1993), pp. 17-33+35-54
Published by: Wiley on behalf of The Aristotelian Society
Stable URL: http://www.jstor.org/stable/4106983
Accessed: 11-03-2015 00:26 UTC

CAN CONTRADICTIONS BE TRUE?

Timothy Smiley and Graham Priest

I—Timothy Smiley

I

When logicians claim that contradictions can be true, one wonders what leads them to say such a thing, how their logic copes with the consequences, and what is their account of negation. There are now to my knowledge three theories of this sort: Routley & Meyer's 'dialectical logic', Rescher & Brandom's 'logic of inconsistency', and Priest's 'dialethism'. I take up the questions of motivation and consequences in sections II and III, concentrating on dialethism and incidentally solving the Liar paradox. I begin with the question about negation, for the fact that a logical system tolerates A and $\sim A$ is only significant if there is reason to think that the tilde means 'not'. Don't we say 'In Australia, the winter is in the summer', 'In Australia, people who stand upright have their heads pointing downwards', 'In Australia, mammals lay eggs', 'In Australia, swans are black'? If 'In Australia' can thus behave like 'not' [1], perhaps the tilde means 'In Australia'?

Dialectical Logic [2] was Routley & Meyer's war work; supplying arms to the Soviet side in 'the ideological logical warfare between East and West'. The Dictionary of Marxist Thought acknowledges their help:

> Even more interesting is the fact that a complete formal system with entailment can be constructed within which certain propositions and their negations are both theorems and yet the classical law of non-contradiction—not (A and not A)—is a theorem (Routley and Meyer 1976). The significance of this is that there is no conclusive reason in formal logic for rejecting the view that the world supports some formal contradictions. [3]

Do they deserve this tribute? Their formal semantics goes: $\sim A$ is true in world a iff A is not true in a^*, where a^* is the image world of a. By itself this 'star rule' is merely a device for preserving a recursive treatment of the connectives (it was used in [4] to define a^* in terms of negation rather than the other way round), and it does nothing to explain their tilde until supplemented by an explanation of a^*. So what has their tilde got to do with negation? Very well, they reply, 'call it a *negatiom* then, in which case dialectical logic has no negation, i.e. classical negation, but does have as a surrogate a negatiom'. Bourgeois formalists!

For a genuine attempt to justify the star rule one has to turn to Routley's [5], which modifies a traditional idea of negation as otherness so as to create a family of models: extensional models in the spirit of Venn diagrams, a 'debate' model in which a and a^* are opposing sides in a debate, and a 'record cabinet' model in which they are opposite sides of a record. These all yield the same rule for the tilde; the snag is that it is the wrong one. We wanted the star rule '$\sim A$ is true in a iff A *is not* true in a^*', but what we get is '$\sim A$ is true in a iff A *is* true in a^*'. Thus the extensional models all include the equation $j(\sim A) = j^*(A)$ (p.215), while in the record model $\sim A$ is on side a iff A is on side a^* (p.219). As for the debate model, it

> can be given a more semantical turn. In the p-issue, $\sim p$ is asserted, or presented as true, on one side, a say (i.e. $a \vDash \sim p$ in obvious notation), while the reverse, namely p, is asserted or presented as true, on the opposite side a^* (i.e. symbolically $a^* \vDash p$). Now one side succeeds in a debate, or establishes its case, iff the opposite side does not; therefore $a \vDash \sim p$ iff $a^* \nvDash p$. That is, a version of the star rule naturally emerges from the debate model more semantically considered. Statement $\sim p$ is made, or presented as, true at side or situation a iff p is made, or presented as, true at its opposite a^*. (p.218)

The unwanted rule (with *is*) is affirmed, correctly, at the beginning of this passage and again at the end. The star rule (with *is not*) is produced in between like a conjuring trick, by switching the meaning of '$a \vDash$' from 'a presents as true' to 'a establishes its case'; and of course these are quite different things, the possible-worlds equivalent of 'a establishes its case' being 'a matches the actual world'.

The Logic of Inconsistency [6] was inspired by the observation that people can hold contradictory, even overtly contradictory beliefs, so long as the dispositions to assert them only manifest themselves in distinct contexts or 'assertion zones'. The logic is therefore designed to block inferences, such as the inference from

A and ~*A* to *A*&~*A*, which illicitly combine theses across the boundaries of different assertion zones. Onto this unexceptionable idea, however, Rescher & Brandom graft one which goes clean counter to it. Namely, when my separate beliefs, say *A* and ~*A*, are considered all together as a set, the fact that they inhabit incompatible assertion zones can be ignored, and they can be treated as if I believed them all simultaneously. Putting it as the authors do in terms of possible worlds, given a world in which *A* obtains and a world in which ~*A* obtains, there is supposed to be another world in which *A* and ~*A* obtain simultaneously. How is this managed? By changing the truth-conditions for ~ and replacing the classical rule by a new rule with six subclauses, fabricated to produce the desired result (pp.6, 148). And similarly, to block the inference to *A*&~*A*, the classical truth-table for & is replaced by another artificial rule with six more subclauses (pp.148–49). In short, what they have done is to solve an abstract mathematical problem: if arbitrary interpretations of the symbols are permitted, can one interpret ~ and & so as to make *A* and ~*A* and all tautologies come out true and *A*&~*A* and all other contra-tautologies false? They leave no doubt how arbitrary the process of re-interpretation is: 'We ourselves are in control of their defining characterization, and we can manipulate this descriptive make-up as we wish' (p.14). But this is not what their prospectus promised (pp. *x*, 2, 4). It said nothing about arbitrary re-interpretation. On the contrary, ~ and & were to be 'the familiar connectives of classical propositional calculus', giving point to the caution 'we ourselves do not declare both *P* and ~*P*, but only $T_w(P)$ and $T_w(\sim P)$'. It promised '*inconsistent* worlds ... as genuinely possible cases' (their italics), worlds such that 'in some perfectly definite way something *both is and is not so*' (my italics). What else can one do but ask for one's money back?

Dialethism [7] stands to the classical idea of negation like special relativity to Newtonian mechanics: they agree in the familiar areas but diverge at the margins (notably the paradoxes). This agreement removes what might otherwise be a difficulty in assessing Priest's own use of negative language, 'exclusive', 'ineffable' and so on. His actual account of negation is a matter of making as little change as necessary and accepting classical principles whenever it seems safe. He does however make two important theoretical points: the need to acknowledge rejection as the polar opposite of acceptance or assertion, and a link between truth and assertion through a 'teleological' account of truth. Both, I shall suggest, should lead him to rejoin the classical club.

The classical idea links negation to acceptance and rejection through the equivalence between accepting ~*A* and rejecting *A*. Indeed it takes the equivalence so much for granted that its adherents are liable to overlook or even deny the separate existence of rejection. For Priest, however, while the joint acceptance and rejection of *A* is impossible, the joint acceptance of *A* and ~*A* is possible or even mandatory. He therefore needs to deny that accepting ~*A* implies rejecting *A*, and he eases the way for this by denying the rest of the equivalence as well, dissociating negation from rejection altogether. For example, he says (pp.123, 128), a scientist may reject a statistical hypothesis without thereby accepting its negation, arguing directly against *A* without making a specific case for ~*A* and without even using 'not'. Again, an intuitionist who rejects the law of excluded middle does not accept its negation.

In reply, one needs to clarify the classical claim that rejecting *A* and accepting ~*A* are equivalent. Priest is quite right to say that a particular act of rejecting *A* may not itself be an act of accepting ~*A*, as in his example of the scientist. But equivalence is not identity. Rather, the classical equivalence involves the relation which he calls 'is rationally committed to' (p.141). So explained, the equivalence is not impugned by his example, in which the operative words are 'thereby' and 'specific'.

Next, one needs to distinguish different attitudes, and attitudes to different objects, which are lumped together under Priest's blanket definition of rejection as 'refusal to accept'. There is a difference between (1) rejecting a proposition as unwarranted, and (2) rejecting a proposition as untrue (outright denial). And there is a difference between rejecting a proposition and (3) rejecting an utterance. We do not use negation in the first case, but there is one to suit each of the others. Thus see Horn's [8] for a loving account by a linguist of two kinds of negation, differing 'not only phonologically, morphologically and syntactically, but also in semantic function'. Polemical negation[1] signifies objection to an (actual or

1 'Négation polemique', the term attributed to Ducrot in [8]. Horn prefers Ducrot's earlier

possible) utterance as inappropriate, whether because misleading, an understatement, untrue, unwarranted, meaningless, misspelt, not p*l*t*c*lly c*rr*ct, or for any other reason. It typically uses focusing devices such as intonation and emphasis (compare 'I'm not Scotch' with 'I'm not Scottish'), or cleft syntax ('It's not a doctor he needs, it's a lawyer', 'It's not that he's innocent, it's just that he's not been proved guilty'). It is often accompanied by a rectification, as in these last examples. And it is compatible with the truth of the affirmative ('I'm not his daughter—he's my father', 'His re-election is not possible, it's certain'), unless the feature objected to happens to be a necessary condition for the affirmative to be true. Propositional negation[2] is too familiar to need comment, except to note that the two negations are distinguishable even when the ground for the objection to an utterance is that it is untrue, e.g. 'It's not that 57 is prime, it's that 57 is not prime'.

Priest's counterexample of the intuitionist can now be set aside as falling under the first case, rejection of a proposition as unwarranted. For present purposes I set aside the third case too, since polemical negation functions solely as a sign of rejection and not as a propositional operator. It will earn its keep in section II. The case that concerns us here is the second, since it seems that propositional negation can function both as a sign of rejection and as a propositional operator, when an utterance of negated *A* doubles as a denial of the proposition expressed by *A* and as an assertion of ~*A*. The classical account accordingly gives negation a double foundation in the ideas of disagreement and incompatibility. I shall not try to rehearse it; if I did I could hardly improve on Huw Price's article 'Why "Not"?' [9], despite its being set in the context of a debate about excluded middle rather than contradiction. My purpose is to adduce some dialethic support for the classical equivalence, in the shape of principle R (p.141), 'If a disjunction is rationally acceptable and one of the disjuncts is rationally rejectable, then the other is rationally acceptable'. For when applied to Priest's ideal rational agent (p.128), the classical equivalence

'négation métalinguistique' because 'polemical' is too loaded, but I fear that 'metalinguistic' is even more loaded.

2 Using the term to indicate the effect achieved rather than the means used (which could well be predicate negation). Cf. [8] on 'descriptive' negation.

becomes '*A* is rationally rejectable iff ~*A* is rationally acceptable'. And by applying principle R to the dialethically accepted law of excluded middle '*A* or ~*A*' we obtain one half of this, namely 'If *A* is rationally rejectable, ~*A* is rationally assertible'. With a firm connection between negation and rejection thus re-established on Priest's own premises, the onus seems to be on him to meet the classical case as put by Price, say; or else to accept it and look for another solution to the paradoxes.

Priest tentatively distinguishes untruth from falsity, i.e. the untruth of *A* from the truth of ~*A*. Either way, untruth in dialethism shares with falsity the crucial feature that it does not exclude truth. His teleological account of truth (pp.74ff.) is intended to endorse Dummett's explanation of the point of introducing the predicate 'true'. It says in a nutshell that the true sentences are those we aim to assert. But the predicate 'true' does more than determine positively which sentences are true; it draws a distinction between truths and untruths, and an explanation of the predicate should cover the distinction, not just the positive side of it. What, one wants to know, was the point of distinguishing between truths and untruths? Again, if 'asserting, like other human activities, has a *telos*' (p.77), so does rejection, understood as before to mean denial of the proposition in question. What then is the *telos* of rejection? The teleological account of truth suggests the same obvious answer to both questions: if the true sentences are those we aim to assert, the untrue ones are those we aim to reject. Since Priest accepts that assertion and rejection are exclusive, should he not conclude that truth and untruth are exclusive too?

II

Instead of finding fault with the derivation of the Liar paradox, says Priest, one should accept it as proving the existence of a true contradiction or dialetheia. This has the great attraction that instead of closing down possibilities it opens them up, like the liberating discovery of irrational numbers. Dialetheias resemble irrationals too in that after the initial discovery they turn up everywhere: in set theory, quantum mechanics, the law, and wherever there is motion or change [7]. I shall stick to the Liar paradox, however, as a test case. Any solution to a paradox implies a criticism of rival ones, just as the diagnosis of an illness implies, if only tacitly, a rejection of alternative diagnoses. This aspect is unusually prominent in Priest's unorthodox solution, for he is driven to it by a general argument that *any* orthodox rival is bound to be self-defeating. I shall try to subvert the case for dialethism by propounding one he has overlooked and showing how it avoids his objection.

Suppose that some grammatically acceptable sentences *malfunction*—that in a particular context, perhaps in any context, they fail to convey any coherent message. Suppose too that the Liar is among them. And suppose finally that the cases of malfunction, and hence also of success, are not regularly distributed.

The last supposition suggests a parallel with Gödel's theorem. Gödel showed that the truths of arithmetic are not regularly distributed (specifically, not recursively enumerable); but the theorems of any formal axiomatic theory are recursively enumerable. The theorems must therefore overshoot or undershoot the truths. Overshooting is not a serious option, since it means that the theory is unsound, so any acceptable theory must undershoot, i.e. be incomplete. A parallel situation will face any logician trying to reform our language on the model of a calculus. If the wffs are a recursive set while the successful sentences are not, the former must overshoot or undershoot the latter. Overshooting is not a serious option. For the rules of any formal logic trade on a correspondence between syntactical features and semantical ones, and if the correspondence breaks down for a sentence overall, as when it malfunctions, it cannot be assumed to apply to its parts either. So any acceptable calculus must undershoot: its expressive power must be narrower than the original language, and some successful sentences must be excluded from the formalisation.

In a natural language, however, overshooting is a possibility to be taken seriously. Tackling the paradoxes in a language which we have to take as we find it is a different project from constructing a calculus. If this still needs saying, it is because so many logicians only pay lip-service to it. The ease with which a practitioner [10] can put natural languages into the hat and pull predicate calculus out is breathtaking—an aside about idealisations in science and how their falsity does not interfere with their applicability, and presto! Even Priest, who for his own purposes postulates that the application of sentences to the world may go quirkily wrong (p.85), cannot quite shake off the idea that a language is a calculus with a human face. Discussing a variant of the medieval paradox in which Socrates says 'What Plato is saying is false' while Plato is saying 'What Socrates is saying is true' (pp.17–18), he dismisses out of hand the possibility of the sentence malfunctioning: 'I understood what he said; I can draw inferences from it; I can act on the information contained in it'. He does not seem to entertain the possibility that these things are exploratory and provisional. Drawing the normal inferences from *A* presupposes that *A* functions normally; but that is no bar to their contributing to a reductio ad absurdum argument that it does not. The mistake is to think of malfunctioning as being like failure to be a wff, something perceptible and inherent and all-or-nothing, whereas it may be inferred and fortuitous and perhaps a matter of degree.

It is common ground that a solution to a paradox must provide some independent reason for the position it adopts, and this is one of Priest's criticisms of orthodox solutions. But it is easy to overstate the requirement. Suppose that the paradox would be blocked if the sentence in question had a certain feature (dialetheia, ambiguity or whatever). Ideally we would both demonstrate the fact and explain it. Short of this we might manage demonstration without explanation, as typically happens with reductio ad absurdum proofs. Falling shorter still, we might have non-conclusive arguments for our position. For example, an analogy between the Liar and the Truth Teller or Curry's paradox would be better than nothing; and from the dialethic point of view reductio ad absurdum should also fall into this category—not conclusive because the pivotal contradiction might turn out to be a dialetheia, but not therefore negligible. Where I part company with Priest is that I think (a) satisfactoriness is a matter of degree and (b) we are all in the same boat. Over (a) I find his criticisms of orthodox solutions too black-and-white. For example, I cannot accept the charge that using reductio ad absurdum to support a solution just begs the question and smacks of fraud (p.17). It could only do that in the context of an argument specifically against dialethism. Over (b) I think that all solutions are liable to fall short of the ideal. All follow Sherlock Holmes's maxim 'when you have eliminated the impossible, whatever remains, however improbable, must be the truth'. But all face the difficulty that eliminating the impossible involves contested judgments about the comparative implausibility of rival solutions. And none match the great detective's ability to explain the affair, i.e. to show that what remains not 'must be' but *is* the truth. The solution I am advocating is no exception, even in the hands of its most persuasive pioneers—Mackie & Smart [11, 12], Kneale [13] and Mackie [14]. But dialethism is no better off either.

Priest's further and chief criticism of orthodox solutions is that they are necessarily self-defeating (pp.18–20 & 27–31). For let them be faced with a strengthened Liar, *A*: '*A* is not true'. This can be spelt out as '*A* is false or ungrounded / truth-valueless / not stably true', utilising the very idea that was used to solve the original. Then they have left themselves no way out except to say that the idea in question, though expressible in the language *in* which the solution is given, is not expressible in the language *for* which it is given. Since the former is English, this is an admission that the problem has not after all been solved for English.

How can we deal with the strengthened Liar? There are two possible answers. The first would deny something Priest takes to be beyond question (p.19), that if *A* is not true then (any occurrence of) '*A* is not true' is true. When we commentators voice the conclusion that *A* is not true, we use a different token of '*A* is not true' from *A*; and what we say can be true though *A* itself goes awry, the reason being that one token is self-referring while the other is not. This is the answer given by Whiteley [15] and Goldstein [16], and the same point has been made by those with other fish to fry, e.g. Hughes [17] and Cargile [18].

This answer may be alright as far as it goes, but what if the paradoxer returns with a Liar strengthened in a different direction, namely *A*: 'every occurrence of the same type as *A* is not true'? Let *B* be *any* token of the same type as *A*, then we can argue that if *B* is true every occurrence of the same type as *A* is not true, so *B* is not true.[3] It follows that some other occurrence of the same type as *A* is true. Call it *C*, and arguing as before that if *C* is true it is not true, we arrive, provisionally, at a contradiction. All this 'reasoning within the paradox', to use Mackie's phrase, is conditional on none

3 *Added after completion*: Graham Priest informs me that this move is anticipated, then criticised, in articles by Hazen and Hinkfuss respectively in *Can.J.Phil.*Vol. 17 (1987) and Vol. 21 (1991).

of the items malfunctioning. But *B* and *C* and the unnamed token of the same type used in the course of the argument, are obviously on a par. And there seems no reason to query '*B* / *C* is true' or '*B* / *C* is not true' if *B* and *C* themselves do not malfunction. So the initial conclusion is, among other things, that *B* malfunctions. This time, therefore, there is no logical high ground for us commentators to occupy.

It seems then that distinguishing between tokens is not in the end a satisfactory answer, and for simplicity's sake we may ignore it even for the standard strengthened Liar. So consider *A*: '*A* is not true', where this is spelt out as '*A* is false or malfunctions'. Concluding that *A* malfunctions, how can we avoid the further conclusion that *A* is not true; with the consequent re-entry into paradox? Answer: this conclusion depends on the inference '*A* malfunctions. Therefore *A* is false or malfunctions'. But it's not that *A* is false or malfunctions; it's that '*A* is false or malfunctions' malfunctions. The fact that the conclusion of the inference fails to express the proposition which its form would suggest, undercuts the appeal to form on which the inference relies for its validity. Sod's law trumps the law of or-introduction, just as it does when sentences are ambiguous or context-dependent. Cf. Mackie & Smart [11] on the inevitable shortcomings of 'a mechanical sort of logic'. As Priest says of his own solution (p.132), 'If this is all disconcertingly non-algorithmic, this is just an unfortunate fact of life'.

To summarise: *A* does not succeed in saying that it is not true; nothing can. Can *we* say that *A* is not true? If the negation is propositional, an attempt to deny the proposition expressed by '*A* is true', then no; there is no proposition to deny. (The argument for this is the reverse of what has just been said: if '*A* is true' did express a proposition, it could be denied, which would lead back into the paradox. It is not suggested that '... is true' necessarily malfunctions whenever... does.) But if the negation is polemical, then yes. Since it malfunctions too, '*A* is true' is not the right thing to say about *A*. *A* is not true—*A* malfunctions.[4] But it's not that *A* is not true; it's that '*A* is true' and '*A* is not true' both malfunction.

4 Alternatively: 'A is not *true*—A malfunctions'. Unless supplemented by phonetic notation, writing is a crude medium for differentiating the two kinds of negation. Cleft syntax apart, the choice is between the same form for both, with the attendant potential

As I said, this solution is far from ideal. But it does not need to be ideal to serve its purpose, which is to undermine the 'there is no alternative' case for dialethism. All it needs for that is (i) to stand comparison with the dialethic solution and (ii) avoid Priest's charge that the Strengthened Liar entails a self-defeating recourse to a metalanguage. As to (i), there is a striking symmetry between the rival solutions. I take the law of contradiction for granted and conclude that the Liar sentence must malfunction, without explaining the phenomenon. Priest takes for granted that grammatical sentences cannot malfunction and concludes that the Liar sentence must be both true and false, without (I say) explaining the phenomenon. As to (ii), I make no claim to be able to say or think something (that the Liar sentence is not true) which is ineffable in English. There is no solver's metalanguage, only the one language, one's only language; and if its sentences malfunction, or what seemed to be a coherent thought turns out not to be, that may be a matter for puzzlement or regret, but scarcely for criticism.

III

Using reductio ad absurdum against someone who accepts true contradictions sounds as futile as Brer Fox's throwing Brer Rabbit into the brier-patch. But contradictions are not the only pivot for a reductio proof. Showing that a theory leads to triviality, in the shape of the proposition that everything is true, would be a knock-down argument against it. For polemical purposes, however, the proof needs to be sound in terms of the opponent's logic. No good taking the classical short cut from A and $\sim A$ to B, for the opponent will naturally have adopted a 'paraconsistent' logic, designed to block this route. I shall offer three such reductio proofs.

Rescher, Priest and Routley differ from the run of paraconsistent logicians in believing that naive set theory is true (and not merely inconsistent but non-trivial, like a corpse with an interesting smell). *The first proof* shows that it leads to triviality in each of their

for confusion, or the plonking emphasis of italics. Speech is altogether better, the difference being marked yet subtle enough for jokes to turn on it; cf. Horn's nice account (p.373) of the play between propositional and polemical negation in 'That was no lady, that was my wife'.

systems. Rescher concedes the point if the troublesome axiom is left as a single formula; consequently he needs to split it into halves which can plausibly be assigned to different assertion zones:

> it becomes necessary on the present approach to paradox avoidance that the Fregean Comprehension Axiom (that for any property ø there exists a set consisting of *all* and *only* those things which possess ø) must be broken apart into its two components (all, only): (i) there is a set containing all ø-bearers, and (ii) there is a set containing only ø-bearers which is a subset of any set containing all ø-bearers. ([6], p.164)

But of course the axiom cannot be broken apart in this or any other useful way. For it says that the *same* set contains all and only the ø-bearers, and this feature is inevitably lost when the axiom is split up. In the quoted passage this becomes obvious once (ii) is stripped of its obfuscating tautology (a set containing only ø-bearers is bound to be a subset of any set containing all ø-bearers). What is left is the naked fallacy of equating $(\exists x)(A \& B)$ with $(\exists x)A$ & $(\exists x)B$.

Priest's semantics for set theory (p.184) includes this extensionality principle: '$\{x \mid \alpha\}=\{x \mid \beta\}$ is false iff for some closed term t, $\alpha(x/t)$ is true and $\beta(x/t\}$ is false or vice versa'. Let t be $\{x \mid x \notin x\}$ and let α and β both be $x \notin x$. By Russell's paradox $\alpha(x/t)$ is true and $\beta(x/t\}$ false, hence $t{\neq}t$. Now let Fx be $x{=}t$, and consider 'there is one F' and 'there are two Fs'. Since dialethism can handle the conceptual apparatus of first-order logic with no major surprises,[5] these can be spelt out in the usual way as $(\exists y)(x)(Fx \leftrightarrow x{=}y)$, and $(\exists y)(\exists z)(y{\neq}z\ \&\ (x)(Fx \leftrightarrow x{=}y \vee x{=}z))$. The first is provable by straightforward existential generalization from the theorem $(x)(x{=}t \leftrightarrow x{=}t)$, and the second likewise from the theorem $t{\neq}t\ \&\ (x)(x{=}t \leftrightarrow x{=}t \vee x{=}t)$. But the fundamental principle for making sense of arithmetic, i.e. for relating the 'pure' use of numerals as nouns to their 'applied' use as adjectives, is this: the number of Fs is n if and only if there are n Fs. Applying it to the present case gives 2=1. By subtraction $n{=}0$ for any n, and taking n to be the number of untruths produces the triviality conclusion. The same argument applies to the extensional dialectical set theory of Brady & Routley's [19]. Their belief that it

5 So Priest says (pp.98, 115), though I think I have found one: because the logic makes $x{\neq}x$ satisfiable, every theory has a finite model. For example, Peano's axioms hold for a single individual, under the interpretation in which $s0 = 0 \neq s0$.

is non-trivial comes from treating it in isolation from the arithmetical principle cited above.

The remaining triviality proofs are aimed at dialethism. *The second proof* is a paradox about disprovability. Where a proof of *A* is a suitable series of assertions (perhaps laced with rejections) culminating in the assertion of *A*, a disproof of *A* is a series of assertions and rejections culminating in the rejection of *A*. For a classical thinker, a disproof of something is identical to a proof of its negation, but a dialethist needs to keep them separate. Priest (pp.50ff.) has argued that the ordinary informal notion of proof leads to contradiction; I shall argue in the same spirit that it leads to triviality. I consider *A*: '*A* is disprovable', show that *A* is both provable and disprovable, and infer *B* for arbitrary *B*. The proof runs like this:

	(1)	If *A* is disprovable, '*A* is disprovable' is provable,
i.e.	(2)	If *A* is disprovable, *A* is provable;
so	(3)	If *A* is disprovable, *A* is provable and disprovable.
But	(4)	'*A* is provable and disprovable' is disprovable,
so	(5)	'*A* is disprovable' is disprovable,
i.e.	(6)	*A* is disprovable.
Hence	(7)	*A* is provable
and so	(8)	'*A* or *B*' is provable.
But	(9)	If '*A* or *B*' is provable and *A* is disprovable, then *B* is provable,
so	(10)	*B* is provable,
so	(11)	*B*.

Only steps (1) and (9) call for extended comment. For the rest, (2) follows from (1) by the definition of *A*, and (3) follows from (2) by straightforward (dialethically acceptable) propositional logic. (4) is an instance of the dialethic principle that joint rational acceptability and rejectability are incompatible (p.128), provability and disprovability being what rational acceptability and rejectability amount to when the context of inquiry is one of demonstration. (5) follows from (3) and (4) by the principle that anything that implies something rationally rejectable is itself rationally rejectable (p.130). (6) follows from (5) by the definition of A, and (7) follows from (2) and (6) by modus ponens, dialethically acceptable here since the 'ifs' are not mere material implications. (8) follows straightforwardly from (7), and (10) from (6), (8) and (9). (11) is immediate from (10).

Step (1) is best considered together with its twin (1b): if A is provable, 'A is provable' is provable. It would be wrong to offer them as instances of general iterative principles about acceptability and rejectability, for in general I can have grounds for accepting or rejecting something without having grounds for recognising that I have those grounds. But demonstration is a special case, since 'it is part of the very notion of proof that a proof should be effectively recognisable as such' (p.51). So compare (1b) with this: if A and 'if A, B' are provable, B is provable. The justification for this latter is that whenever I have a proof of A and a proof of 'if A, B', I automatically obtain a proof of B by juxtaposing them and adding an assertion of B. But (1b) can be justified in a similar way. For whenever I have a proof of A then, since a proof is a proof only when it is recognised as such (p.52), I automatically obtain a proof of 'A is provable' by adding its assertion. Cf. Heyting's 'if p is proved, the provability of p is proved' [20]. And what goes for proof goes for disproof. Whenever I have a disproof of A, I automatically obtain a proof of 'A is disprovable' by adding its assertion; therefore (1).

Step (9) is an application of the principle R quoted in section I. Priest has stressed that even if joint acceptance and rejection were possible, as is envisaged here with respect to A, 'principle R would not be undercut...Some confusion may arise from the thought that something's being rationally acceptable (as well as rejectable) "cancels out" its rational rejectability. This is just a confusion. If something is rationally acceptable and rejectable, it is still rationally rejectable. Any consequences that this fact has still, therefore, stand' (p.142).

The third proof is a strengthened Liar paradox. Classically, semantic evaluations are functions mapping sentences to truth and falsity, conventionally represented by 1 and 0. To allow for the overlap between truth and falsity, dialethism (p.94) substitutes functions to three truth-values $\{1\}$, $\{0\}$ and $\{1, 0\}$. Using the evaluation v which gives sentences their actual truth-values, the functional notation can express 'A is true' and 'A is false' as

'$1 \in v(A)$' and '$0 \in v(A)$' (p.95), and the Tarski T-scheme becomes '$1 \in v(A)$ iff A'.

Now consider A: 'A is false only'. 'False only' is expressed in the function notation by '$v(A)=\{0\}$'. Since the truth-values are exhaustive, $v(A) = \{1\}$ or $\{0\}$ or $\{1, 0\}$. In the first and last cases, given that $1 \in v(\mathrm{A})$, we can derive A by the T-scheme, then $v(A)=\{0\}$ by the definition of A. In the middle case, given that $v(A)=\{0\}$ we derive A by definition, then $1 \in v(A)$ by the T-scheme. In each case, therefore, $v(A)=\{0\}$and $1 \in v(A)$, whence $1 \in \{0\}$ and so $1=0$. By dialethically acceptable reasoning we conclude unconditionally that $1=0$. It follows that, whatever B may be, $1 \in v(B)$ and consequently B.

Priest had at one time [21] considered A, concluding that some sentences 'are so contradictory as to take impossible values such as both true and false ($\{1, 0\}$) and true only ($\{1\}$)'. But he seems not to have been alarmed, and one sees why. Dialethism is not simply a theory *about* contradictions; it requires the theorist himself to assert some, and the discovery that a sentence both must and cannot take exclusive values looks like just one more contradiction to take on board. So instead of milking the conclusion that $\{1,0\}=\{1\}$, he observed that 'if we allow a sentence to take two values which are mutually exclusive, there seems to be no reason why we should not allow them to take an arbitrary number'. He therefore explored the effect of repeating his original manoeuvre (the construction which substituted 2^2-1 values for the classical 2), by substituting 2^3-1 values for 3 and so on. But having shown that after the first dramatic step the construction makes no difference to the resulting logic, he seems not to have thought it worth persevering with, and only three values survive in [7].

The lesson I draw from pressing the paradox to yield triviality is that a dialethist must stop talking about 'the truth-value of A', and stop treating evaluations as functions with truth-values as objects. Instead of truth-tables, truth-conditions must be formulated strictly adjectivally: '$\sim A$ is true iff A is false' etc. If that were all it might not be too bad, but evaluation functions are only a special case of this pattern

$f(x)$ = such-and-such if $A(x)$

$f(x)$ = so-and-so if not.

Mathematics teems with such functions. Textbooks use dozens to establish the facts about continuity and differentiability, and a notable example from logic is the construction in Henkin's completeness proof: $\Delta_{n+1} = \Delta_n + A_n$ if...; $\Delta_{n+1} = \Delta_n$ otherwise. For examples from Priest's own writings see [7], p.160, and [22], pp.250 & 258.

The dialethist's problem is that he can never be sure that these are genuine, single-valued functions. For if $A(x)$ and $\sim A(x)$ can both be true, the single-valuedness condition $y=fx$ & $z=fx \rightarrow y=z$ leads to intolerable results like our 1=0 above. A parallel condition with $\rightarrow$ read as material implication is provable, but because of the failure of modus ponens it is too weak to qualify as a satisfactory expression of single-valuedness. And substituting sets as values would only defer the problem, for at the end there is no way back to a unique member.

REFERENCES

[1] R.L.Sturch, 'Is there any reason for saying that in Australia the winter is in the summer?', *Why?*, June 1958.
[2] R.Routley & R.K.Meyer, 'Dialectical logic, classical logic and the consistency of the world', *Studies in Soviet Thought*, Vol.16, 1976.
[3] M.I.Finley, 'Logic', in *A Dictionary of Marxist Thought*, ed. T.Bottomore (Oxford, Blackwell, 1983).
[4] R.Routley & R.K.Meyer, 'The semantics of entailment', in *Truth, Syntax and Modality*, ed. H.Leblanc (Amsterdam: North-Holland, 1973).
[5] R. & V.Routley, 'Negation and contradiction', *Revista Colombiana de Matematicas*, Vol. XIX, 1985.
[6] N.Rescher & R.Brandom, *The Logic of Inconsistency* (Oxford: Blackwell, 1980).
[7] G.Priest, *In Contradiction* (Dordrecht: Nijhoff, 1987).
[8] L.R.Horn, *A Natural History of Negation* (Chicago: Univ. of Chicago Press, 1989).
[9] H.Price, 'Why "Not"?', *Mind*, Vol.XCIX, 1990.
[10] A.Gupta, 'Truth and paradox', *Journal of Philosophical Logic*, Vol.11, 1982.
[11] J.L.Mackie & J.J.C.Smart, 'A variant of the "heterological" paradox', *Analysis*, Vol. XIII, 1952–3.
[12] J.L.Mackie & J.J.C.Smart, 'A variant of the "heterological" paradox: a further note', *Analysis*, vol. XIV, 1953–4.
[13] W.C.Kneale, 'Propositions and truth in natural languages', *Mind*, Vol. LXXXI, 1972.
[14] J.L.Mackie, *Truth, Probability and Paradox* (Oxford: OUP, 1973)
[15] C.H.Whiteley, 'Let Epimenides lie!', *Analysis*, Vol.19, 1958–9
[16] L.Goldstein, '"This statement is not true" is not true', *Analysis*, Vol. 52, 1992.
[17] G.E.Hughes, *John Buridan on self-reference* (Cambridge: CUP, 1982)
[18] J.T.Cargile, review of [7], *Canadian Philosophical Reviews*, Vol.IX, 1989.

[19] R.T.Brady & R.Routley, 'The non-triviality of extensional dialectical set theory', in *Paraconsistent Logic*, ed. G.Priest, R.Routley & J.Norman (Munich, Philosophia Verlag, 1989)
[20] A.Heyting, 'The intuitionist foundations of mathematics', in *Philosophy of Mathematics*, ed. P.Benacerraf & H.Putnam (Oxford, Blackwell, 1964)
[21] G.Priest, 'Hyper-contradictions', *Logique et Analyse*, No.107, 1984
[22] G.Priest, 'To be and not to be: dialectical tense logic', *Studia Logica*, Vol.XLI, 1982.

CAN CONTRADICTIONS BE TRUE?

Timothy Smiley and Graham Priest

II—Graham Priest

> *Smiley returned to the kitchen and rinsed his face, then remembered that he had come to fetch water for the whiskey. Settling again in his arm chair, he trained his magnifying glass on the second of the men... The whiskey was keeping him awake, but it was also putting him to sleep.*
>
> John le Carré[1]

I

Introduction. The view that some claims are neither true nor false is of ancient ancestry, going back at least to Aristotle, and has been discussed by logicians ancient, medieval and modern; appropriate modern formal logics, those with truth value gaps, also go back to the foundation years of the subject, with the work of Łukasiewicz. The dual view that some claims are both true and false (dialetheism) is of equally ancient lineage, going back at least to some pre-Socratics. With a few exceptions, however, it has been largely ignored by logicians, ancient, medieval and modern; and the appropriate modern formal logic, paraconsistent logic, is a creature of the last 30 years.

Why there is such a disparity concerning views that are so obviously symmetrical is an interesting socio-historical question. At any rate, the situation is now starting to change. Over recent years a handful of logicians have argued the dialetheic case. By and large, critics have not been swift in taking up the challenge. This is a pity, at least from one perspective: whether or not dialetheism is correct, a discussion of the questions it raises, concerning fundamental notions like negation, truth and rationality—questions that have been little asked for two millennia—can hardly fail to deepen our understanding of these notions.

1 *Smiley's People*, p.130.

Smiley's paper[2] provides easily the most acute critique of the view so far, and raises a number of these questions. There is no hope of addressing all the issues raised in a reply of this length. The bulk of Smiley's arguments target my views (mainly those expressed in *In Contradiction*[3]), and I will restrict myself to discussing what he says about these. Even so, there is not space to say everything pertinent. I will take up what seem to me to be the major issues (in the order he raises them) and leave the minor ones. For these reasons, silence should not be construed as consent.

II

Negation and Denial. The first issue concerns negation, as raised by Smiley in his Section I. Smiley and I do not disagree over the definition of this. For both of us, negation is an operator that toggles truth and falsity. Where we disagree is with its relation to rejection and denial. We should start, in fact, by clearly distinguishing these two notions. Denial, and its dual, assertion, are linguistic acts. As Frege, Austin, and others have taught us, they are force operators, or kinds of illocutionary act. Rejection, and its dual, acceptance, are, by contrast, mental states. Paradigmatically, of course, assertion and denial express acceptance and rejection, or some Gricean sophistication thereof (*IC*, p.79). But we still need to keep both the distinction and the connection clear.

According to Smiley, negation and rejection are connected by the following principle:

accepting $\neg\alpha$ is equivalent to rejecting α (*)

where equivalence is to be understood in the sense that anyone who does one of these things is rationally committed to doing the other. And certainly, this is a view that any dialetheist must find problematic, since they want to accept certain claims and their negations. In fact, Smiley endorses the stronger claim that any acceptance of $\neg\alpha$ actually *is* a rejection of α, though this is never essential to his argument.

Smiley does not defend (*) himself, but appeals to Price [1990], where a broadly evolutionary story about negation is told and the

2 [1993]. Page references are to this, unless otherwise indicated.

3 Priest [1987]. I will refer to this in what follows as *IC*.

strong form of (*) is certainly endorsed—or, rather, (*) for assertion and denial, though I do not think the distinction significant in this context. Price's story about the need for denial is exemplary, and I have no desire to take issue with it. However, he is quite clear that he gives no arguments for (*) itself. He says (p.224f):

> [These considerations] do not show that what the argument needs is something with the detailed characteristics of ordinary negation. For one thing, they do not yet explain the fact that the sign of denial seems to function interchangeably as a force modifier and as a sense modifier. Denying *P* seems to be equivalent to asserting ~*P*. If negation is primarily a sign of denial, it needs to be explained how this equivalence can hold. I mention this problem mainly to set is aside…

There seems, then, to be no case to answer here. Moreover, we can even agree that the assertion of $\neg\alpha$ *sometimes* amounts to a denial of α, without having to accept that it always does. I am unsure what value evolutionary stories have in logic, but let us pursue Price's evolutionary ideas a little further. We may suppose that a need for denial was felt, and that negation entered the language as a convenient way of expressing this. Evolution, as they say, is an ongoing process. Even those who, like Smiley and Price, think that negation always expresses denial must hold that there is more to negation than is given by (*). (*) is silent, for example, about the function of negation in contexts where it does not attach to a whole utterance. Negation must therefore have evolved further properties. But in evolution things not only gain properties; they may lose them. Hence, we may suppose, our Neolithic speakers recognised that they were, on occasion, disposed to assert $\neg\alpha$ when they accepted α, and so did not want to deny it. This might arise, for example, for no more bizarre reason than that there was persuasive evidence for both sides of the case. (We will have more *outré* examples in a moment.) In such cases the denying function of negation would have to be suspended.

So much for the argument from Price. Smiley backs this up with an *ad hominem* argument to the effect that I accept half of (*), namely, that if one rejects α one should accept $\neg\alpha$; and so the onus is on me to explain why I reject the other half. I find this a rather strange argument. I don't see why, if someone accepts that all cows are animals, they incur an onus to explain why they don't think that all animals are cows. The rationale for the half of (*) I accept is clearly explained in *IC* and depends on a careful argument that there are no truth value gaps. Whether or not this argument is correct, it is not one that can be reworked to support the converse claim.

Smiley's arguments for (*) seem to me, therefore, to have little force. Moreover, there appear to be good reasons why (*) is false; even, perhaps, by Smiley's own lights. (And if the weaker thesis is false, the stronger one certainly is.[4]) Dialetheism itself provides counter-examples. These obviously beg the question in this context. However, dually, the existence of truth value gaps (which Smiley himself endorses, as we shall see in Section IV) threatens (*). Suppose that α is neither true nor false (i.e., that neither α nor $\neg\alpha$ is true) and that we have good evidence for this. We certainly do not want to accept $\neg\alpha$. What is the rational attitude to have with respect to α? It would seem that it should be rejected, that is, one should refuse to accept it (*IC*, p.122).

But (*) is problematic even from a purely orthodox position. As many people have noted, there are situations where we have excellent evidence for certain contradictions; and where, moreover, this seems implicit in rationality itself. The Paradox of the Preface is an example of this: we have evidence (as good as you like) for the joint truth of a certain number of claims (e.g., those in a book), $\alpha_1,\ldots,\alpha_n$, and so their conjunction, α; but we also have very strong inductive evidence that at least one of the claims is false: $\neg\alpha$. In such situations, the rational thing to do is to accept both contradictories (whilst, possibly, noting that the situation is anomalous). If this is right—and philosophers as orthodox as Arthur Prior have thought so—someone who accepts $\neg\alpha$ certainly need not (rationally) reject α.[5]

It might be suggested that it is not rational to assert the conjunction of the α_i's, but only each severally. I think that this is incorrect: a book is a unified entity. We do not complain if, for example, someone takes assertions from different places in the book and argues from them jointly. Compare this with the Lottery

4 I still think that the intuitionist provides a refutation of the strong form of (*). The intuitionist rejects instances of the Law of Excluded Middle because there are arguments *against* them, without accepting their negations. E.g., van Dalen: 'Let us consider [various instances of the Law of Excluded Middle] and weigh the grounds for accepting or *refuting* the principle in each separate case' (Fraenkel *et al.* [1973], p.229; my italics).

5 For references, see *IC*, 7.4.

Paradox, where we certainly do not want to assert a corresponding conjunction. In any case, the point at issue is one about rationality; and if it is rational to accept each of an obviously inconsistent collection of sentences, I cannot see that the fact that the set has cardinality two makes much difference.[6]

III

Truth and Denial. I turn now to the second of Smiley's arguments in his Section I. Anyone who believes that there are truth value gaps or 'gluts', needs to distinguish between falsity and untruth. Dialetheists hold that truth and falsity overlap. They need not, I suppose, hold that truth and untruth overlap (that is, that a contradiction of a certain kind is true). But if their rationale for dialetheism includes the semantic paradoxes, and in particular the liar paradox, then exactly the same considerations will lead them to this position. The sentence λ: $\underline{\lambda}$ is untrue, would appear to be both true and untrue.[7]

Smiley has an (*ad hominem*, I take it) argument against the position, which goes as follows (p.22):

1. Truth is the *telos* of acceptance.
2. Untruth is the *telos* of rejection.
3. Acceptance and rejection are exclusive.
4. Hence truth and untruth are exclusive.[8]

Smiley, in fact, mixes the psychological and linguistic categories. I think it better to run the argument in terms of the psychological categories. Otherwise premise 3 is obviously false. It is clear that one can both assert and deny the same thing, even at the same time: I can, for example, deny over the phone that I went to the Whiskey-a-Go-Go last night whilst simultaneously assert it to you, who are watching, with a wink. Premises 1 and 3 are taken from *IC*. (Premise 1, as stated there, involves assertion rather than

6 In this connection, there are numerous occasions on which there has been a well grounded acceptance by the scientific community (and so, presumably, a rational one) of mutually inconsistent theories. See, e.g., *IC*, p.126.

7 I use underlining as a name-forming device.

8 Note that 'exclusive' here must mean more than that the conjunction cannot be true. This is something I agree with. See *IC*, p.91.

acceptance; but, again, I do not think the distinction important here.) Premise 2 is, I think, false. To say that truth is the *telos* of acceptance is to say that if a sentence appears to be true (in the light of all the evidence, etc.) one ought to accept it. Similarly, to say that untruth is the *telos* of rejection is to say that if a sentence appears to be untrue (in the light of all the evidence, etc.) one ought to reject it. Now, why should one suppose this? Smiley supports it with nothing more than a rhetorical question. It may seem plausible when one thinks of run-of-the-mill examples. But what of those singular situations, such as the one given by λ, where one seems to have evidence that a sentence is *both* true and untrue. Is it right to reject a claim of this kind? Not obviously. More plausibly, one should reject a statement if it appears to be untrue and does not appear to be true. *That* is the *telos* of rejection. As usual, untruth and falsity seem to behave in very similar ways. (See *IC*, p.124.)

Another problem with the argument is the validity of the inference to 4. If two kinds of actions are incompatible, why should one suppose that their *teloi* are also incompatible? Notoriously, there is more than one way to skin a cat: we may, in fact, achieve not merely compatible, but even the same end by very different, and incompatible, means. One may strive for peace by arming to the teeth as a deterrent, or by totally disarming, and persuading others that one is no threat.

One might try to rework the argument as follows: if something is true one ought to accept it; if something is not true one ought to reject it; but one cannot do both, so something cannot be true and not true. This reworking would fail. For a start, both conditionals are false. *Evidence* of truth (or untruth) is required. Moreover, the argument obviously appeals to the principle that ought implies can, which fails. (See *IC*, p.240, ff.) But though the argument fails, it raises an interesting question. Suppose that untruth is the telos of rejection. If λ is both true and untrue, and we have evidence of this fact, should one accept or reject it? An answer is implicit in *IC*, ch.13. There, I argued that we can be put in a bind where we are obliged to do the impossible. The main kind of obligation discussed there is legal obligation, but I made it clear that there is nothing special about this. Maybe rational obligation can produce similar binds too; rationally, we are damned if we do and damned if we don't. λ, then, should be both accepted and rejected. This is impossible. *C'est la vie.*[9] Note, however, that $\underline{\lambda}$ is very different from the majority of the paradoxes of self-reference. Most, though establishing that something is both true and false, do not establish that something is both true and untrue.

IV

A Test Case? Let us now turn to Section II of Smiley's paper. In this, he tries to undercut the case for dialetheism by taking the Liar Paradox as a 'test case' and trying to solve it. We will look at the solution in a moment, but first a few words to put the issue into perspective.

Calling the Liar Paradox a test case is rather misleading. In *IC* a case is built for dialetheism based on a number of different arguments, drawing on the logical paradoxes, Gödel's Theorem, motion, legal binds and other phenomena. Calling the Liar a 'test case' implies that it will decide the other cases. Quite clearly, it will not. Considerations relevant to the arguments concerning the logical paradoxes, for example, are unlikely to be relevant in the case of motion—or if they are, this requires to be shown.

This is true even within the category of logical paradoxes. Even within the class of *semantic* paradoxes, the Liar has quite atypical features, as we have noted. The Liar is also radically unlike the definability paradoxes, such as Berry's. Crucially, the Law of Excluded Middle (the rejection of which is closely associated with the *type* of solution Smiley espouses) is not appealed to in such paradoxes. (See *IC*, 1.8.) The same is true in spades once we consider the set-theoretic paradoxes too. It is not at all clear that considerations that are relevant in the case of the semantic paradoxes will have any force against the set theoretic paradoxes. Post-Ramsey, logicians have thought that the two sorts of paradox require quite different considerations. Hence, even if Smiley were successful in solving the semantic paradoxes, the case for dialetheism based on the set-theoretic paradoxes remains to be answered. What is more, even if separate solutions for the two sorts

9 If we took this line then Principle R (*IC*, p.141) would have to go, despite what is said in *IC*, on pain of a triviality argument of Smiley's second kind. (See Section VII.) I think it can be argued that this is both reasonable in the context, and does no significant damage to the main claims of *IC*. But I cannot take on this issue here.

of paradox can be given, the dialetheic solution still has a case, based on the fact that it gives a *uniform* solution to what has traditionally appeared to be a *single* family of paradoxes.[10]

V

The Liar. With these preliminary comments, let us turn to Smiley's solution to the Liar. This is a sophistication of a truth value gap solution, as run by Mackie and others.[11] Smiley often talks as though it is sentences that are true/false; however, I take it that, officially, he thinks that declarative sentences normally express propositions, and that it is these that are primarily true/false. Certain sentences may, however, 'malfunction', and so fail to express any proposition, or 'convey any coherent message' (p.23).

Let λ be the sentence: $\underline{\lambda}$ expresses a false proposition. If $\underline{\lambda}$ expresses a true or a false proposition we have a contradiction. However, we can deny both possibilities now: $\underline{\lambda}$ malfunctions, and so expresses no proposition. Call this *Stratagem 1*. Invoking propositions as truth bearers is a perfectly natural (though, of course, contentious[12]) move: sentences may be ambiguous, and so be used to express different messages. Even supposing that certain grammatical declarative sentences may fail to express a proposition has some rationale in the context of category mistakes (though, again, this is contentious). The work that the notion of proposition is doing here is, however, more demanding. *Prima facie*, the Liar sentence does express a proposition—and a unique one. This, according to Smiley, is an illusion (though one that even he finds difficult to shake off[13]). Now this really does require an independent argument, or we may legitimately suppose that the notion of expressing a proposition is doing something that goes beyond anything it was introduced to do, and so we lose all grip on what it is to express a proposition.[14]

10 See Priest [199+].

11 Even Mackie, however, expressed doubts about its adequacy. See Mackie [1973], p.295.

12 See, e.g., Haack [1974], ch.6.

13 He asks (p.26) 'Can we say that [$\underline{\lambda}$] is not true?'. If '$\underline{\lambda}$ is not true' does not express a proposition, it is not at all clear that this makes sense; compare: 'Can we say that quadruplicity drinks procrastination?'.

14 Note that the argument cannot be that the sentence is self-referential in some way. On this line, the sentence 'This sentence expresses no proposition' must be taken to express

But let us pass on; for we have only just started with the problems. Stratagem 1 solves the problem because, in the theoretical context provided, the liar is not formulated correctly. Let λ now be the sentence: $\underline{\lambda}$ does not express a true proposition. In the context, this can be glossed as: $\underline{\lambda}$ expresses no proposition or a false one. The claim that $\underline{\lambda}$ does not express a proposition seems to entail λ, and so contradiction ensues. What to say about this? Smiley's solution is to deny the correctness of the inference from '$\underline{\lambda}$ does not express a proposition' to λ, i.e., '$\underline{\lambda}$ does not express a proposition or expresses a false one'. In fact, any inference may fail if one of the sentences involved malfunctions. Call this *Stratagem 2*; and note that Stratagem 2 is quite distinct from Stratagem 1: there is nothing in the notion of malfunctioning, as such, which requires the rejection of a formal logic with or-introduction. (See, e.g., Goddard and Routley [1973].)[15]

I have two comments on Stratagem 2. The first is that the notion of expressing a proposition now seems to have passed beyond breaking point. '$\underline{\lambda}$ does not express a proposition' expresses a clear and true proposition. '$\underline{\lambda}$ does not express a true proposition' appears to have a content that *includes* that of the first sentence, and so it must have content. If it does not, I am at a loss to know what expressing a proposition means. Without some independent argument here, the notion of expressing a proposition seems just to have gone on holiday.

The second comment concerns the price of this solution. The class of malfunctioning utterances is not decidable (or regular in any other way, p.23); an assumption of 'functioning' may be provisional (p.24). Hence we must wave goodbye to the project of formal logic, that is, of determining a (non-empty) class of inferences that are guaranteed to be truth-preserving in virtue of their form. Dialetheism does not endorse the formal validity of *reductio ad absurdum*. Some people find this the major stumbling block to

a false proposition. Notice also the contrast with a dialetheic solution here. According to this, the liar argument does not fail, and so the issue of giving an independent reason for locating the site of failure does not arise. (Of course, dialetheists may have to explain why other arguments fail, but that is another matter.)

15 Smiley also considers, but rejects, another stratagem: different tokens of the same type (with co-referring subjects) can fail to express the same proposition. Call this *Stratagem 1.5*. Stratagem 1.5 faces all kinds of problems of its own, and I agree that it goes nowhere.

accepting it. Smiley's proposal has exactly the same effect. Any *reductio* argument will fail if one of the steps malfunctions. And *exactly* the same is true of all arguments. This makes dialetheism look quite conservative. At least *most* inferences that classical logic takes to be formally correct are so![16]

But someone who espouses Stratagems 1 and 2 is still not out of the woods yet. Never mind or-introduction; just consider the sentence: λ expresses a true proposition. In some way, we have to mark our rejection of this. If it were to express a truth, the consequences would be quite unacceptable. The natural way to do this is, of course, simply to assert that λ does not express a true proposition; but that way lies madness. How to solve this problem? Following Horn, Smiley distinguishes between two kinds of negation. Call this *Stratagem 3*. Not_1 is the familiar truth/falsity toggling operator. Not_2 (polemical negation) is less familiar. We assert not_2-α when we want to express the fact that the utterance of α is 'inappropriate' (p.21). We cannot assert that $\underline{\lambda}$ does not_1 express a true proposition: this, as we have seen, malfunctions. But we can assert 'λ does not_2 express a true proposition'. The trouble with this solution is simply that not_2 will not do the job that is required of it. One may, in uttering not_2-α, be doing no more than rejecting certain connotations or conversational implicatures of α. This is quite compatible with the sentence negated expressing a truth. Hence, asserting 'λ does not_2 express a true proposition' does not, in itself, express the appropriate attitude. Moreover, adding the rider 'In fact, it malfunctions' won't do the job either. It would if it entailed that it does not express a true proposition. But by Stratagem 2, it doesn't.

Where does this leave us? We have had to employ three separate stratagems, of varying degrees of implausibility; but, in the end, the proponent of this view is unable to express their attitude to certain key sentences. All that is left is the silence of Cratylus. This fits the familiar picture: ineffability or contradiction. (See *IC*, 1.7.)

16 Of course, as Smiley notes, assessing the formal validity of a natural language argument always presupposes a certain amount of regimentation: that ambiguities have been resolved, indexical references made uniform, etc. But Smiley's proposal disposes of the notion of formal validity even for formal languages (and regimented natural languages if these are different).

Possibly, the position can be saved, say by invoking a *Stratagem 4*; but even if this were the case, it is clear that the solution is complex and convoluted. The dialetheic solution to the Liar is, by contrast, bold and simple. I cannot deny that it requires the rejection of something to which logicians are pretty firmly attached; but after that, everything falls into place. A similar point was true of the simplification provided by helio-centric astronomy. And the conservative position would be a lot stronger if philosophers could do what they have not yet done: come up with some non-question-begging arguments as to why contradictions cannot be true; and so show that this is not simply a piece of dogma.

VI

Numbers. Let us now turn to Smiley's third section. Although certain contradictions may be acceptable, *all* contradictions are certainly not. Hence an *ad hominem* argument that demonstrated this (triviality) would be the most damning of blows. Smiley attempts three such arguments in this section. I will address them in the same order.

The first concerns numbers. Let n be any natural number, and let $\exists^n x\varphi$ be the usual first-order sentence expressing the fact that there are n x's satisfying φ. By a certain argument, Smiley constructs a formula, ψ, such that we can establish both $\exists^1 x\psi$ and $\exists^2 x\psi$. We then infer that 1=2, by appealing to 'the fundamental principle' relating the use of numerals as nouns and as adjectives, viz: the number of φ's is n iff $\exists^n x\varphi$. Triviality is supposed to follow.

I am not convinced by the triviality *dénouement* of the argument: the number of untrue statements is, presumably, not n for any finite n. However, 1=2 is bad enough. Smiley's construction of ψ depends on a certain version of the principle of extensionality which was given as optional in *IC*; but sets with an inconsistent number of members can be constructed in other plausible ways.[17] The problem with Smiley's argument, is not, therefore, here. The problem is with the appeal to the connection between numerical adjectives and nouns. The fundamental connection is not what Smiley says, but the following:

17 See, e.g., Goldstein [1992], p.110.

$\{x;\varphi\}\varepsilon n \leftrightarrow \exists^n x\varphi$

the set of φ's has number n iff there are n φ's, an analytic truth, which can even be taken as the definition of the set n. (I take numbers to be sets, but nothing much, as far as I can see, hangs on this.) Smiley slides in the assumption that a set can be a member of only one number. Without this assumption we can infer that $\{x;\psi\}\varepsilon 1$ and $\{x;\psi\}\varepsilon 2$; but we cannot infer that 1=2. Now, why, in the present context, should we suppose that sets can belong to only one number? i.e., that:

$y\varepsilon n\cap m \rightarrow n=m$ (**)

I see no reason to suppose this; indeed, the very example in question is a counter-example.

Does the fact that a set can have more than one number not play havoc with our normal practice of counting? No! We have no normal practice of counting inconsistent collections. Our normal practice is of counting quite consistent collections, such as the number of marbles in a tin; and dialetheism gives no reason at all to suppose that the numbers of such collections are not unique. Uniqueness is violated only in the case in which we are counting inconsistent totalities. And if you will try to count these, what do you expect?!

At any rate, the onus is at least on the proponent of the argument to establish (**). The most plausible argument for (**), it seems to me, is as follows. We assume, for a start, the Frege/Cantor principle that two sets have the same cardinal size iff there is a 1-1 correspondence between their members. Now suppose that $y\varepsilon n\cap m$. Take any sets of size n and m, x and z, respectively. There is a 1-1 correspondence between x and y and one between y and z. Hence there is a 1-1 correspondence between x and z. (Let us call this *Correspondence Transitivity*.) Hence n=m. (The last step is fairly immediate if numbers are defined as equivalence classes of sets under 1-1 correspondence. If they are not, more needs to be said.)

The argument I have just sketched requires a longer discussion (both logical and philosophical) than can be attempted here. However, a central problem with it concerns Correspondence Transitivity. Again, I can only indicate it here. The definition of R's being a 1-1 correspondence between x and y is a conjunction of a number of clauses, one of which is:

$$\forall a(a\varepsilon x \supset \exists b(b\varepsilon y \wedge Rab))$$

Note that there are good reasons to suppose that the conditional here must be a material one. For suppose that $\emptyset_1=\{w;\ w\neq w\wedge p\}$ and $\emptyset_2=\{w;\ w\neq w\wedge q\}$. $\emptyset_1$ and $\emptyset_2$ are empty, and hence by the Frege/Cantor condition, there must be a 1-1 correspondence between them. Standardly, of course, it is the empty correspondence, $R_\emptyset$. But to establish that:

$$\forall a(a\varepsilon\emptyset_1 \supset \exists b(b\varepsilon\emptyset_2 \wedge R_\emptyset ab))$$

we need to appeal to the fact that $\neg\exists a\ a\varepsilon\emptyset_1$ and a paradox of implication. Hence, the conditional must be material.

Now, suppose that R and S are 1-1 correspondences between x and y, and y and z respectively. We need to show that R.S is a 1-1 correspondence between x and z (where (R.S)ac iff $\exists b(aRb \wedge Sbc)$). The inference that would normally establish the relevant clause is:

$$\forall a(a\varepsilon x \supset \exists b(b\varepsilon y \wedge Rab)),\ \forall b(b\varepsilon y \supset \exists c(c\varepsilon z \wedge Rbc))$$
$$\vdash \forall a(a\varepsilon x \supset \exists c(c\varepsilon z \wedge (R.S)ac))$$

A moment's reflection shows that this inference depends on the validity of the propositional inference:

$$\alpha \supset(\beta \wedge \gamma),\ \beta \supset \delta \vdash \alpha \supset (\beta \wedge \gamma \wedge \delta)$$

or the more fundamental:

$$\alpha \supset \beta,\ \beta \supset \delta \vdash \alpha \supset \delta$$

and this is invalid in the dialetheic logic of *IC*, as may easily be checked.

Stripped of its technicalities, the point is a simple one. The transitivity of correspondence depends, unsurprisingly, on the transitivity of the conditional; and the material conditional is not transitive. Transitivity fails when the middle term, β, is both true and false. In our particular case, the middle term is of the form $b\varepsilon y$, where y is a peculiar set with both 1 and 2 members. Hence one of the things in it must not be in it. It is therefore a set of exactly the kind that one should expect to screw up the inference.

VII

Disprovability. I now turn to Smiley's second argument. A proof is a sequence of statements of a certain kind, which provides a justification for accepting the content of the assertion that is its last member. Dually, Smiley invites us to consider a disproof to be a similar sequence which provides a justification for rejecting the content of the denial that is its last member. A statement is provable if there exists a proof, and disprovable if there exists a disproof. An argument involving the statement 'This sentence is disprovable' ends, according to Smiley, in the proof of an arbitrary statement.

The argument is ingenious, but has a number of weak points. For example, the argument uses the principle:

If α then β; $\underline{\beta}$ is disprovable; hence $\underline{\alpha}$ is disprovable.

at line 5.[18] Now the principle may work if the conditional in the major premise has the force of an entailment.[19] A disproof of $\underline{\beta}$ together with the fact that $\underline{\alpha}$ entails $\underline{\beta}$ constitutes a disproof of $\underline{\alpha}$. However, a weaker conditional will not do. If the conditional is an enthymematic one, for example, with suppressed premise γ, then this is no longer the case. We no longer have a disproof of $\underline{\alpha}$ unless we have, at the very least, a proof of $\underline{\gamma}$. (And even then, a proof of $\underline{\alpha}$ would use the disjunctive syllogism.)

Now the conditional in question is inherited from line 1: if $\underline{\alpha}$ is disprovable then '$\underline{\alpha}$ is disprovable' is provable. I do not think that this is an entailment. This is easiest to see for the dual principle: if $\underline{\alpha}$ is provable then '$\underline{\alpha}$ is provable' is provable. (I agree with Smiley that these are of a piece.) The antecedent of this tells us that there is something that establishes that $\underline{\alpha}$ is true. Does it follow from this, on its own, that there is something that establishes that '$\underline{\alpha}$ is provable' is true, i.e., that $\underline{\alpha}$ is provable? No: truth does not entail provability; and the existence of something establishing truth does not entail the existence of something establishing provability. What sort of conditional we have here is an interesting question, and I am not sure that I know the answer. What grounds its truth is the fact that if the premise is true we can produce something that *shows* the consequent to be true; but it does not *say* it is (to use a happy Wittgensteinian distinction). At any rate it is not an entailment, any more than the showing of a red object entails that it is red.

18 Line numbers refer to Smiley's numbering.

19 Though even this may be doubted. It is natural to suppose that it should be required to be a provable entailment.

An even more problematic step is the premise invoked at line 4: '$\underline{\alpha}$ is provable and disprovable' is disprovable. Smiley's justification of this is simply a gloss on a quotation from *IC*. The premise is 'an instance of the dialetheic principle that joint rational acceptability and rejectability are incompatible, provability and disprovability being what rational acceptability and rejectability amount to in the context of demonstrative argumentation' (p.29). This is far too swift. Provability and disprovability may entail rational acceptability and rejectability; and if the latter pair are incompatible, so are the former. How do we get to disprovability, however? At the very least, we need an argument for incompatibility, and one, moreover, that *guarantees* the truth of its conclusion. (See the final step of the argument: $\underline{\beta}$ is provable; hence $\underline{\beta}$.) Now an argument is offered in *IC* (p.128), but even if it is good (and I now have my doubts about it) it is far from a proof. Nor do I see any hope, at present, of producing such a thing.

It might be thought that this is a simple evasion. After all, why can't we just run the argument replacing 'provable/disprovable' with 'rationally acceptable/rejectable'? The final step of the argument would fail, it is true; but that everything is rationally acceptable is already bad enough. If we do this, however, the argument fails elsewhere, as Smiley observes (p.30). Corresponding to step 1 of the argument, we would then have: If $\underline{\alpha}$ is rationally rejectable '$\underline{\alpha}$ is rationally rejectable' is rationally acceptable. This and its mate (if $\underline{\alpha}$ is rationally acceptable '$\underline{\alpha}$ is rationally acceptable' is rationally acceptable) are both false. $\underline{\alpha}$ is rationally acceptable if there is evidence for it; this evidence is not necessarily evidence that it is rational to accept $\underline{\alpha}$. There is good evidence for the General Theory of Relativity; but the bending of light in a gravitational field has no consequences for rationality at all.

VIII

Definition by Cases. Let us turn, finally, to Smiley's third triviality argument. As *IC* (1.7) argues, 'extended paradoxes' return to haunt all solutions to the liar paradox; we have already seen this with Smiley's own attempt. Dialetheism escapes this problem. In this context, a strengthened liar is a sentence, α, of the form: $\underline{\alpha}$ is false only, i.e., false and not true. Applying the T-schema, with obvious notation, we have:

$$T\underline{\alpha} \leftrightarrow (F\underline{\alpha} \wedge \neg T\underline{\alpha})$$

And using the facts that $T\underline{\alpha} \vee \neg T\underline{\alpha}$ and $(\neg T\underline{\alpha} \rightarrow F\underline{\alpha})$, we establish that $(T\underline{\alpha} \wedge \neg T\underline{\alpha})$. But this is no problem. The aim of a dialetheic approach to the paradoxes is to accommodate contradictions such as this, not eliminate them.

Smiley raises the question of what happens when we consider a parallel argument concerning, not truth, simpliciter, but truth-in-an-interpretation, where interpretations are formulated as in *IC*. Letting ν_0 be the assignment of truth values that is in accord with the actual; the T-schema then becomes:

$$1 \varepsilon \nu_0(\underline{\alpha}) \leftrightarrow \alpha \qquad (***)$$

a self-referential instance of which gives us:

$$1 \varepsilon \nu_0(\underline{\alpha}) \leftrightarrow \nu_0(\underline{\alpha})=\{0\}$$

A bit of calculation then gives 0=1, and all Hell breaks loose. One way to deal with this problem is simply to admit the inadequacy of the semantics of *IC*. Classical truth values are in the set $t_0=\{0,1\}$. LP truth values are in the set, t_1, of all non-empty subsets of t_0. To accommodate the kind of phenomenon in question, it might be thought, we should take our truth values to be in the set, t_2, of all non-empty subsets of t_1. Of course, the same kind of phenomenon will force us to ascend to the next level of the hierarchy, and so on. What happens at the limit is investigated in Priest [1984], as Smiley notes. This construction accommodates extended liars of any ordinal level. Possibly, an extended liar of absolute level is constructible; but the details of the construction are complex enough to make this not at all obvious.

A second line of solution is suggested by Smiley himself. Accounts of truth-in-an-interpretation obviously have an element of conventionality about them. For example, the choice of 0 and 1 as truth values is obviously arbitrary; more importantly, here, the representation of truth values as sets is equally arbitrary. We could do *exactly* the same job if we expressed truth-in-an-interpretation conditions in relational terms. More specifically, an interpretation is now conceptualised as a binary relation, V, of a certain kind, with domain the set of formulas and co-domain the set $\{1,0\}$. We can then state truth conditions, etc., in the obvious way.

When we run the argument now, we arrive at a sentence, α, such that:

$$V_0\underline{\alpha}1 \leftrightarrow (V_0\underline{\alpha}0 \wedge \neg V_0\underline{\alpha}1)$$

Arguably, as before, we then obtain:

$$V_0\underline{\alpha}1 \wedge \neg V_0\underline{\alpha}1$$

But this seems no more problematic than the absolute truth version. The catastrophic results therefore appear as spin-offs of a conventional form of the representation used, not of the facts themselves.

It might be suggested that this approach does not really solve the problem, but merely hides it. In particular, even if we do not give our semantic values as sets, there can be no objection to deploying set-theoretic machinery in conjunction with what we already have; and then the problem reappears. Does it? Let $d=\{x;\ V_0\underline{\alpha}x\}$. If we can show that $1\varepsilon d$ and $d=\{0\}$, the problem does, indeed, reappear. The first of these is equivalent to $V_0\underline{\alpha}1$, which we have. The second is equivalent to:

$$x=0 \leftrightarrow V_0\underline{\alpha}x$$

I see no way of establishing this validly from things that are uncontentiously true. For example, the natural argument from right to left goes: suppose that $V_0\underline{\alpha}x$; then since $\neg V_0\underline{\alpha}1$, $x\neq 1$; hence $x=0$ since $V_0\underline{\alpha}x \rightarrow (x=0 \vee x=1)$; but the first inference is dialetheically invalid (see Priest [1991a], p.194) and the second uses the notorious disjunctive syllogism.

Let us now turn to what Smiley himself makes of this situation. The dialetheist must, according to him, stop treating truth values as objects, and evaluations as functions. The first of these points is false: though we give truth conditions 'adjectivally', there is nothing to stop us talking about the set d and its ilk; if we exercise sufficient care, there is no harm in this. The second point is, however, correct. Smiley concludes that dialetheists must eschew functional definition by cases. This is far too swift a jump: from the fact that evaluations are no longer thought of as functions, it certainly follows that we cannot use functional definition by cases to specify them. It does not follow that functional definition by cases must be eschewed altogether. However, the claim that if we were allowed the normal use of functional definition by cases, trouble would arise, is correct. Consider, for example:

$$f(x) = 1 \text{ if } \varphi(x)$$
$$\quad = 0 \text{ if } \neg\varphi(x)$$

Take x to be the Russell set, r, and $\varphi(x)$ to be $x \varepsilon x$. Then we immediately have $f(r)=1$ and $f(r)=0$, and so $1=0$. Of course, what has gone wrong here is quite transparent; a presupposition of definition by cases is that the cases are exclusive and exhaustive. Given dialetheism, this presupposition fails. So, must definition by cases be eschewed? Yes and no. We may always think of functions, officially at least, as set-theoretic entities of a certain kind. A functional definition by cases is then a simple instance of the abstraction scheme. For example, the above becomes:

$$<x, y> \varepsilon f \leftrightarrow (\varphi(x) \wedge y=1) \vee (\neg\varphi(x) \wedge y=0)$$

In this sense, we may always use definition by cases. But we cannot assume, as we always can classically, that the relation thus defined is functional, i.e., that $(<x,y> \varepsilon f \wedge <x,z> \varepsilon f) \rightarrow y=z$. To establish this in the above case, for example, would require a quite illicit use of the disjunctive syllogism. In that sense we cannot use unrestricted functional definition by cases.

This points to a final flaw in Smiley's argument. We must define the evaluation ν_0, used in (***), as follows:

$$<\underline{\alpha}, x> \varepsilon \nu_0 \leftrightarrow (T\underline{\alpha} \wedge x=1) \vee (T\neg\underline{\alpha} \wedge x=0)$$

The standard T-schema then gives us that:

$$<\underline{\alpha}, x> \varepsilon \nu_0 \leftrightarrow (\alpha \wedge x=1) \vee (\neg\alpha \wedge x=0)$$

But (***) itself (i.e., $<\underline{\alpha}, 1> \varepsilon \nu_0 \leftrightarrow \alpha$) can no longer be established; the argument for it uses the disjunctive syllogism.

Does the fact that we cannot unreservedly use functional definition by cases matter? No. In many cases where we define a function by cases, its functionality is never really used, especially if the value of the function is a set. Instead of reasoning about the value of the function for a certain argument, we can reason about the union of all its relata. However, if uniqueness cannot be established dialetheically, and further reasoning depends essentially on this, we will have to admit that it is invalid. Does this cripple reasoning about functions? Not at all. Wherever we have a classically valid argument for uniqueness, and a consistent context—and these are the only ones that make much sense to a classical logician anyway—the argument will be perfectly acceptable to a dialetheist (as explained in *IC*, 8.5, 8.6).[20] Thus, a dialetheist has a quite unproblematic understanding of the legitimate use of classical definition by cases. As ever, a dialetheist can do anything a classical logician does; but where classical reasoning fails, in inconsistent situations, the dialetheist is not left bereft of the means to proceed, but can explore further.

IX

Conclusion. I have now addressed all the major arguments deployed by Smiley. As is clear, even if what I have said is correct, there is a good deal more to be said about them. And the issues are complex enough to make dispute about whether it is correct, entirely proper. Dialetheism is a view that has been widely (though quite incorrectly) viewed as absurd, since Aristotle stamped his magisterial authority on logic. It would be remarkable indeed, if, in crafting a case for it, one managed to get it exactly right first time. However, two other things will, I hope, also be clear. The first is that Smiley's paper raises issues of fundamental importance for anyone concerned with the foundations of logic. The second is that they leave the case for dialetheism constructed in *IC* undented.[21]

REFERENCES

A. Fraenkel, Y. Bar-Hillel and A. Levy [1973], *Foundations of Set Theory* (2nd ed), North Holland

L. Goldstein [1992], 'Smooth and Rough Logic', *Philosophical Investigations* 15, 93–110

L. Goddard and R.Routley [1973], *Logic of Significance and Context*, Scottish Academic Press

S. Haack [1974], *Philosophy of Logics*, Cambridge UP

J. Mackie [1973], *Truth, Probability and Paradox*, Clarendon Press

H. Price [1990], 'Why "Not"?', *Mind* 99, 221–38

20 I now think that a much better account of the classical recapture is to be given in terms of minimal inconsistency. See Priest [1991b].

21 I would like to thank a number of people for helpful comments, but especially Stewart Candlish, André Gallois, Roger Lamb, Greg Restall, Huw Price, Richard Sylvan and last, but by no means least, Timothy Smiley himself.

G. Priest [1984], 'Hyper-Contradictions', *Logique et Analyse* 27, 237–243
G. Priest [1987], *In Contradiction*, Nijhoff
G. Priest [1991a], 'Intensional Paradoxes', *Notre Dame Journal of Formal Logic* 32, 193–211
G. Priest [1991b], 'Minimally Inconsistent LP', *Studia Logica* 50, 321–31
G. Priest [199+], 'The Structure of Paradox', a paper read to the 1991 meeting of the AAL; to appear
T. Smiley [1993], 'Can Contradictions be True?, I', this volume