Everett's Trilogy

GRAHAM PRIEST

1. Introduction

In three recent papers, Everett (1993, 1994 and 1996), Anthony Everett has produced a number of interesting and interrelated arguments against a dialetheic solution to the semantic paradoxes. The purpose of this paper is to assess the effectiveness of these arguments. The relationship between the papers is roughly this: Everett (1994) argues that a dialetheic solution to the paradoxes falls foul of certain kinds of Curry paradox (left pincer); Everett (1993) argues that it falls foul of extended paradoxes (right pincer); Everett (1996) argues that any way of avoiding one pincer drives one firmly into the other (the fork). In the next section I will set up the background for the discussion. I will then take the papers in turn. I will argue that Everett's strategic onslaught fails, on all fronts. There are, however, interesting lessons to be learned from his campaign.

2. Background

The paradigm semantic paradoxes are set up with a truth predicate, T, and some sort of fixed-point construction. If we use angle brackets as notation for a naming device, then a liar sentence is one of the form λ, where $\lambda \Leftrightarrow \neg T\langle\lambda\rangle$. Applying an instance of the T-schema, $T\langle\lambda\rangle \Leftrightarrow \lambda$, and transitivity gives $T\langle\lambda\rangle \Leftrightarrow \neg T\langle\lambda\rangle$, whence the law of excluded middle (or various other principles) gives $T\langle\lambda\rangle \wedge \neg T\langle\lambda\rangle$. If we let $\perp$ be the absurdity constant, which entails everything,[1] then the Curry paradox is a sentence of the form κ, where $\kappa \Leftrightarrow (T\langle\kappa\rangle \Rightarrow \perp)$. Using the T-schema for κ and transitivity gives us $T\langle\kappa\rangle \Leftrightarrow (T\langle\kappa\rangle \Rightarrow \perp)$. The principle of absorption (or contraction)

[1] This makes perfectly good sense dialetheically. See Priest 1987, §8.5.

$(\alpha \Rightarrow (\alpha \Rightarrow \beta) \vdash \alpha \Rightarrow \beta)$ then gives us $T\langle\kappa\rangle \Rightarrow \bot$, and a couple of applications of *modus ponens* give $\bot$.

In Priest (1987) and elsewhere, I have given an account of how I think these paradoxes should be handled. The liar paradox is sound: the contradiction at its conclusion is true. What is challenged is the view that this is unacceptable. In particular, the principle of inference *ex contradictione quodlibet* $(\alpha \wedge \neg\alpha \vdash \beta)$ is invalid. This is demonstrated using a semantics according to which sentences may take some nonempty *subset* of the usual truth values, $\{t, f\}$, but otherwise life is as normal in classical logic.

The situation concerning the Curry paradox is rather different. We are not at liberty to accept its conclusion. It delivers us triviality without the detour through *ECQ*. The solution advocated in Priest (1987) is simply that the conditional operator, $\Rightarrow$, of the *T*-schema, does not satisfy absorption.[2] Now there are many semantics for the conditional that invalidate absorption, e.g., algebraic or (ternary relation) world semantics for relevant logics, various semantics for linear logic, continuum-valued semantics for fuzzy logics, etc. And the philosophical foundations of these semantics is a matter of active research.[3] Pretty much any account of the conditional that invalidates absorption would suffice (Priest, 1987, p. 102). However, for the sake of definiteness, I there contented myself with a simple (binary relation) possible-worlds account. The conditional of the *T*-schema, $\Rightarrow$, is taken to be a strict implication. (The worlds are also 3-valued, with truth values $\{t\}$, $\{t,f\}$ and $\{f\}$—though this is not pertinent to the discussion of conditionality.) The accessibility relation on worlds is not unrestrictedly reflexive, and a counter-model to absorption is then easy to construct. If (and only if) reflexivity fails at world w of some interpretation, then it is quite possi-

[2] There is another solution, recently advocated by Goodship (1996), that is in some ways more appealing. This is that the conditional of the *T*-schema is simply a material conditional, $\supset$. Absorption for the material conditional holds, but detachment fails. Hence, the Curry paradox is blocked. In fact, $T\langle\kappa\rangle \supset \bot$ is logically equivalent to $\neg T\langle\kappa\rangle$ and so the Curry paradox collapses into the Liar paradox.

[3] Everett (1996, fn. 9) notes the work on contraction-free logics, but claims that this does not undermine the intuitive plausibility of absorption, since the various semantics are not plausible candidates for those of natural language conditionals. Now many of the semantics in question operate with highly intuitive notions such as sense and information-flow, whose relationships with conditionality are intimate ones. Though I do not claim that this is yet the case, I think it very likely that developments in these semantics may come to undermine the apparent plausibility of absorption for some kinds of conditionals, in just the way that Seventeenth Century developments in dynamics undermined the apparent plausibility of various Aristotelian claims about motion, such as that things in motion naturally come to rest.

ble to have α and $\alpha \Rightarrow \beta$ true at w, without having β true there. The rest is routine.[4]

Let me finish these background considerations with one further comment. There is an important conceptual distinction between truth and truth-in-an-interpretation, or in the case of languages with intensional connectives, truth-at-a-world-in-an-interpretation. And this is often forgotten in discussions of paradox. Truth is a property (or at least a monadic predicate); it may be invoked for many different purposes; so far in this paper, it has appeared only in the object language in which the paradoxical reasoning is performed. Truth (at a world) in an interpretation is a set-theoretic relation between a sentence and the value t; unlike truth, it is a model-theoretic notion, and has one primary use: to provide a metatheoretic definition of validity. Maybe it is a constraint on notions of interpretation that there be (a world in) an interpretation, truth at which is extensionally equivalent to truth *simpliciter*. But that is another matter.

3. *Absorption regained*

Let us now turn to Everett's (1994) left pincer which concerns Curry paradoxes. The aim here is to define some new conditional, $\sqsupset$, such that:

(a) $\sqsupset$ satisfies contraction;

(b) we have the version of the T-schema with $\sqsupset$ (or rather, its corresponding biconditional) as the main connective.

The Curry paradox then goes through as before.

Everett gives two such conditionals. Let us take the second first. (In Everett (1994) this is written as a pair of stacked '$\rightarrow$'s.) Take any interpretation; if the accessibility relation is reflexive at a world, let us call the world itself "reflexive". The truth conditions for $\sqsupset$ are simply that $\alpha \sqsupset \beta$ is true at world w iff for all reflexive worlds accessible to w, if α is true as w, so is β. If $\alpha \Rightarrow \beta$ is true at world w then at all worlds accessible to w where α is true, β is true. *A fortiori*, then, $\alpha \sqsupset \beta$ is true at w. Thus we have (b) since it follows from the T-schema couched in terms of $\Rightarrow$.

[4] Only a little is said in Priest 1987 about why the accessibility relation is not generally reflexive. I now think that the best story to tell is one according to which it fails at non-normal worlds, these being worlds where the laws of logic are different (making it possible to consider counter-logical conditionals); in particular, in this case, where *modus ponens* fails. See Priest 1992. The fact that it fails only at worlds where logic is *different* shows why such a failure does not count against $\Rightarrow$ being a conditional operator. (This line of thought indicates a reply to Everett 1994, fn. 3.)

Turning to (a), Everett claims that $\sqsupset$ satisfies absorption.[5] The reason given for this is that even models that invalidate the rule validate it (p. 415). For justification of this we are referred back to a passage where Everett explains why a counter-model for absorption requires reflexivity to fail. The exact nature of this argument is less than clear, since it goes by way of discussing one particular counter-model. However, there are, as far as I can see, only two distinct ways of spelling out the general argument, and both fail, as I shall now explain.

The first way is as a direct argument. Suppose that w is a world of some interpretation, and that $\alpha \sqsupset (\alpha \sqsupset \beta)$, is true at it. We need to show that $\alpha \sqsupset \beta$ is also true at it, i.e., that for all accessible reflexive w', if α is true at w' so is β. So suppose that w' is reflexive and that α is true there. Then by applications of *modus ponens*—crucially, more than one—we infer that β holds at w'. We now apply the rule of Conditional Proof (CP, if-introduction) to infer: if α is true at w', then β is true at w'.

This reasoning is classically unimpeachable. However, it applies CP in the metalanguage. Moreover, it does this where an antecedent has been used *multiple times* to deduce the consequent. This is exactly a contraction. If one were to spell out the reasoning without using natural deduction techniques one would see that at the crucial point absorption is used. (I leave this as an exercise.) Alternatively, one might just note that if one is allowed CP in this form in a natural deduction system (together with *modus ponens*) a proof of absorption is quickly forthcoming. So an unrestricted form of CP is at least as strong as absorption.

We see, then, that this metatheoretic reasoning for the validity of absorption uses absorption (or something at least as strong). As I have insisted in many places,[6] the logic of meta-theoretic reasoning must be the same as that of the object-language. Hence, reasoning in such a way is not dialetheically acceptable; and to insist on it would simply beg the question. The argument therefore fails.[7] There are, in fact, general reasons why any attempt to define a conditional that satisfies absorption must suffer the same fate, though I will not go in to this here. (A discussion can be found in Priest 1990, §7.)

[5] Or strictly speaking, the corresponding conditional, but the difference is inessential here. Note, though, that Everett's discussion starts off with the semantics being three-valued, but slides without warning into using two-valued semantics.

[6] E.g., Priest 1989, 1990, and forthcoming.

[7] Others have given trivialisation arguments which commit exactly this fallacy. For example, Denyer (1989) attempts such an argument which begs the question over the disjunctive syllogism. See Priest 1989.

4. *A semantic limbo*

The second way of interpreting Everett's argument is not as a direct, but as a *reductio*, argument and goes as follows. Suppose that we had a countermodel to absorption for $\sqsupset$. Then at some reflexive world, w, of the interpretation, we would have to have α and $\alpha \sqsupset \beta$ true, and β untrue. But since w is reflexive and $\alpha \sqsupset \beta$ is true at w, so is β. Contradiction.

The problems with this argument are twofold. First, *reductio* is not generally valid in dialetheic logic. This argument therefore fails since it begs the question, just as the first one did. Secondly, even if the *reductio* were correct, it would show only that there is no interpretation, I, and world w, in I, such that $\alpha \sqsupset (\alpha \sqsupset \beta)$ is true at w, and $\alpha \sqsupset \beta$ is not true at w. The logical form of this is: $\neg\exists I\exists w(\gamma \wedge \neg\delta)$. The statement that the inference is valid is, in the same notation, $\forall I\forall w(\gamma \Rightarrow \delta)$. And the inference from the first of these to the second is not valid for an intensional conditional, $\Rightarrow$. Again, therefore, the argument fails.[8]

This argument does raise an interesting question, however. The connective $\sqsupset$, as defined, appears to make perfectly good dialetheic sense, though it might not have the properties that a classical logician takes it to have. What properties do the semantics show it to have? It is not difficult to check through the reasoning that shows that, e.g., identity, $\alpha \sqsupset \alpha$, is valid. It is also simple to come up with counter-models for principles such as $\alpha \sqsupset (\alpha \wedge \beta)$. Now consider absorption itself. We cannot show that this is valid—provided we stick to valid reasoning. But if we try to give a counter-model, we run into contradiction. There would therefore appear to be no counter-models to it either.

It would seem, then, that there must be a class of inferences, including absorption, that can neither be shown to be valid nor be shown to be invalid by these semantics. In particular, the dialetheist is committed to saying that absorption for $\sqsupset$ is invalid, and hence to claiming that there are invalid inferences that the semantics cannot show to be so—at least, as these are interpreted classically. Situations of this kind are not unknown in non-classical logics. For example, Kripke semantics for intuitionist predicate logic are known to be complete classically, but incomplete intuitionistically, provided that Church's Thesis is intuitionistically correct.[9] Hence, for the intuitionist, the dual phenomenon arises: inferences that are valid, but a classical interpretation of the semantics cannot demonstrate them to be so.

[8] One can infer the material conditional, $\forall I\forall w(\gamma \supset \delta)$, but this will not do as a paraconsistent statement of validity; for since $\supset$ does not support detachment, neither would valid arguments. See, further, Priest 1990, §5.

[9] See Dummett 1977, §5.6, esp. p. 259.

In this situation, there are various possibilities. One is that we give up the claim that semantics are constitutive of validity. Validity is to be defined proof-theoretically; semantics are a heuristic tool only, useful in many contexts, but with limitations of which we are aware. The second is that we retain attachment to the primacy of semantics, but change the semantics to other, classically intelligible, semantics. This is always an option.[10]

The third option is the most interesting, but also the most problematic. Non-classical logics make possible mathematical structures that are classically unintelligible. It may be possible to use this space. For example, the Law of Excluded Middle, $\forall p(p \vee \neg p)$, is intuitionistically false, but no classical counter-example can be given, or this would provide an argument for $\exists p(p \wedge \neg p)$. However, using versions of the theory of choice sequences, inconsistent with the classical theory of reals, it is possible to give a counterexample to the Law.[11]

In the dialetheic case, we might try to produce counter-examples to principles such as absorption for $\sqsupset$ using inconsistent structures. In the light of the preceding discussion, it is not difficult to see how this might go. Consider the model with one reflexive world, w, such that p is true at w, and q both is and is not true at w. (Note that this specification is inconsistent.) Since p and q are true at the only accessible world, $p \sqsupset q$ is true there, as, therefore, is $p \sqsupset (p \sqsupset q)$. But since p, is true at w and q is not, $p \sqsupset q$, is not true there. Hence, the inference is invalid.

The status of this model is problematic, even from a dialetheic point of view. It requires more than that some sentence is both true and false at some world of a model, that is, that it takes the value $\{t,f\}$ there (which is quite consistent given the semantics); it requires the evaluation function itself to be inconsistent, so that a sentence both is and is not true at a world, i.e., t both is and is not in its value at that world. It is not clear that one can prove the existence of this kind of interpretation. Moreover, if models of this kind *are* legitimate, then a lot more than absorption is going to turn out to be invalid. For example the conditional $p \Rightarrow p$ is invalid (though valid as well), since in the above model p is true at w but p isn't. (This may not be as damaging as it sounds. We reason correctly with formally invalid inferences all the time: every inference is of the form $p \vdash q$, which is invalid. A good argument is one which instantiates *some valid*

[10] Boolean negation gives rise to exactly the same situation as Everett's $\sqsupset$. In this case, appropriate counter-models can be defined by changing from world semantics to an algebraic semantics. (See the appendix of Priest (forthcoming).) The same trick will not work for $\sqsupset$. Its definition makes sense only given the resources of world semantics, and so it has no obvious algebraic counterpart. Maybe so much the worse for it.

[11] See Fraenkel Bar-Hillel and Levy 1973, p. 260, or Dummett 1977, p. 84.

form of inference.) The viability of this line of thought obviously requires much further consideration; what turns out to be invalid may depend on sensitive questions about how, exactly, the semantics are set up. Moreover, to consider the issue here would take us a long way out of our way. So let us drop the subject, and return to Everett's arguments.

5. *Enthymematic conditionals*

The second way that Everett (1994) suggests of obtaining a conditional that will allow the Curry argument is with the use of a logical constant, $\mathfrak{R}$, which holds in just reflexive worlds. In effect, $\alpha \sqsupset \beta$ is now defined, enthymematically, as $(\alpha \wedge \mathfrak{R}) \Rightarrow \beta$.[12] Condition (a) is satisfied since the enthymematic biconditional now follows simply from the straight one; and, Everett (1994) claims, so is condition (b).

Before we consider this, it is worth noting that it is not at all obvious that there is a logical constant of the required kind. Why should we suppose there to be one? We are told (Everett 1994, p. 416) that we can define one explicitly, e.g., by the phrase "This world is reflexive". Now, the vocabulary of this definition is metatheoretic, and not normally taken to be a part of the object language. But as I have already argued, there can be no objection to this. It also contains demonstratives, which may be more problematic. Can we be sure that in an arbitrary world denotations of demonstratives get fixed correctly? But leaving this issue aside, this definition will not do. Assuming that demonstratives do function as required, $\mathfrak{R}$, so defined, is true at w iff "wRw" is true at w. It does not follow from this that wRw, since we do not have: φ is true at w iff φ; truth-at-w and truth are not the same thing.[13]

If language contains no expression that will already do the job, maybe we can just add a new one with the appropriate truth conditions: $\mathfrak{R}$ is true at world w iff w is reflexive. It is not clear, form a dialetheic point of view, that truth conditions of this kind succeed in giving the constant a determi-

[12] This is not quite the way that Everett (1994) sets it up. The fixed point is defined as κ, where $\kappa \Leftrightarrow (T\langle \kappa \wedge \mathfrak{R}\rangle \Rightarrow \bot)$, but since T distributes over conjunction, $T\langle\mathfrak{R}\rangle$ effectively becomes the suppressed premise of the enthymeme, and this holds at all and only the worlds that $\mathfrak{R}$ does, by the instance of the T-schema: $T\langle\mathfrak{R}\rangle \Leftrightarrow \mathfrak{R}$. I set things up in the way I have in order to make the discussion more economical.

[13] Everett (1994, fn. 5) suggests using $\forall wwRw$ as an alternative suppressed premise. This is clearly a lot stronger, but if it is true at w, so is wRw, and the rest of the argument proceeds as before. The problem here is even more transparent. From the fact that $\forall wwRw$ is true at a world, we certainly cannot infer that $\forall wwRw$ (is true).

nate sense. If they did, we could just as well introduce a constant, $\wp$, true in just those worlds that access @, the actual world. Now consider a world, w, where everything is in the extension of the truth predicate. At such a world, for any α, we have $T\langle \wp \wedge \Box\, \alpha\rangle$, and so, by the T-schema, $\wp \wedge \Box\, \alpha$. By assumption, w accesses @; hence @ is trivial. There can therefore be no such constant.[14]

Anyway, setting these worries aside, the main problem with the argument again concerns condition (b). Everett (1994) claims that $\sqsupset$, defined enthymematically using $\mathfrak{R}$, satisfies absorption. The situation is, however, exactly the same as for the first conditional we considered. In fact, this case is essentially the same as the first. The constant $\mathfrak{R}$ simply allows us to pack the truth conditions of the first conditional into the object-language. But moving from the metalanguage to the object language, like transposing the key of a piece of music, changes nothing structural.

For example, consider the direct argument for the validity of absorption, which goes as follows. Suppose that $(\alpha \wedge \mathfrak{R}) \Rightarrow ((\alpha \wedge \mathfrak{R}) \Rightarrow \beta)$ is true at world w of some interpretation. Let w' be any accessible world where $\alpha \wedge \mathfrak{R}$ is true. By *modus ponens*, $(\alpha \wedge \mathfrak{R}) \Rightarrow \beta$ is true at w'. Since $\alpha \wedge \mathfrak{R}$ is true at w', w' is reflexive. Hence by *modus ponens* again, β, is true at w'. Thus if $\alpha \wedge \mathfrak{R}$ is true at w', so is β, i.e., $(\alpha \wedge \mathfrak{R}) \Rightarrow \beta$ is true at w. Note that this argument appeals to the supposition that $\alpha \wedge \mathfrak{R}$ is true at w' *multiple* times before applying Conditional Proof. Hence it uses absorption.

6. Hypercontradictions

Let us move on to Everett's (1993) right pincer. This concerns extended paradoxes. A number of people have thought that extended paradoxes sink a dialetheic account of the semantic paradoxes, just as they sink all consistent accounts: we need only consider the sentence that says of itself that it is false and not also true. This thought is incorrect, however. If we consider a fixed point, φ, of the form $F\langle\varphi\rangle \wedge \neg T\langle\varphi\rangle$, then, reasoning in the usual way we can establish that $T\langle\varphi\rangle$, $\neg T\langle\varphi\rangle$—and, for good measure,

[14] I suggested (Priest 1990), following Belnap, that the fact that the addition of a piece of logical machinery to a theory gives a non-conservative extension may show that it lacks determinate sense. The constant $\wp$—and, if the rest of Everett's argument is right, $\mathfrak{R}$—obviously fails this test. As Everett (1994, p. 418) notes, this argument presupposes that the theory contains the T-schema; but that's exactly what I *am* presupposing.

$F\langle\varphi\rangle$. This is not a problem: the aim was never to get rid of contradictions, but to accommodate them.

Timothy Smiley (Priest and Smiley 1993) has given an apparently more virulent extended paradox employing, not the notion of truth, but the notion of truth in an interpretation. He invites us to consider a fixed point, ψ, of the form $v_{@}(\psi) = \{f\}$, where $v_{@}$ is that evaluation whose assignments are in accord with the truth (*simpliciter*). Reasoning employing this, we conclude that $t = f$, and triviality rapidly ensues (Priest and Smiley, 1993, pp. 30f.).

There are a couple of possible responses to this argument (Priest and Smiley, 1993, p. 50). One is to modify one's semantics so that the set of non-empty subsets of $\{t,f\}$ does not exhaust the semantic values. Any member of the set of non-empty subsets of *this* set may also be a value, as may non-empty subsets of this, and so on. This allows the semantic value of ψ to be one of these hypercontradictions, and the argument to triviality is broken. A second response we will come to in a moment.

Everett (1993) argues that even given this semantic machinery, there is an extended paradox.[15] Given any set, s, let $\eta(s)$ be the set of its urelements. We consider the fixed point, θ, of the form $\eta(v_{@}(\theta)) = \{f\}$. Simple reasoning then yields $t = f$ as before. The argument, as does Smiley's original, requires us to suppose that we have the equivalence between $t \in v_{@}(\theta)$ and θ in a detachable form, which there is ground to doubt (Priest and Smiley, 1993, p. 52). However, with this reservation, it seems to me that the argument of Everett (1993) is correct here.

7. *Functionality*

A second response (given in Priest and Smiley 1993, p. 50f.) is to retain the original semantics, but to formulate semantic evaluations as *relations*, not as *functions*. As is shown there, this allows for the expression of everything that needs to be said, whilst invalidating Smiley's argument. There is an obvious repair of Smiley's argument, formulated with the help of additional set-theoretic machinery. This is also shown to fail. Everett (1993, fn. 6), gives a very similar argument, but using truth instead of truth-in-an-interpretation. We define a fixed point, θ, of the form $X_\theta = \{\langle F\rangle\}$, where $X_\theta = \{P; (P = \langle T\rangle \vee P = \langle F\rangle) \wedge T\langle P\langle\theta\rangle\rangle\}$. (Angle brackets again are used to form names, this time of predicates.) From

[15] Priest and Smiley (1993, p. 50) envisage the possibility of this. Everett's claim (1993, p. 6) that the paper says that it will be impossible, given the extended semantics, to construct a triviality-inducing extended paradox is just incorrect.

this it is supposed to follow that $\langle T\rangle = \langle F\rangle$. Triviality follows in a few steps.

The argument, as provided by Everett (in correspondence), depends on the premise that $X_\theta = \{\langle F\rangle\} \vee \langle T\rangle \in X_\theta$.[16] Now, I see no hope of proving this premise without using dialetheically invalid moves. Even given that we can establish that $\langle F\rangle \in X_\theta \vee \langle T\rangle \in X_\theta$, we need to show that when the first disjunct, but not the second, obtains, $X_\theta = \{\langle F\rangle\}$, i.e., $z \in X_\theta \Leftrightarrow z = \langle F\rangle$. And there are real problems about showing this. From right to left, for example, we may have that $(\langle F\rangle \in X_\theta \wedge z = \langle F\rangle) \Rightarrow z \in X_\theta$; but to infer that $\langle F\rangle \in X_\theta \Rightarrow (z = \langle F\rangle \Rightarrow z \in X_\theta)$, and so $(z = \langle F\rangle \Rightarrow z \in X_\theta)$, requires the invalid $(\alpha \wedge \beta) \Rightarrow \gamma \vdash \alpha \Rightarrow (\beta \Rightarrow \gamma)$. This fails in the semantics of Priest (1987, p. 112) even when α is a necessary truth.[17]

This objection depends on formulating set theory using intensional connectives; and it might be thought that the argument will work if we formulate it using extensional connectives, as usual. Now there may, indeed, be good reasons for supposing that set-theory ought to be formulated with extensional connectives (see Goodship, 1996). However, if we do this, then the abstraction schema of set theory takes the form:

$$x \in \{y\colon \varphi(y)\} \equiv \varphi(x)$$

(where $\equiv$ is the material biconditional, defined as usual: $\alpha \equiv \beta$ is $(\alpha \supset \beta) \wedge (\beta \supset \alpha)$, where $\alpha \supset \beta$ is $\neg\alpha \vee \beta$) and we can no longer detach from left to right, or vice versa, without using the invalid disjunctive syllogism. Inferences of this form are employed at several places in the above proof, and so it still fails.

[16] For the record, it is an argument by dilemma, and goes as follows. Suppose that $X_\theta = \{\langle F\rangle\}$, i.e., θ; then by the *T*-schema and a couple of simple moves, $\langle T\rangle$ satisfies the defining condition of X_θ, i.e., $\langle T\rangle \in X_\theta = \{\langle F\rangle\}$, and so $\langle T\rangle = \langle F\rangle$. Alternatively, suppose that $\langle T\rangle \in X_\theta$; then by the definition of X_θ and a couple of applications of the *T*-schema, θ, i.e., $X_\theta = \{\langle F\rangle\}$; so again $\langle T\rangle = \langle F\rangle$.

[17] With a slightly different account of the conditional, it can be shown that any argument of this kind must fail. Brady (1989) demonstrates that naive set theory with an underlying relevant logic is non-trivial, even when the abstraction schema holds in the very strong form:

$$\exists y \forall x (x \in y \Leftrightarrow \alpha)$$

where α is *any* condition, even one that contains y. For any closed α, we may define $\langle\alpha\rangle$ as $\{z;\alpha\}$, for some fixed variable, z, and Tx as $\phi \in x$. It follows straight away that for any closed α, $T(\alpha) \Leftrightarrow \alpha$. (See Priest 1990, fn. 9.). The theory therefore includes a theory of truth. Moreover, the strong abstraction schema gives us fixed points of the kind required for self-reference. Let $\alpha(x)$ be any formula of one free variable, x. By the schema, there is a set, s, such that:

$$\forall x (x \in s \Leftrightarrow \alpha(\{z;\ \phi \in s\})).$$

It follows that:

$$\phi \in s \Leftrightarrow \alpha(\{z;\ \phi \in s\})$$

whence there is a fixed point β such that $\beta \Leftrightarrow \alpha(\langle\beta\rangle)$. Hence, there is a non-trivial theory of sethood, truth and self-reference.

The second line of response to the original problem is therefore untouched by Everett's arguments. Before considering the rest of Everett (1993), let me digress for a moment. As Priest and Smiley (1993, p. 49f.) set up the situation, a contrast is drawn between extended paradoxes employing the notion of truth, *simpliciter*, which are anodyne, and extended paradoxes employing the notion of truth-in-an-interpretation, which need not be. This now seems to me to be an incorrect antithesis.

It is quite possible to set up triviality-inducing extended paradox arguments using the notion of truth, *simpliciter*.[18] Suppose that we are allowed to define the function, τ, defined on sentences, as follows:

$$\begin{aligned} \tau(\langle\alpha\rangle) &= 1 \text{ if } T\langle\alpha\rangle \\ &= 0 \text{ if } \neg T\langle\alpha\rangle \end{aligned}$$

Now consider the fixed point, λ, of the form $\neg T\langle\lambda\rangle$; we establish, as usual, that $T\langle\lambda\rangle$ and $\neg T\langle\lambda\rangle$, and hence that $\tau(\langle\lambda\rangle) = 1 = 0$. As formulated, the argument does not work, simply because the definition of τ is manifestly illegitimate. If both of the cases in the *definiens* may arise, as they may, dialetheically, the function is not well-defined. (See, further, Priest and Smiley 1993, p. 52.)

However, what all the above considerations do show is that it is not the distinction between truth conditions and truth-in-an-interpretation conditions that is pertinent to the possibility of generating triviality-inducing extended paradoxes. Rather, it is whether or not functionality, in its classical form, is present. Without it there are no problems; with it there are. Functionality must be handled very sensitively in a dialetheic context. This is an important lesson.

Anyway, and to return to Everett (1993), the rest of this argues that attempts to avoid the force of an extended paradox of the kind given can lead only to further problems. Since these arguments are repeated and generalised in the final paper (Everett 1996), let us move on to that.

8. *The fork*

Everett (1996) starts by rehearsing the arguments of the other two papers. He then presents his fork. As best I can formulate it, the argument goes as follows:

1. To provide a satisfactory solution to the Curry paradoxes, absorption must be shown to be invalid.

[18] This was pointed out to me by Uwe Peterson.

2. To show it to be invalid, an appropriate counter-model must be constructed.
3. Such a model must "guarantee" that contraction is invalid, i.e., "rule out" the possibility of its validity.
4. This can be done only if we have a way of expressing the claim that something is so which rules out the possibility of its not being so.
5. If we have such a way then this can be used to construct a triviality-inducing extended-paradox.

Let us take this argument step by step. (1) is quite correct. Any paradox solution that *merely* claims that some premise is false, or that some step is invalid, is cheap and worthless; reasons need to be given. It seems to me, however, arbitrary to talk just of Curry paradoxes here. As far as I understand it, the argument of Everett (1996) would work just as well if it targeted the Liar paradox and *ex contradictione quodlibet* instead. To this extent, the appearance of the argument as a fork is misleading.[19]

(2) need not detain us long either. There may be other ways of showing an inference to be invalid, e.g., by using a *reductio*; but a counter-model is by far the most satisfactory way.

(3) is where the problems start. A counter-model shows an inference to be invalid by showing that it is not truth-preserving in some interpretation. Now, if it is logically possible for contradictions to be true, then it is logically possible that the inference is *also* truth-preserving in that model, and maybe even valid. Why, however, is it a failing of the interpretation if it does not rule out this possibility? If it is a failing, then, since it hangs on a quite general point about claims not ruling out their negations, it applies to all arguments a dialetheist might give. But in any case, it hardly seems a failing. It is logically possible that I am now on the moon. This does not show that it is wrong for me to believe that I am not, that arguments that I am not are worthless, and so on. Of course, it would be different if this were more than a logical possibility; for example, if there were good evidence that I am now on the moon. Similarly, if it could be shown that absorption is valid, as well as invalid, then the dialetheist would indeed be in deep trouble. But once the argument for invalidity is accepted, the onus of proof here is on those who claim it to be valid.[20] If this were all there is to the argument, it would fold here; but it is not.

(4) This is the most sensitive part of the argument. The claim is that the dialetheist requires a way of expressing something that rules out its nega-

[19] When the argument is rehearsed more briefly in Everett 1993, it is, in fact, given in this more general form.

[20] See Priest 1987, §8.4 for a defence of the claim that consistency is a default assumption.

tion.[21] This follows from (3), which, as we have seen, is false. However, Everett (1996) gives two additional reasons in fn. 15.[22]

The first is that the semantics must be able to express the thought that a sentence is false in an interpretation and not also true in it, since this fact is expressible in English, an account of whose semantic concepts we are supposed to be giving. This argument just confuses being able to express something with being able to express it in such a way that it rules out its negation. What English can do, uncontentiously, is express the claim that something, α, is false and not also true (in an interpretation), in those very words. A formal language can do exactly the same: if v is an interpretation (conceived as a relationship between sentences and truth values) then the claim is expressed by the sentence $v(\alpha,f) \wedge \neg v(\alpha,t)$. It is not at all obvious that a natural language can express this fact in such a way as to exclude its negation. If dialetheism is correct, maybe it cannot. A speaker may utter those words with that intention, but intentions are not guaranteed realisation.[23]

The second argument Everett (1996) gives is that we need to be able to have this form of expression in order to mark the difference between a statement that is plain false, say α, and one which is both true and false, say β (in an interpretation, I take it). This argument also fails. To the extent that we need to express this distinction, it is expressed in the obvious way: β satisfies the predicates $v(x,t)$ and $v(x,f)$; α does not satisfy the first of these. If one argues that this does not adequately express the distinction, because the fact that α does not satisfy $v(x,t)$ does not rule out its satisfying it, then, as in the previous case, this just begs the question; we were supposed to be giving an argument as to *why* a mode of expression without this consequence is inadequate.

But let us grant that we need a way of ruling things out, anyway. Does a dialetheist have a way of doing this? At this point we need to ask what it means to rule something out. Arguably, several things might fall under this rubric. Classically, the negation of α rules it out in this sense: someone who asserts it cannot, without retraction, thereafter endorse α without becoming rationally committed to everything; which is, presumably,

[21] Similar claims have been made by Parsons (1990) and Batens (1990). In particular, Parsons also claims that a dialetheic solution to the paradoxes is in exactly the same situation concerning the inexpressibility of certain notions as are consistent solutions. These claims are discussed and rejected in Priest 1995.

[22] Everett (1996) tends to put the point in terms of truth, *simpliciter*, but since it is model-theoretic validity we are talking about, it should be truth in an interpretation.

[23] It might, of course, be argued that a sentence cannot express any claim unless it rules out something, and in particular, its negation. Such an argument would just be fallacious. See Priest forthcoming, §10.

something that no sane person would wish to do.[24] A dialetheist can rule out something in this sense. $\alpha \Rightarrow \bot$ has exactly this effect.

Another thing that "ruling out" is often taken to mean is saying something which expresses the mental attitude of rejection (refusal to accept). But a dialetheist can do this too. Assertion and denial are distinct illocutionary acts. Canonically, assertion expresses acceptance, and denial expresses rejection. Some (e.g., Frege) have thought that to assert a negated sentence is *ipso facto* to deny it. But there are certainly counter-examples to this. For a start, when a dialetheist asserts "the liar sentence is not true", they are certainly not expressing a rejection of the thought that it is also true, and so denying it. Conceivably, this example might be thought to beg the question, but there are plenty of others; for example, someone may come to realise that their beliefs, say about religion or politics, are inconsistent, by being made to assert an explicit contradiction. The assertion of $\neg\alpha$ in this context does not express a rejection of α. *Ex hypothesi*, they *do* accept α—at least until they revise their beliefs.[25]

The question of when an utterance expresses a denial is an interesting one, which I shall not pursue here. Certainly, an utterance of $\neg\alpha$ may express a denial of α, as, nearly always, does an utterance of $\alpha \Rightarrow \bot$. Even this need not do so, however. In the mouth of someone who believed everything, $\alpha \Rightarrow \bot$ would not express a denial—nothing would.[26] As a category, denial is *sui generis*, and cannot be reduced to a kind of assertion. And it is one whose employment is open to the dialetheist just as much as the non-dialetheist.

To summarise the discussion concerning (4), what we have seen is that the case given for the claim that we must have a way of expressing things that rules other things out does not stand up. Yet the dialetheist does have

[24] This is what Everett means by "ruling out", as may be inferred from Everett (1996, p. 14). He says there that a semantics with a simple truth predicate cannot express the claim that a sentence is not true (or takes an undesignated value) in such a way as to rule out its negation, whilst one employing model-theoretic apparatus can. There is a running together here of truth and truth-in-an-interpretation. (The language of designated values belongs to model theory.) But setting this aside, why does Everett think that this is so? Given a paraconsistent logic for the model-theoretic metatheory, $\nu(\alpha) = \{f\}$ no more makes it logically impossible for $t \in \nu(\alpha)$ to hold than $F\langle\alpha\rangle \wedge \neg T\langle\alpha\rangle$ makes it logically impossible for $T\langle\alpha\rangle$ to hold. But from $\nu(\alpha) = \{\mathrm{f}\}$ and $t \in \nu(\alpha)$ it does follow that $t = f$, and triviality ensues. Hence, the first expression does rule out the possibility that α is true (in ν) in the sense in question.

[25] For further discussion of the issue, see Priest and Smiley 1993, p. 36ff. and, especially, Priest forthcoming, §§9, 10.

[26] A different counter-example was provided by Everett himself in correspondence. Consider a Curry paradox, κ, of the form $T\langle\kappa\rangle \Rightarrow \bot$. This sentence ought to be rejected, but one cannot do so by uttering $\kappa \Rightarrow \bot$, since this commits the utterer to κ itself.

ways of ruling things out anyway, in whichever of the above senses this is supposed to be taken. So this part of the argument fails.

Let us turn, finally, to (5). The claim here is that if there is a way of ruling something out, this will give rise to triviality-inducing extended paradoxes. Is this true? We noted that $\alpha \Rightarrow \bot$ rules α out, in one sense. If we formulate an extended paradox with this notion, we obtain a fixed point, κ, of the form $T(\kappa) \Rightarrow \bot$. This is, of course, exactly Curry's paradox; and it poses no problem if absorption fails, which, according to me, it does.

The other way of ruling something out that we considered was denial, but denial is an illocutionary act, not a connective, and does not give rise to extended paradoxes in the way one might expect. About the closest one can get is a sentence like "I deny this sentence" (call this β) uttered as a performative. This would be the denial of something true, and therefore, presumably, something not to be uttered. If, on the other hand, one asserts β then it is false, and so not to be uttered either. The rational person, it would seem, neither asserts nor denies β.[27] Once one throws in a normative element, however, new paradoxes concerning assertion and denial are forthcoming. Consider, for example, the sentence "It is irrational to assert this sentence". Call this γ. Someone who asserts this is asserting something, and at the same time asserting that it is irrational to assert it. This is irrational. Hence, it is irrational to assert γ. But we have just established this (i.e., γ). Hence it *is* rational to assert it.[28] This is certainly not a triviality-generating paradox, however, and so not a problem for the dialetheist. It *is* a problem for those who espouse consistency. But I will leave others to worry about that. The important point is that the final step of the argument we have been considering fails.

9. Conclusion

I have now finished discussing the arguments of Everett (1996), and with it, all three of Everett's papers.[29] As I have argued, the central contentions of each fail. In particular, there is no connection of the kind envisaged between extended paradoxes and Curry paradoxes (though there is a connection of a different kind, as we saw in the last section). The discussion

[27] Various other near-paradoxes of this kind are discussed in Parsons 1984.

[28] See Priest 1995, p. 61.

[29] The final part Everett (1996) generalises his argument to other paradoxes, such as the heterological paradox and its currification. There seem to me to be some more moves in this which are rather too fast. But this part raises nothing essentially novel, so I will not discuss it.

has raised some issues that require further consideration, but that may be left for another occasion.[30]

Department of Philosophy
The University of Queensland
Brisbane, Queensland
4072 Australia

GRAHAM PRIEST

REFERENCEC

Batens, D. 1990: "Against Global Paraconsistency". *Studies in Soviet Thought*, 39, pp. 209–29.

Brady, R. 1989: "The Non-Triviality of Dialectical Set Theory", Ch. 16 of G. Priest, R. Routley and J. Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent*. Munich: Philosophia Verlag.

Denyer, N. 1989: "Dialetheism and Trivialisation". *Mind*, 98, pp. 259–68.

Dummett, M. 1977: *Elements of Intuitionism*. Oxford: Oxford University Press.

Everett, A. 1993: "A Note on Priest's Hypercontradictions". *Logique et Analyse*, 36, pp. 39-43.

——1994: "Absorbing Dialetheias". *Mind*, 103, pp. 414–9

——(1996 forthcoming): "A Dilemma for Priest's Dialetheism". *Australasian Journal of Philosophy*.

Fraenkel, A., Bar-Hillel, Y. and Levy, A. 1973: *Foundations of Set Theory*, 2nd edition. Amsterdam: North Holland.

Goodship, L. 1996: "On Dialethism". *Australasian Journal of Philosophy*, 74, pp. 153-61

Parsons, T. 1984: "Assertion, Denial and the Liar Paradox". *Journal of Philosophical Logic*, 13, pp. 137–52.

——1990: "True Contradictions". *Canadian Journal of Philosophy*, 20, pp. 335–53.

Priest, G. 1987: *In Contradiction*. The Hague: Nijhoff Martinus.

——1989: "Denyer's $ Not Backed by Sterling Arguments". *Mind*, 98, pp. 265–68.

——1990: "Boolean Negation and All That". *Journal of Philosophical Logic*, 19, pp. 201–15.

——1992: "What is a Non-Normal World?". *Logique et Analyse*, 35, pp. 291–302.

——1995: "Gaps and Gluts: Reply to Parsons". *Canadian Journal of Philosophy*, 25, pp. 57–66.

[30] I would like to thank Anthony Everett for helpful discussions on earlier drafts of this essay, which clarified a number of issues. I would also like to thank the editor of *Mind* for a very generous and helpful set of written comments.

——(forthcoming): "What Not?—A Defense of a Dialetheic Theory of Negation", in D. Gabbay and H. Wansing (eds.), *Negation*.
Priest, G. and Smiley, T. 1993: "Can Contradictions be True?". *Proceedings of the Aristotelian Society*, Supplementary Volume 65, pp. 17–54.