An Anti-Realist Critique of Dialetheism

by

Neil Tennant*

Department of Philosophy

The Ohio State University

Columbus, Ohio 43210

email tennant.9@osu.edu

May 20, 2004

Abstract

We criticize dialetheism from the point of view of an anti-realist with sympathy for relevantism in logical reasoning.

*Forthcoming in G. Priest, J. C. Beall and B. Armour-Garb, eds., *The Law of Non-Contradiction: New Philosophical Essays*, Oxford University Press.

We argue that the view that there are true contradictions suffers both from an improper understanding of the interrelations among absurdity, contrariety, falsity and negation, and from an incorrect diagnosis of what gives rise to the well-known contradictions in semantics and mathematical foundations.

Anti-realism emerges as a better reflective equilibrium than dialetheism in confrontation with all these phenomena. Both positions require logical revisions of classical logic. But anti-realism's logical reforms are better motivated than those of the dialetheist, and the resulting logic is more adequate for the methodological demands of mathematics and empirical science. Priest's prospect of an 'intuitionist dialetheism' is unconvincing, both because of important features of intuitionistic logic, such as the independence of the logical operators and the normalizability of proof, and because the intuitionist (or anti-realist) disagrees so strongly on the actual alleged examples of dialetheias in logic and foundations.

1 Introduction

Dummett once suggested that certain metaphysical theories about given discourses were 'mere pictures'; and that what was needed instead was a theory of meaning for, and an account of the correct logic underlying, the discourse in question. Dialetheism—the view that there are true contradictions, and that these contradictions are rationally believable—is a metaphysical theory not without its own arresting pictures. Consider:

> Truth and falsity come inextricably intermingled, like a constant boiling mixture. One cannot, therefore, accept all truths and reject all falsehoods ...

and

> ... natural language being what it is, we should not necessarily expect the pieces of language to fit together neatly, like some multi-dimensional jig-saw puzzle. There may well be mismatches. In particular, the conditions of application of a sentence may well overlap those of the application of its negation, especially if the world arranges itself in an unkind fashion. At such spots in the weft and warp of language we have dialetheias.

The dialetheic images do not come solely from the kitchen and the weaver's loft. They have an out-of-doors counterpart:

> ... the logical paradoxes are the site of a fault-line in the whole tectonic of "classical" logic. Though painfully aware of it, logicians this century have had as little success with it as their geological counterparts have had with the San Andreas fault. By applying a little pressure along the crack, I hope to blow the whole configuration asunder.

These quotes are taken from Priest (1987) (henceforth: *IC*)[1], at pp. 124, 85 and 7 respectively. The book combines flair for rhetoric with attention to formal detail, and is the best-known manifesto for dialetheism.

The suggestion that there could be true contradictions is the most radical that has ever been made within the field known generally as 'deviant logic'. A common reaction is out-of-hand dismissal. The suggestion, however, is worth looking at very closely—so that its eventual dismissal may be properly justified.

That there could be true contradictions is a view that requires yet another major deviation from the standard systems of classical or intuitionistic logic. Since the latter systems contain the absurdity rule (*ex falso quodlibet*):

$$\frac{\bot}{A}$$

they would allow the asserter of a true contradiction validly to infer any proposition whatsoever, and assert it. The dialetheist's suggestion, however, is not that it could be the case that every proposition were true; rather, it is just that there might be at least one true contradiction. He would not wish thereby to be committed to asserting every proposition whatsoever.[2] Therefore he must, and does, abandon *ex falso quodlibet*. He must, and does, believe that there can be interestingly different inconsistent theories. This is because, having asserted his chosen allegedly true contradiction $P \wedge \neg P$, he will still be able to derive absurdity from it, by means of conjunction- and negation-elimination:

$$\frac{\dfrac{P \wedge \neg P}{\neg P} \quad \dfrac{P \wedge \neg P}{P}}{\bot}$$

Contradictions are still inconsistent; they still imply absurdity.[3]

By abandoning or rejecting the absurdity rule, however, one avoids being committed, by asserting a contradiction, to asserting every proposition whatsoever. Thus one cannot perform a *modus tollens* from the ridiculousness of such an alleged commitment, to the ridiculousness of a contradiction's being true. Any relevantist (and I am one) who rejects the absurdity rule, but who finds dialetheism objectionable, must therefore rest their objection on something other than the ridiculousness of being committed to asserting every proposition. For—to repeat—by abandoning the absurdity rule, the dialetheist avoids the latter commitment.[4]

The dialetheist's suggestion is not that contradictions are consistent (that is, that they do not imply absurdity); rather, the suggestion is that some of them are *true*.

2 The interpretation of absurdity

The anti-realist who is also a relevantist might be thought to be somewhat under-equipped to undertake a really trenchant critique of dialetheism. After all, Priest himself said he thought that 'intuitionist dialetheism' may well bear further investigation (*IC*, p. 87). (We shall see below, however, that Priest's own formulation of the rules of inference for his Logic of Paradox do

not make the task of 'intuitionizing' that logic at all easy.) In this section we consider whether the dialetheist can expect to make an easy convert of the relevantist anti-realist.

As a relevantist, the latter will eschew the absurdity rule, and not acknowledge Lewis's first paradox $A, \neg A : B$ as a valid argument form. So, it might be thought, she will not be able to complain very convincingly against any dialetheist who chooses to assert both P and $\neg P$. By the relevantist's own lights, that contradiction could be 'localized'. It need not 'blow up' into completely promiscuous belief. Thus—so this line of thought continues—generally correct beliefs will still be able to lead to generally survival-enhancing actions, despite the occasional presence, in one's belief-system, of these localized dialetheias.

To this the anti-realist has a perhaps surprisingly heated response. She will say (or, I am arguing, ought to say) something along these lines:

> Look, Mr. Dialetheist. Contradictions are *such bad things* that all we need to do is *locate* them. This we do by means of the rule of $\neg$-Elimination, precipitating absurdity ($\perp$) as conclusion. Now when that happens, *that is terrible enough.* Having located them, we know that whatever gave rise to them is impossible for

a rational agent to believe. We don't *need* to 'convince' ourselves, by means of quite unnecessary applications of the absurdity rule, that anything whatsoever would follow from a contradiction. For $\perp$ is, logically, as horrific a conclusion as one can possibly get. Indeed, $\perp$ is *so* bad that, funnily enough, nothing *can* really follow from it the way the absurdity rule would otherwise maintain. $\perp$ is like a logical black hole: no possible thought that makes any sense could ever escape from it. So the absurdity rule is really silly, inappropriate and unnecessary. Thinking that one needs the absurdity rule in order to 'bring out' the terribleness of $\perp$ is about as naive as it would have been to remonstrate with Adolf Hitler over the murder of six milion souls by saying 'You really ought not to behave towards other people this way, you know, because that can set a bad example for others to follow.' Logically speaking, as soon as one encounters $\perp$, one ought to cry out 'Enough already!'

This, then, is the problem: how could a proposition that implies absurdity be true? The dialetheist's suggestion means that we would be either

(1) countenancing the occasional truth of the absurdity symbol

—or, should (1) not be the case—

(2) countenancing the occasional failure of application of rules of inference to transmit truth—strangely, in just those situations where an allegedly true contradiction has been asserted as a premiss for the ensuing rule-applications that now suffer this peculiar lapse from the truth-transmitting role that is their very *raison d'être*.

Ad (1): can the absurdity symbol occasionally denote the True? Surely not: absurdity (the constant $\bot$) is just that: a *constant*. It is *always false*.

The dialetheist, however, replies, 'Aha!—but is this enough to ensure that $\bot$ is *never true*? What if $\bot$ *can* (sometimes) be true?'

Well, yes, it ought to be enough to ensure this; provided only that one can make $\bot$ obey appropriate rules. The anti-dialetheist strategy ought to be to ensure that $\bot$ is a 'pure' constant. That is, it should not only always denote The False; it should also never denote The True.

Our manner of speaking here makes it seem as though we view the absurdity symbol as iconic: as standing for a truth-value (namely, The False, as opposed to The True). On this model one might conceive of the problem as being that of how to ensure that the symbol denotes its 'essential referent', and nothing else. On an analogy employing proper names of persons, the

problem would be like that of ensuring that 'Julius Caesar' denotes Julius Caesar, and no one else (for example: not Brutus). The problem is *not* like that of ensuring that 'Julius Caesar' denotes Julius Caesar, and not, for example, any of Brutus's murder victims. That the term 'Brutus's most famous murder victim' happens to denote Julius Caesar is no problem for the referential essentialist. For Julius Caesar need not have been Brutus's most famous murder victim. Baptisms do not ordain the future. They do, however, *single out* bearers for the name employed, and in such a way that it would be metaphysically impossible for a distinct individual to be denoted correctly by that name.

So our problem can best be understood as that of securing the *rigidity* of $\perp$ when it is construed as a name-like constant symbol, standing for The False.

There is another construal of $\perp$ as a constant, on which it is not name-like. On this alternative construal, $\perp$ signifies the joint impossibility of some hypothesized propositions $P_1, \ldots, P_n$ whose identity can be gathered from the context in which $\perp$ makes its inferred appearance. Thus $\perp$ is taken to have *scope*, and be *modal* in character—a combination of features that we can summarize handily in the phrase '*scopily modal*':

$$\begin{array}{c} \underbrace{P_1, \ldots, P_n} \\ \vdots \\ \bot \end{array}$$

One who has reasoned thus from the assumptions $P_1 \ldots, P_n$ takes herself to be in a position to say *it is not possible for all of* $P_1, \ldots, P_n$ *to be true.* Or: *it cannot be the case that* $P_1 \ldots, P_n$; or: *it is never the case that* $P_1, \ldots, P_n$. In all these synonymous expressions one sees the combination of *modality* ('not possible'; 'cannot be'; 'never') and *scope* $(P_1, \ldots, P_n)$.

The most obvious application of the foregoing is to a proposition and its negation, captured by a single step of negation-elimination:

$$\begin{array}{c} \underline{\neg P \quad P} \\ \bot \end{array}$$

Here we want to say *it is impossible for both* P *and* $\neg P$ *to be true*; *it cannot be the case that both* P *and* $\neg P$; *it is never be the case that both* P *and* $\neg P$. And yet the dialetheist is suggesting that indeed it *is* possible for both P and $\neg P$ to be true; it *could indeed* be the case that both P and $\neg P$; it *is sometimes* the case that both P and $\neg P$. Naturally, he is not holding out such a prospect for arbitrary propositions P : his suggestion is only

that there are certain propositions P that behave in this remarkable way. That behaviour arises from the subject-matter of P and the logico-linguistic construction of P.

The dialetheist will further concede that this entails (for the kinds of proposition P in question) that P could be both true and false. (For how, otherwise, could $(P \wedge \neg P)$ be true?) But note the converse, and the price it extracts. As soon as one contemplates the possibility of, say, a contingent proposition P being both true and false, the destruction of polarity becomes severe, in that the a priori, analytic falsehood $(P \wedge \neg P)$ must now (for the particular choice of P in question) be held out as also *true*. Letting in any degree of overlap between truth and falsity puts one's semantic lenses right out of focus. The falsest of false claims, namely claims of the form $(P \wedge \neg P)$, can now, outrageously, also be true.

Ad (2): can our rules of inference occasionally fail to transmit truth? Again, surely not. Logical laws, as laws of rational thought, admit of absolutely no exceptions. Valid rules of inference *always* transmit truth. (More precisely: if the subproofs for the application of a rule are truth-preserving, then the proof formed by the application of the rule is truth-preserving. The preservation in question is from the undischarged assumptions of the proof to its conclusion.) Laws apply on the basis of form alone—more narrowly, on the

basis of *one logical operator-occurrence at a time*—namely, the occurrence of whatever operator is *dominant* in the major premiss (for an elimination) or the conclusion (of an introduction). Since no would-be dialetheia $P \wedge \neg P$ has any way of advertising itself as such to the rule of $\wedge$-Elimination, the application of that rule will be as truth-preserving as ever. So too will be the subsequent application of $\neg$-Elimination to derive $\bot$. That is to say, the formal proof of $\bot$ from $P \wedge \neg P$ shows that *it is impossible for* $P \wedge \neg P$ *to be true*.

We have just employed the following argumentative strategy: we claimed that the dialetheist's suggestion implies a certain disjunction; then we explored each disjunct in an attempt to reduce it to absurdity. Hence the dialetheist's suggestion is reduced to absurdity. In normal polite circles this would be enough to silence the opposition. But what happens in this particular debate?—the dialetheist merely smiles knowingly and avers that his position is indeed inconsistent—but still true. Nevertheless, he owes us a definite answer: which of (1) and (2) above is he committed to holding?[5] Since Priest himself does not use or countenance the absurdity symbol, he has never raised or answered this question. But for the anti-realist the question is pressing, and the answer to it would be most revealing.

3 The dialectical grip of dialetheism

The dialetheist's suggestion is even more disabling than the contention that certain claims of the logical form $P \wedge \neg P$ can be true. On the assumption that there could be true contradictions 'all the way up', one finds oneself in a rather invidious position as the opponent of the dialetheist. For one would like to be able to bring certain canons of counterargument to bear. One of the most important of these is the method of *reductio ad absurdum*. Providing a *reductio* of an opponent's position (expressed by a set of propositions) usually clinches the case against him. But now we have the dialetheist maintaining the possible truth of contradictions. Confronted with a *reductio* of his philosophical position, he is able simply to smile and say that his very own philosophical position is both true and false—thereby vindicating itself, thanks to the successful *reductio* supplied by the opposition!

If we want to preserve the orthodox understanding of absurdity against the dialetheist's insidious suggestion, and do so on the construal of $\perp$ as scopily modal—that is, as showing the 'noncompossibility' of what precipitates it—then we have to develop an account that makes *reductio ad absurdum* a dialectically effective tool in argument. By 'dialectically effective' here we mean that it will provide a compelling reason to *abandon* any the-

oretical viewpoint that is committed to the propositions assumed for the *reductio*.

One is tempted to ask, as Popper did for empirical theories, what could possibly count as a falsification of the dialetheist's theory. But now we are treating of a philosophical theory, not an empirical one. So it is not enough to ask this question with the ordinary notion of falsehood implicit in the term 'falsification'. For one wants to know what sort of falsification would be regarded by the dialetheist as one that *ought to persuade him to abandon his theory*, rather than as one that could be thrown back in his opponent's face as a superb piece of supporting evidence for that very theory.

4 Are absurdity and negation really coeval?

Priest writes (*IC*, §4.7, p. 81) that 'Negation is that sentential function which turns a true sentence into a false one, and vice versa.' And he defines falsity (as a metalinguistic predicate F of object-language sentences α) as follows:

$$F\underline{\alpha} \leftrightarrow T\underline{\neg\alpha}.$$

There is no need, however, to use the truth-predicate T in this definition. One could just as well give the definition

$$F\underline{\alpha} \leftrightarrow \neg\alpha.$$

This would better justify Priest's claim that 'It would seem that falsity and negation can be defined in terms of each other, but neither can be defined without the other.'

Priest is thinking, however, only of explicit definitions, in which the central biconditional can be read as 'means by definition', and can indeed be read in either direction, depending on which concept (negation or falsity) one takes oneself (or one's interlocutor) to have grasped first.

But there is another way to attain the concept of negation, which I have described at length in a paper referred to earlier.[6] On this account the first logical notion (related, but not equivalent, to negation) to make its appearance is absurdity, as the conclusion of *antonymic inferences*:

$$(\dagger) \qquad \frac{A \quad B}{\bot}$$

Here the antonyms A and B are so simple and primitive that there cannot be any question of their 'dialetheically' holding simultaneously. Such antonyms A and B are antonymic not on the basis of their logical form, but on the basis of their primitive non-logical contents. The tension between them—their mutual exclusivity—is a matter of deep metaphysical necessity. Some antonym-pairs derive from the structure of our phenomenology. An example

would be color incompatibilities:

$$\frac{\text{(Solidly) Red} \quad \text{(Solidly) Green}}{\bot}$$

where the feature-placings are understood to concern the same area of one's visual field at the same time.

Other antonym-pairs reflect fundamental features of our categorization of physical objects in space and time:

$$\frac{\text{X is right here, right now} \quad \text{X is way over there, right now}}{\bot}$$

where X is a physical object too small to straddle both the spatial locations indicated.

I shall not bore the reader with further examples of this kind. Natural language teems with them—a usefully *ad hominem* consideration, since Priest himself continually stresses the need to account for natural language with all its expressive power, and not just expressively limited formal languages. Moreover, the contents of the various As and Bs that can feature in such A-B clashes of the form (†) can be learned and grasped without adverting to the other, and especially without use of the negation particle. I contend that an intelligent child can learn the meaning of 'hot' (as applied

to water in a basin, say) without needing to be told that it is 'not cold'—and vice versa. Once both 'hot' and 'cold' have been mastered, however, the child would immediately be apprised of the transition

$$\frac{\text{Hot} \quad \text{Cold}}{\bot}$$

where $\bot$ is 'scopily modal' in the sense explained earlier. Reflecting on the nature of one's sensory experience, and the kinds of sensations involved, can make one aware of these incompatibilities as *necessary*. (If you do not like the 'hot'/'cold' example, never mind; it suffices for my argument if there is *just one* scopily modal antonym-pair that can be grasped without recourse to any symbol for propositional or sentential negation.)

The conception of contrariety is expressed by means of an inferential transition *from* the contraries in question *to* absurdity. Because of the structure of both our sensibility and our understanding, these contrarieties are deep, primitive and necessary. They are exceptionless. They are a priori. It matters not that one needs sensory experience in order first to *acquire* a grasp of each member of an antonym-pair. A prioricity has nothing to do with acquisition, and everything to do with mastery. What matters is only that once the members of an antonym-pair have been grasped, their contrariety is something that their understander can grasp without the need

for any further sensory experience.

Once one has the notion of contrariety (via the notion of absurdity), one can proceed to introduce the concept of negation. This is done by means of the introduction rule, which does not furnish an explicit definition, but rather fixes the sense of the negation sign within an inferential context:

$$\begin{array}{c} \\ A \\ \vdots \\ \underline{\perp} \\ \neg A \end{array}$$

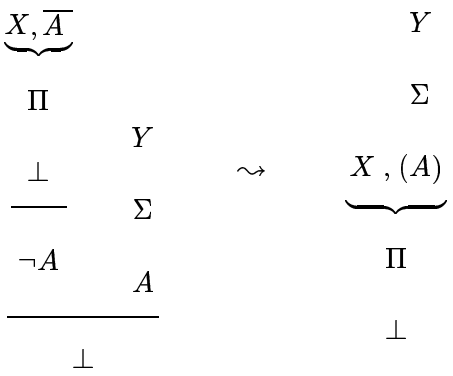
It is of course understood that the side-assumptions, other than A, on which $\perp$ depends will remain undischarged, so that the newly inferred conclusion of the form $\neg A$ rests on them.

Now, harmoniously balancing this introduction rule is the *elimination* rule, which can be 'read off' the introduction rule:

$$\begin{array}{c} \underline{\neg A \quad A} \\ \perp \end{array}$$

The reason why I say that the elimination rule can be 'read off' the intro-

duction rule is that they obey the following reduction procedure:



This is why the introduction rule (which tells one how to introduce a dominant occurrence of the negation sign in an inferred conclusion) serves to introduce the very concept of negation. There is no explicit definition here; there is only an explicit recipe for the *use* of the negation sign in inference.

In the statement of both the introduction rule and the elimination rule, the notion of absurdity is stable. It is always invested with its full import of *necessary falsehood*. Remember, this derives directly from its use to register *deep and primitive* contrarieties. There is no question—the possibility simply cannot arise—of $\perp$, taken as a sentential constant, ever being *true*. And *that* is why negation works in such a way that it could never be the case that both P and $\neg P$ were true.

5 Alleged examples of true contradictions

Those who still take consistency as a regulative ideal—because it is necessary, even if not sufficient, for truth—will wish to examine closely the arguments given by Priest in Part I of *IC*. These are arguments purporting to deliver actual examples of honest-to-goodness true contradictions. They are the familiar semantic paradoxes, the familiar set-theoretic paradoxes, and an unfamiliar one thought up by Priest: the Gödel-sentence for naive provability. Priest's strategy is to take on all-comers who seek to defuse or explain away the familiar paradoxes, and to insist that the only natural way to respond to his examples is simply to acknowledge them for what they appear to be: namely, true statements of the form $P \wedge \neg P$. (Because his example of the alleged Gödel-sentence for naive provability is newly invented, there are no well-known counterarguments, against his use of this example, for Priest to consider; he does, however, as we shall argue below, fail to anticipate an important one.)

When I was a graduate student, I discovered a disconnected, but now relevant, phenomenon. While expecting our first child, I suddenly became aware of pregnant women all over the place.[7] Had that first pregnancy been a phantom one, I might nevertheless, as a heady expectant father, have

falsely *imagined* that various obese women on the streets were pregnant. The analogy I am seeking here is that the would-be dialetheist, on 'discovering' what he believes is his first true contradiction, suddenly takes himself to be confronted by them everywhere. But we must not allow ourselves to be overly impressed by the sheer weight of numbers, and variety, of these would-be true contradictions. Each candidate has to be examined rigorously, to see whether it really passes muster. I should add here that I am not at all impressed or consoled by Priest's impish claims that 'contradictions should not be multiplied beyond necessity' (p. 90—such Ockhamite restraint ... as though dialetheias form just another natural kind, another handy class of 'posits'!) and 'contradictions are *a priori* improbable' (p. 132—but watch out, you never know when one might hit you!).

When one looks closely at Priest's argumentative strategies with his various examples, a common pattern emerges, which is interestingly related to the vexing one noted above (namely, co-opting any *reductio ad absurdum* of dialetheism as a vindication of dialetheism). In each case—be it a semantic paradox, or a set-theoretic one—Priest's method is to show how a contradiction apparently ensues from certain plausible-seeming principles taken together. He then insists on their plausibility, and on the validity of every step of the reductio, and concludes that the reductio must, after all, estab-

lish a true contradiction. What is not revealed to the reader, and what the reader has to uncover for herself, is the fact that there are *other*, much less plausible principles being tacitly employed in generating the contradiction in question. Priest's reductio arguments, that is to say, are enthymematic. And once the hidden premisses are made explicit, it is easy to challenge them, and defuse the paradoxes in question.

Doing this systematically, across the range of examples that Priest adduces, is an admittedly time-consuming and tendentious task. But it is a task that has to be carried out to an exhaustive conclusion, in any proper test of the merits of dialetheism as the existential claim that *there are* true contradictions. We shall be examining below, in §9, what we take to be the disputable hidden but crucial premisses in Priest's main examples of purported dialetheias in the area of logic and the foundations of mathematics. But first we turn to some context-setting remarks about reflective equilibrium.

6 Reflective equilibrium

Seldom is there any knock-down argument in philosophy for any global viewpoint. Those committed to particular viewpoints would of course be de-

lighted to have definitive reductios of all competing viewpoints; but hardly ever are these available. Every reductio is, after all, relative to certain side-assumptions, and the dispute can always be shifted so as to focus on these. Seldom is a well-represented philosophical viewpoint genuinely reduced to absurdity in a 'herme(neu)tically sealed' fashion, using only its own explicitly acknowledged principles as premisses for the reductio, and without allowing any contentious side-assumptions to seep in.

Global viewpoints such as realism or anti-realism almost always appeal to their proponents as *reflective equilibria*, chosen because of the particular way that the viewpoint in question distributes its emphases, shapes its main concepts, resolves known tensions among competing claims (so-called aporias), organizes widely shared intuitions, accommodates common-sense knowledge and scientific principles, etc. On occasion a chosen reflective equilibrium can involve reforming certain intuitions, re-shaping central concepts, and abandoning certain fundamental-looking principles, because of significant off-setting gains in clarity, systematicity, elegance, economy, and so on.

Speaking entirely as an interested party, I would offer the attainment of the reflective equilibrium of anti-realism as a striking example of this process. The realist and the anti-realist both have to deal with an important

tension among certain central principles: Bivalence of Truth, Knowability of Truth, and Epistemic Modesty.[8] The realist holds to bivalence, and rejects knowability; whereas the anti-realist insists on knowability, and gives up bivalence. In the anti-realist's reflective equilibrium, the classical, pre-theoretic concept of truth is re-shaped in so far as all truths are held to be knowable-in-principle; the principle of bivalence is abandoned; and the resulting methodological 'loss' is offset by the gains of systematicity and elegance to be had from intuitionistic logic and a nicely 'separable' inferential meaning-theory for the logical operators.[9]

Priest himself reveals quite clearly how his own viewpoint of dialetheism is, for him, the outcome of his own search for reflective equilibrium in dealing with phenomena such as semantic closure, the set-theoretic paradoxes, the Gödel-phenomena, the metaphysics of change, and the logic of norms. He writes (*IC*, p. 112)

> Intuition may provide an important part of the data against which a logical theory is measured. But a theory which is strong and satisfactory in other respects can itself show the data to be wrong.

He also emphasizes what he regards as the 'simplicity and philosophical

perspicuity' of his semantics for entailment, and regards these features as outweighing various objections to that semantics based on precise technical considerations (*IC*, p. 114). He stresses the importance of both direct arguments and indirect ones, to the effect that extant rival views are inadequate (p. 126). He attributes to his naysayers the vices of narrow-mindedness and dogmatism for rejecting inconsistency 'thoughtlessly and out of hand' (p. 132). The totality of considerations leading one to rest at a certain reflective equilibrium need not, for Priest, be overwhelmingly compelling or apodeictic. After surveying a variety of arguments attempting to show 'directly that contradictions can be rationally believed' (p. 124), he sums up by saying (p. 126)

> Perhaps no *single* argument from this collection may suffice to make naive set theory and semantics acceptable in preference to their consistent rivals; but it seems to me that the combined array is quite sufficient to make the inconsistent theories rationally preferable.

The only way, it seems to me, to counter this overly swift adoption of dialetheism as an intellectual resting point is to re-visit the various topics one by one, and to show much more aggressively, from the 'consistentist'

point of view, how deeply mistaken the dialetheist's account of each one really is. One has to undermine the pragmatic complacency of the theorist who thinks that, having reached a non-definitive 'standoff', or position of stalemate, with the consistentist on each of the disputed topics, some sort of global summation-on-balance can tip the reflective scale in favor of dialetheism. I submit that this is a grave misrepresentation of the true state of affairs, argumentatively. I contend that there are knockdown arguments *against* the dialetheist's accommodation of the various problematic topics, arguments that Priest, as the main proponent of dialetheism, has simply not addressed.

In addition to these several victories that, I claim, the consistentist can score, I maintain that there are some methodological considerations that have not yet been taken explicitly into account by the dialetheist, and which—when properly investigated—provide strong suasive currents that should bear the thinker decisively *away* from dialetheism as her final stopping point. The methodological considerations that I have in mind here are the following:

(1_D) Are the logical reforms of the dialetheist able to accommodate the great bulk of ordinary reasoning in mathematics and the natural sci-

ences?

(2_D) Does the dialetheist's 'logic of paradox' have a satisfying proof theory?

(3_D) Does it do justice to logic as a science of *inference* (as opposed to providing a fancy deviant model of semantic evaluations)?

(4_D) Does it accommodate the very arguments that the dialetheist uses when trying to show that certain propositions are indeed dialetheias?

Question (4_D) is highly non-trivial. An affirmative answer is required in the interests of the reflexive stability of the whole enterprise. But, as will emerge in due course, an affirmative answer is lacking.

The corresponding questions, concerning intuitionistic logic, dominate the anti-realist's case for intuitionistic logic (or something close to it) as the correct logic. For compare:

(1_A) Are the logical reforms of the anti-realist able to accommodate the great bulk of ordinary reasoning in mathematics and the natural sciences?

(2_A) Does the anti-realist's 'logic of warranted assertability' have a satisfying proof theory?

(3_A) Does it do justice to logic as a science of *inference*?

(4_A) Does it accommodate the very arguments that the anti-realist uses when trying to show that there is no reason to believe, on the basis of logic and the theory of meaning alone, that all propositions are determinately true or false, independently of our means of coming to know what their truth-values are?

When arguing for anti-realism, and arguing for intuitionistic relevant logic as the correct logic (given the anti-realist's theory of the meanings of the logical operators), I was concerned to justify affirmative answers to questions (1_A)–(4_A).[10] I believe it is important that the dialetheist should be able to do likewise with questions (1_D)–(4_D).

We shall examine below the extent to which Priest may be said to have developed his (classical) logic of paradox as a clear *inferential* alternative to classical logic.

7 Priest's Logic of Paradox

The only concession that Priest appears to make to the foregoing concern for a thorough and proper treatment of inference is a passing comment on p. 93 of *IC*:

> ... a good way of conceiving formal languages *and their seman-*

> *tics* is as a model for, or abstraction of, certain aspects of natural language: specifically, those aspects which are central to (deductive) inference. [Emphasis added.]

Priest does not, however, get directly to grips with deduction itself, by providing a detailed proof-theory for his logic of paradox. In *IC*, he provides a dialetheic semantics for the extensional connectives $\neg$, $\vee$ and $\wedge$, and simply remarks 'It is ... straightforward to produce a natural deduction system with respect to which these semantics are sound and complete. The details of this need not concern us here.' (p. 95)

But the details *should* concern the reader at this point, and indeed throughout the whole investigation. For, if one is being told to restrict one's logical principles, then it had better be the case that the arguments for doing so can themselves be framed wholly by means of the logical principles that survive the recommended restriction. Now we do not ordinarily reason in natural language, or any of its regimentations, by calculating algebraic assignments of (sets of) truth-values to the various propositions involved in our trains of deductive reasoning, and then checking the resulting distributions of values across premisses and conclusions. Rather, we make assumptions, sometimes only hypothetically, and draw conclusions from them, based on

their logical forms alone. *That* is what 'deductive inference' is all about. So we had better be shown, very clearly, exactly what forms of *inference* remain licit—as far as the dialetheist is concerned—after the 'whole configuration' of classical logic has been 'blow[n] ... asunder'.

This section is devoted to revealing reasons for dissatisfaction with Priest's logic of paradox as a canon of inference that can be incorporated into a reasonable reflective equilibrium of the kind described above. Readers who do not wish to engage in the ensuing logical details can skip this section, provided only that they bear in mind that the gist of the technical discussion is that Priscilla is pouty about the logic of paradox.[11]

In a footnote on p. 95 of *IC*, Priest tells his reader that the details 'can be found, in effect, in Priest (1982)'. If we turn to that paper, we find the same dialetheic semantics for the three connectives just mentioned, and the following 'natural deduction system in the style of Prawitz'. Here I take pains to state every rule in a single direction, rather than use Priest's abbreviations of two-way rules. (I also take the liberty to correct the logical operator in Priest's rule (2), which I am sure was meant to be the rule of $\wedge$-Introduction, rather than a strange two-premiss rule of $\vee$-Introduction.)

Introduction and Elimination Rules for $\wedge$ and $\vee$:

$$\frac{A \quad B}{A \wedge B} \qquad \frac{A \wedge B}{A} \qquad \frac{A \wedge B}{B}$$

$$\frac{A}{A \vee B} \qquad \frac{B}{A \vee B} \qquad \frac{\begin{array}{ccc} & \overline{A} & \overline{B} \\ A \vee B & \vdots & \vdots \\ & C & C \end{array}}{C}$$

De Morgan Inferences:

$$\frac{\neg(A \vee B)}{\neg A \wedge \neg B} \qquad \frac{\neg A \wedge \neg B}{\neg(A \vee B)} \qquad \frac{\neg A \vee \neg B}{\neg(A \wedge B)} \qquad \frac{\neg(A \wedge B)}{\neg A \vee \neg B}$$

Double Negation Introduction:

$$\frac{A}{\neg\neg A}$$

Law of Excluded Middle:

$$\frac{}{A \vee \neg A}$$

Priest says of this 'proof theory' that it 'could hardly be said to be simple or natural'. What we have here, however, is a proposed systems of rules for the construction of natural deductions; we do not as yet have any *proof theory*. For a proof theory furnishes such results as normalization theorems, based on reduction procedures; and examines the relationship between the chosen system and other well-known systems such as classical or intuitionistic logic. Such relationships might involve translations that preserve deducibilities, such as the double-negation translation of classical into intuitionistic logic. But we have none of these things from Priest's treatment. All we get is the set of rules of inference just stated, and a sketch of a (soundness and) completeness proof with respect to the dialetheic semantics. The soundness proof is said to be 'straightforward and left to the reader'. The completeness proof for the propositional logic we are interested in actually omits the crucial details for the connectives, concentrating instead on the interesting tense operators that dominate the discussion of the paper in question.

7.1 Priest's rules are not separable

The first thing to note about Priest's system of rules is that it characterizes $\neg$ only via its connection with $\wedge$, $\vee$ and itself. While $\wedge$ and $\vee$ have their own introduction and elimination rules that single them out for specific treatment, $\neg$ has no such proprietorial rules of its own. Instead, we have (1) a rule of Double Negation *Introduction*, which is quite unusual in a natural deduction setting; (2) four De Morgan inferences, the last of which is classically but not intuitionistically valid; and (3) the Law of Excluded Middle, which again is classically but not intuitionistically valid. The great theoretical advantage of 'Prawitz style' natural deduction is that it shows how minimal logic is nested within intuitionistic logic, which in turn is nested within classical logic. There is no similar way, with Priest's Logic of Paradox as currently presented, to identify its intuitionistic fragment by simply dropping certain strictly classical rules. For it would appear that there is only one way to prove the Law of Non-Contradiction $\neg(P \wedge \neg P)$ in this system, and it begins by helping oneself to the Law of Excluded Middle:

$$\frac{}{A \vee \neg A}$$

$$\frac{A \vee \neg A}{\neg\neg(A \vee \neg A)}$$

$$\frac{\neg\neg(A \vee \neg A)}{\neg(\neg A \wedge \neg\neg A)}$$

$$\neg(A \wedge \neg A)$$

The third line is derived by a De Morgan interdeducibility *and* a tacit use of the substitution rule (which Priest does not state) whereby interdeducibles are intersubstitutable (here, within the scope of the dominant negation). The fourth line also requires this substution rule, the interdeducibles in question being A and $\neg\neg A$.

But wait a minute!—are the sentences A and $\neg\neg A$ really interdeducible, even in the supposedly classical system that Priest has furnished? 'Well of course they must be!' the reader might retort, 'Has not Priest shown that his system is complete for ordinary classical deducibilities?' The answer, unfortunately, is negative. The completeness proof involves a crucial unproved assertion, Lemma 1 (v) on p. 258:

$$A \in \Delta \ \textit{iff} \ \neg\neg A \in \Delta,$$

for any prime and deductively closed set Δ. Primeness is irrelevant here; it is the assumed deductive closure of Δ that must be doing the job. Now,

while the left-right direction of the displayed claim is immediate by Priest's unusual rule of Double Negation Introduction, he clearly seems to have assumed, incorrectly, that the right-left direction is true. I cannot see any way to derive A from $\neg\neg A$ by means of Priest's rules above for his (classical) Logic of Paradox. Indeed, inspection of the rules reveals that there cannot be such a derivation. Any such derivation would have to end with an application of $\vee$-Elimination. The only available major premiss would have to be an instance of the Law of Excluded Middle. Of course, in ordinary intuitionistic logic there is just such a derivation:

$$\dfrac{A \vee \neg A \qquad \dfrac{\dfrac{}{A}(1) \qquad \dfrac{\dfrac{}{\neg A}(1) \qquad \neg\neg A}{\bot}}{A}}{A}(1)$$

The problem, however, is that this derivation involves an application of *ex falso quodlibet*, which Priest does not allow. Moreover, the derivation to the same effect in the system *IR* eschews *ex falso quodlibet*, but uses a liberalized form of $\vee$-Elimination, according to which if one case ends with

absurdity, then one can bring down the conclusion of the other case as the main conclusion:

$$\dfrac{A \vee \neg A \qquad \dfrac{\dfrac{\dfrac{}{\neg A}(1) \qquad \neg\neg A}{\bot}}{A}(1)}{A}$$

This liberalized $\vee$-Elimination rule has no place, however, in Priest's system; for it yields a proof of disjunctive syllogism, which is not valid in that system.

It would appear, then, that Priest must have intended his stated rule of Double Negation Introduction to be a two-way rule, like his two-way De Morgan rules. In other words, he must have been intending to help himself to the rule of Double Negation Elimination as primitive, in addition to Double Negation Introduction.

7.2 Implication is defined, not primitive

The second thing to note about Priest's inferential system is that the conditional connective $\rightarrow$ is defined as follows:

$$A \rightarrow B \quad =_{df} \quad \neg A \vee B.$$

Now while this is fine for classical logic, it will not do for intuitionistic logic. As is well known, every one of the connectives $\neg$, $\wedge$, $\vee$, and $\rightarrow$ is independent in intuitionistic logic. Without a primitive conditional, the propositional part of intuitionistic logic is expressively incomplete. Hence this is a further consideration against Priest's choice of logical primitives and of inference rules governing them. The isolation of an intuitionistic fragment of the Logic of Paradox would appear to be even more remote because of it.

7.3 Quantifier rules are absent

The third thing to note about Priest's inferential system if that, though he provides a dialetheic semantics for first-order logic (*IC*, §5.3, pp. 96–8), he never states any rules of inference for the quantifiers $\exists$ and $\forall$. Thus we are unable to assess whether his logic can indeed furnish regimentations of all the arguments that he himself puts forward in order to convince his reader that certain propositions are indeed dialetheias. For these arguments are conducted at least within a first-order fragment of English, and would definitely need the two quantifiers for their proper regimentation.

The most educated guess as to the appropriate form that quantifier rules might take in Priest's inferential system is that the existential quantifier should be treated analogously to disjunction, and the universal quantifier

analogously to conjunction. Thus they would have their usual introduction and elimination rules, as supplied by Gentzen and Prawitz. In addition, there would be the quantificational analogues of the De Morgan rules:

$$\frac{\neg\exists x F}{\forall x \neg F} \qquad \frac{\forall x \neg F}{\neg\exists x F} \qquad \frac{\exists x \neg F}{\neg\forall x F} \qquad \frac{\neg\forall x F}{\exists x \neg F x}$$

7.4 The deduction theorem fails

The fourth thing to note about Priest's inferential system is that the deduction theorem fails. For, by his Fact 1 on p. 95 of *IC*, every two-valued logical truth is a logical truth in the dialetheic sense. Hence the sentence

$$(A \wedge \neg A) \rightarrow B$$

is a dialetheic logical truth. But, by his Fact 3, we do not in general have that B is a dialetheic logical consequence of $A \wedge \neg A$. Since any correct deductive system would have to match its deducibility relation $\vdash$ to the semantic consequence relation, this means

$$A \wedge \neg A \nvdash B.$$

Hence the deduction theorem fails in the direction

$$X \vdash A \rightarrow B \;\Rightarrow\; X, A \vdash B.$$

For my own part, I do not think that this is of any great concern, since exactly the same thing occurs with the system IR of intuitionistic relevant logic. But in the case of IR, a story can be told of how such 'failure' of the deduction theorem is offset by an *epistemic gain*. (In IR, what is involved is essentially a failure of unrestricted transivity of deduction, since in IR—unlike Priest's Logic of Paradox—one has $A, A \to B \vdash B$.) The epistemic gain in question is that when one is deprived of a certain 'result' $Y : B$ that one would have expected, given transitivity of deduction, one can prove, instead, a *strengthening* of that result—one of the form $Z : \bot$ or $Z : B$, for some subset Z of Y. An immediate corollary is that, if intuitionistic mathematics is consistent, then every intuitionistic mathematical theorem can be proved from the mathematical axioms using only IR. (Likewise for the classical case.) There is, so far as I know, no such metatheorem concerning epistemic gain for Priest's system, inferential or semantic.

There is, however, the following result, showing how closely dialetheic logic can mimic classical logic when one's theory is consistent. (See IC, §8.6, p. 149. We are about to take a closer look at the left-right direction of Priest's Theorem 0.) In order to state this result, we introduce some notation. We shall assume that Priest's dialetheic logic obeys both a dialetheic soundness theorem and a strong completeness theorem with respect

to dialetheic semantical consequence. Thus we shall speak indifferently of dialetheic deducibility and/or consequence, and symbolize this by $\vdash_D$. Let $\vdash_C$ likewise be classical deducibility (which of course is the same as classical consequence). Let Δ be a set of sentences, and let φ be a single sentence.

Theorem. If $\Delta \vdash_C \varphi$, then for some conjunction ψ of members of Δ, we have $\psi \vdash_D \varphi \vee (\psi \wedge \neg\psi)$.

Proof. Suppose that $\Delta \vdash_C \varphi$. Then for finitely many $\psi_1, \ldots, \psi_n$ in Δ, we have $\psi_1, \ldots, \psi_n \vdash_C \varphi$. Let ψ be $(\psi_1 \wedge \ldots \wedge \psi_n)$. So $\psi \vdash_C \varphi$. Hence $\vdash_C \varphi \vee \neg\psi$. But every classical theorem is a dialetheic theorem. So $\vdash_D \varphi \vee \neg\psi$. Call the dialetheic proof in question Π. Now, by virtue of the dialetheic proof

$$\dfrac{\dfrac{\Pi}{\varphi \vee \neg\psi} \qquad \dfrac{\dfrac{\dfrac{}{\varphi}(1)}{\varphi \vee (\psi \wedge \neg\psi)} \qquad \dfrac{\dfrac{\psi \qquad \neg\psi}{\psi \wedge \neg\psi}}{\varphi \vee (\psi \wedge \neg\psi)}(1)}{\varphi \vee (\psi \wedge \neg\psi)}}{\varphi \vee (\psi \wedge \neg\psi)}$$

we have $\psi \vdash_D \varphi \vee (\psi \wedge \neg\psi)$, as required. *QED*

The upshot of this metatheorem is that whenever the classicist proves a

consequence φ from (a conjunction ψ of certain of) his axioms, the dialetheist can claim to be able to prove, from the same axioms, the surrogate result $\varphi \vee (\psi \wedge \neg\psi)$. Now, this would deprive us of the sought result φ if ψ were inconsistent. For in that case we would have $\psi \vdash_C \bot$, whence $\vdash_D \neg\psi$, whence $\psi \vdash_D (\psi \wedge \neg\psi)$. So we could not be sure that $\varphi \vee (\psi \wedge \neg\psi)$ was dialetheically deducible from ψ only because φ itself was dialetheically deducible from ψ. That, however, is how it ought to be. For, if ψ is inconsistent, how can we (as relevantists, at least) hope to be able to deduce φ from it anyway?

What about the case where the conjunction ψ of axioms is consistent? Here, the classicist, by virtue of having deduced φ from ψ, can claim to be able to assert φ. But the dialetheist can claim at most to be able to assert his surrogate result, $\varphi \vee (\psi \wedge \neg\psi)$. Classically, of course, this logically implies φ. But does it do so dialetheically? It would appear not. The inference from $\varphi \vee (\psi \wedge \neg\psi)$ to φ fails *if* ψ *is a dialetheia*. But we are assuming here that ψ is consistent, hence not a dialetheia. One wants to say, on behalf of the dialetheist, that *if* ψ is consistent, then the surrogate result—the existence of a dialetheic deduction of $\varphi \vee (\psi \wedge \neg\psi)$ from ψ—should really *show* that φ is indeed a *dialetheic*, and not just a classical, consequence of ψ. But such good intentions founder on the fact that the dialetheist himself can provide no way of saying this, hence no way of justifying one's saying it. For, as

Priest concedes (*IC*, p. 140),

> There is no statement that can be made which *forces* $[\psi]$ to behave consistently. This is one of the hard facts of dialetheic life.

Priest proceeds to use his notion of dialetheic consequence to define a more complicated notion of so-called *-consequence (*IC*, p. 150), which is sandwiched between dialetheic consequence and classical consequence. It affords the result (Theorem 6, p. 152) that the classical consequences of any classically consistent set of sentences are *-consequences thereof. The trouble, however, is that *-consequence does not admit of a proof-reduction (via a complete notion of effective proof) the way the Σ^0_1-relations of classical and dialetheic consequence do. This is because within the definition of *-consequence one existentially quantifies over dialetheic non-deducibilities. The resulting relation of *-consequence is therefore Σ^0_2, as Priest notes, and is of no epistemic use.

8 Gaps v. gluts

Logicians' slang for the failure of bivalence is to say that 'truth-value gaps' are being admitted. Of course, this is not entirely accurate, and not exactly

the way the anti-realist or intuitionist would wish to phrase the matter. After all, any claim of the form $\neg(\varphi \vee \neg\varphi)$ is intuitionistically inconsistent. Nevertheless, one can see what the slang is getting at. (At least the talk is of gaps, and not third values, which of course cannot be used to make sense of intuitionistic logic.) Now as Priest once put it to me in conversation many years ago, all that the dialetheist is doing is recommending truth-value *gluts*, which should be just as acceptable—according to Priest—as the intuitionist's truth-value *gaps*. This disarming and insouciant suggestion derives its plausibility entirely from a false picture of the philosopher as some sort of conceptual and logical engineer, who can fashion intellectual tools according to any kind of plan. The suggestion is that one is contemplating two quite comparable kinds of 'tweaking'—analogous, say, to serrating the edges of pieces of sheet metal, and/or punching dimples into them—and that there could be no objection to performing these tweaks in isolation, or in combination.

I believe nothing could be further from the truth. As an anti-realist who 'accepts' truth-value gaps (in the slang sense explained above), I advocate giving up the Law of Excluded Middle and all its equivalents. Indeed, I also advocate giving up the absurdity rule (*ex falso quodlibet*), thereby claiming that the correct logic is *intuitionistic relevant* logic. *But*—and this is a

crucial 'but'—I cannot fathom *what it would be* to *acknowledge* both P and $\neg P$ as true, for *any* choice of P. Hence I do not believe that any sentence of the form $P \wedge \neg P$ can *be* true. Note that *acknowledging* both P and $\neg P$ to be true would be something quite different from discovering that one's current beliefs logically committed one, by means of suasive proofs of which one had only just been made aware, to asserting both P and $\neg P$. In such a situation *one does not acknowledge both P and $\neg P$ as truths.* Quite the contrary: one acknowledges that their joint assertion would be an assertion of a logical impossibility, an assertion of something that must be false and *cannot* be true; and *that* is why—if one is rational—one immediately suspends belief in the overall conjunction of all the premiss-beliefs used in the proofs of P and of $\neg P$. The task of the rational agent, in such circumstances, is to start looking immediately for a premiss that can be *given up*—a former belief that is to be banished from one's stock of beliefs, in the hope that this will restore consistency to the remainder.[12] Of course, we have no effective test, in suitably rich languages, for the consistency of even finite sets of sentences. We can have no guarantee that upon giving up certain premisses of those proofs of P and of $\neg P$, we shall have attained a consistent, because reduced, set of beliefs. The spectre of inconsistency always lurks, as we try to form and reform our beliefs. But a spectre it indeed is, to which we remain

rationally averse. We recognize inconsistency as terrible, as *doxastically disastrous*, as something to be avoided at all costs.

This is why it is *so* crucial for Priest's project that he should succeed, definitively, on *at least one* of the examples of alleged dialetheia mentioned above. He has to show his opposition that there is at least one true contradiction—in the sense that no possible 'consistentist' story can dissolve the conflict or tension or absurdity revealed by the train of reasoning in question. Let us therefore now turn to a more detailed consideration of Priest's overly swift dialetheist embrace of the paradoxes, semantic and set-theoretic, and the Gödel-sentence for naive provability.

9 There are no dialetheias!

Priest writes as follows (*IC*, p. 11):

> The paradoxes are all arguments starting with apparently analytic principles concerning truth, membership etc., and proceeding via apparently valid reasoning, to a conclusion of the form 'α and not-α'. *Prima facie*, therefore, they show the existence of dialetheias. Those who would deny dialetheism have to show what is wrong with the arguments—of every single argument,

> that is. For every single argument they must locate a premise that is untrue, or a step that is invalid.

Here at last I intend to make good on my earlier claim that Priest detects dialetheia only by overlooking vital clues that point to the real culprits. I agree with Priest that the opponent of dialetheism has to address every single one of the paradoxical arguments. Given enough space, I would do that. But, since space is limited, I shall have to content myself here with briefer remarks aimed at persuading the reader that every such argument can be dealt with in the principled ways to be described below.

At the outset I should point out that Priest's last claim in the foregoing quote is in error. It is possible to scotch a paradoxical argument for a conclusion of the form $P \wedge \neg P$ *not* by locating a premise that is untrue, *nor* by locating a step that is invalid, but rather: by showing that *those steps are not put together, and cannot be rearranged, in the manner required for genuine conferral of truth-value on the conclusion(s) involved.* This is crucial for dealing with the semantic paradoxes, as we shall see.

9.1 There are no semantic dialetheias

Probably the most plausible candidate for the status of 'dialetheia' is the Liar paradox: *This sentence is false.* As everyone familiar with this paradox

knows, as soon as one reaches a (tentative) decision as to its truth-value, further reflection on the 'truth-conditions' of the Liar sentence leads one to toggle that value. As soon as it 'becomes' true, we see immediately that it 'becomes' false; and vice versa. Thus one never reaches a stable truth-value assignment—or so it seems. Priest takes from this phenomenon of toggling truth-values the seldom-drawn lesson that the Liar sentence is *both* true *and* false (for it takes two to toggle), rather than neither.

It is tempting to regard the Liar sentence as a quirk of language, having nothing to do with the relationship between language and the world. In a world devoid of intelligent beings using language there would be, as it were, no instantiation of any Liar-type phenomena. But, once we acknowledge that our world is one of language-users *and linguistic expressions*, the naive thought that the proper subject-matter of language does not include language itself loses its appeal. To the extent that the 'truth-conditions' of the Liar sentence are then 'in the world' (containing, as it does, the language of the Liar sentence, and the relations of reference involved), the Liar nevertheless seems to say nothing about the world *beyond that.* It is, as it were, obsessively autobiographical. As with the worst kind of party guest, one learns from it nothing about anyone or anything besides itself—if one learns anything at all. It is a nasty little knot in Priest's 'weft and warp of

language', about nothing but that same nasty little knot.

There are, however, other semantic paradoxes that do implicate, in their paradoxical truth-conditions, genuinely empirical facts in the world. Such is the case with Epimenedes's version of the Liar, the statement that *all Cretans are liars*. That Epimenedes was slagging his compatriots—that is, that Epimenedes was a Cretan, and therefore fell within the intended scope of his own generalization—is an empirical fact crucial to the statement's paradoxical status.

This consideration counsels against any attempted solution to the semantic paradoxes that tries to show that they are purely and essentially linguistic phenomena with such 'unworldly' truth-conditions as to justify the view that they can be safely disregarded. If they *do* have 'worldly' truth- (and falsity-) conditions, then these need to be accounted for. Moreover, we need to account for the truth- and falsity-conditions of the *non-paradoxical* sentences that talk about *both* empirical facts *and* the relationship between linguistic expressions and the world (including those expressions). Not every sentence of a semantically closed language is paradoxical. Many of them can be stably evaluated, unlike the Liar and its ilk.

It is beyond the scope of this section to survey the many and various proposals as to how one might 'solve' or 'accommodate', or 'avoid' or 'banish'

the semantic paradoxes. I want instead to stress just one line of approach to semantic paradox that I believe has the edge on all others, and *especially* on Priest's approach (which is simply to throw in the towel and regard them as dialetheias).

On the approach I recommend, one takes seriously the idea that any evaluation of a sentence as true or as false must be 'well-founded'. It is the main idea of the semantical treatment of semantically closed languages offered by Kripke, Woodruff and Herzberger. Moreover—a point seldom appreciated, and little essayed by writers more concerned with formal details of *semantic* evaluation—this idea has a very nice proof-theoretic expression, as follows. Any evaluation of a sentence as true (or as false) should take the form of a proof or justification that is in, or can be brought into, 'normal form'. Let us call a proof in normal form that shows a claim φ to be true a *truth-warrant*. And let us call a proof in normal form that shows a claim φ to be false a *falsity-warrant*. This notion of 'proof in normal form' is an informal one, pending further explication. A falsity-warrant for φ will take the form of a reductio ad absurdum: φ will be an assumption, and the conclusion of the reductio will be absurdity ($\bot$).

A careful examination of the various well-known semantic paradoxes reveals that the proofs of absurdity obtained from the would-be evaluations

of the sentences involved as true and/or as false *cannot be brought into normal form.* It is my conjecture that this is the distinguishing proof-theoretic feature of paradoxicality. The non-normalizability of the 'proofs' involved shows that *evaluations of paradoxes are not well-founded.* And *that* is why they are unstable. The vicious way in which linguistic self-reference and/or semantic closure makes itself manifest in the paradoxes is by subverting the canonical structure of truth-value conferral. Such structure, to be stable in its outcomes, has to be well-founded (that is, reducible, via a sequence of finitely many steps, to normal form). And this structure—when it obtains—can be laid bare in the appropriate value-warrant. With paradoxes, however, these warrants are wanting. The great majority of paradoxes have 'proofs' and/or 'refutations' yielding proofs of absurdity whose reduction-sequences enter into 'loops'; whereas a paradox like that of Yablo has in the same way a reduction-sequence that 'spirals' infinitely, ratcheting up a numerical parameter with every complete twist.[13]

This proof-theoretic diagnosis actually gives one a much-needed prophylactic against the imagined harm that a newly-discovered 'dialetheia' might bring; and it obviates the need to resort to Priest's Logic of Paradox as the 'correct' way to reason in the shadow of such possibilities. If someone identifies an erstwhile assertion as a dialetheia (because of its paradoxicality), then

the very justification that he will be obliged to provide for his claim that it is indeed a dialetheia will immediately yield the kind of non-normalizable 'proof of absurdity' (or 'reductio') that I have been discussing above. Moreover, *that* the construct in question is not normalizable is something that can be effectively discovered (just as, if one is given a method for computing the decimal expansion of a number, and the number happens to be rational, then that method is actually an effective method for discovering that the number in question is indeed rational).

This proof-theoretic diagnosis of paradoxicality affords one the intellectual luxury of being able to work with a semantically closed language (such as natural language) without having things end in explicit contradiction (as with the dialetheic logician), or blow up (as with the non-relevantist logician), in the presence of paradox. The paradoxes reveal themselves as *radically truth-valueless*. They are ultimately gappy. They could not be further from enjoying truth-value glut.

It remains to anticipate, and defuse, one potential objection that might be leveled against my putting forward this account as an anti-realist. It is related to our earlier observation that, for the intuitionist, any claim of the form $\neg(\varphi \vee \neg\varphi)$ is inconsistent. That is, the intuitionist cannot say of any particular sentence that it lacks a truth-value. But, the objector will now

ask, is not that exactly what one is saying of a paradoxical sentence (such as the Liar) when one claims that (by virtue of the normalizability test) it lacks a truth-value?

The answer is negative. Intuitionistically, the content of the two claims 'I have shown that φ is true' and 'I have shown that φ is false' is as follows:

(T) I have a truth-warrant for φ (call it $\begin{matrix}\Pi\\ \varphi\end{matrix}$); and

(F) I have a falsity-warrant for φ (in the form of a reductio $\begin{matrix}\varphi\\ \Sigma\\ \bot\end{matrix}$, say).

In general, a falsity-warrant for φ has to provide an effective method for turning any truth-warrant for φ into a normal proof of $\bot$. When the two warrants are given as natural deductions, this is effected by forming their accumulation $\begin{matrix}\Pi\\ (\varphi)\\ \Sigma\\ \bot\end{matrix}$ and normalizing it. So, in this case, the accumulation has to be reducible to a normal proof of $\bot$.

In the case where φ is the Liar sentence λ, however, this is not the case.

The Liar sentence affords the axiom

$$\lambda \leftrightarrow \neg T\lambda.$$

The 'truth-warrant' on offer from the dialetheist (to justify his claim that he has shown that λ is true) is the proof

$$\begin{matrix} \Pi \\ \vdots \\ \lambda \end{matrix} \qquad \dfrac{\dfrac{\dfrac{\overline{T\lambda}^{(1)} \qquad \dfrac{\dfrac{\overline{T\lambda}^{(1)}}{\lambda} \qquad \lambda \leftrightarrow \neg T\lambda}{\neg T\lambda}}{\bot}}{\neg T\lambda}\,(1) \qquad \lambda \leftrightarrow \neg T\lambda}{\lambda}$$

and the 'falsity-warrant' on offer (to justify his claim that he has shown that λ is false) is the reductio

$$\begin{matrix} \lambda \\ \Sigma \;\vdots \\ \bot \end{matrix} \qquad \dfrac{\dfrac{\lambda}{T\lambda} \qquad \dfrac{\lambda \qquad \lambda \leftrightarrow \neg T\lambda}{\neg T\lambda}}{\bot}$$

When we form their accumulation $\begin{matrix}\Pi\\(\lambda)\\\Sigma\\\bot\end{matrix}$ we find that it *cannot* be reduced to a normal proof of $\bot$. So it is not the case both that the would-be truth-warrant Π is genuinely truth-conferring (for the Liar sentence λ) and the would-be falsity-warrant Σ is genuinely falsity-conferring. Thus the conjunction of (T) and (F) fails when φ is the Liar sentence. Hence $\neg(\lambda \wedge \neg\lambda)$—just as the anti-realist would expect.

9.2 There are no set-theoretic dialetheias

Turning now to the set-theoretic paradoxes, we have to argue against Priest's view that, say, Russell's Paradox is both true and false. Priest wrote (*IC*, p. 120)

> I ... believe that the Russell set is both a member of itself and not a member of itself. I do not deny that it was difficult to convince myself of this, that is, to get myself to believe it. It seemed, after all, so unlikely. But many arguments ... convinced me of it. It is difficult to come to believe something that goes against everything that you have ever been taught or accepted,

> in logic and philosophy as elsewhere. This is just a psychological fact about the power of received views on the human mind.

The crux of Priest's case for the dialetheic status of

$$\{x|x \notin x\} \in \{x|x \notin x\}$$

is to be found on p. 37 of his book:

> I wish to claim that *Abs* and *Ext* are true, and in fact that they analytically *characterise* the notion of set.

These are the principles of naive Abstraction, and Extensionality, respectively (p. 35):

(Abs) $\exists y \forall x(x \in y \leftrightarrow \beta)$

(Ext) $\forall x(x \in z \leftrightarrow x \in y) \rightarrow z = y$

The consistentist finds it unfathomable that one would sooner believe that a set-theoretic proposition is both true and false than believe, instead, that it is mistaken to simply assume that any set-theoretic singular term (such as $\{x|x \notin x\}$) must have a denotation. Why insist on retaining naive abstraction at such philosophical (and logical) cost as dialetheism, rather than *learning* from Russell's paradox (as Frege himself did) that simplicity of

naive postulation cannot triumph over the scientific need for consistency in one's search for the truth? This is an area where the consistentist refuses to tout 'pragmatic' considerations of pithy postulation as trumping the need for a more careful analysis of the postulational bases of set-existence.

Naive abstraction cannot be analytic of the notion of set, unless that notion is itself inconsistent—which is what the dialetheist of course maintains. The consistentist, by contrast, believes that there *is* a consistent notion of set to be had, and that it *can* serve satisfactorily as the universal basis for mathematics. Such a notion of set will enjoy deep and analytic connections with fundamental features of rational thought about abstract things. The sought (consistent) theory will link the notions of predication, membership, set and existence in illuminating (and still analytic) ways.

It is an overly historicist dogma (which one would expect Priest, given his anti-dogmatism in general, to want to question) that naive abstraction is *analytic* of the notion of set. His taking it to be analytic is an important part of his case for Russell's paradox being both true and false. But here, by contrast, is another set of principles that we can take to be analytic of the notion of set. They are the introduction and elimination rules for the set-term-forming operator $\{x|\Phi(x)\}$ in a universally free logic. First, the introduction rule is as follows:

$$(\{\}I)\qquad \dfrac{\overset{(i)}{\underbrace{\Phi a\ ,\ \exists!a}} \qquad\qquad a \in t}{\vdots \qquad\qquad \vdots}$$

$$\dfrac{a \in t \qquad \exists!t \qquad \Phi a}{t = \{x|\Phi x\}}\,(i)$$

Three elimination inferences for the set-term-forming operator can be determined from (the subproofs for) its introduction rule as follows:

$$(\{\}E)\qquad \dfrac{t = \{x|\Phi x\} \quad \Phi u \quad \exists!u}{u \in t} \qquad \dfrac{t = \{x|\Phi x\}}{\exists!t} \qquad \dfrac{t = \{x|\Phi x\} \quad u \in t}{\Phi u}$$

The leftmost elimination inference unpacks the commitments generated by the need to have the leftmost subproof in the introduction rule. If the speaker asserts that t is the set of all Φ's, then, if we can establish of an existent u that it has property Φ, we may conclude that u is a member of t. This is the move from predication to set-membership, qualified by an appropriate existential presupposition.

The middle elimination inference is a special case of the Rule of Denotation in free logic, according to which the truth of any atomic predication

requires that each singular term involved has a denotation. (This makes the truth-conditions of such claims the standard Russellian ones.)

The rightmost elimination inference unpacks the commitments generated by the need to have the rightmost subproof in the introduction rule. If the speaker asserts that t is the set of all Φ's, then, if we can establish that u is a member of t, we may conclude that u has the property Φ, the defining property of the set t. This is the converse move from set-membership to predication.

The leftmost and rightmost elimination inferences are therefore tantamount to Church's conversion schema for set theory based on a free logic. They express the non-naive kernel of truth in naive abstraction. The existential presupposition of the leftmost rule is crucial; it ensures that the proof of the so-called Russell Paradox is no longer a proof that the system is inconsistent, but is simply a proof that there can be no such thing as the set of all things that are not members of themselves. So much for Russell's paradox being a dialetheia!

One can use just the rules given (plus the logical rule of substitutivity of identicals) to derive the principle of extensionality for sets:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{}{a \in d}(1) \quad \dfrac{}{a \in c}(1)}{\exists!a \quad \forall z(z \in c \leftrightarrow z \in d)}(3) \quad \dfrac{\dfrac{}{\exists!a \quad \forall z(z \in c \leftrightarrow z \in d)}(5)}{\exists!c}(1)}{a \in c \leftrightarrow a \in d \qquad a \in d} \quad \dfrac{}{a \in c \leftrightarrow a \in d \qquad a \in c}(1)}{a \in c \qquad\qquad a \in d} \quad \dfrac{\dfrac{}{b \in d \quad \exists!d \quad b \in d}(2)}{d = \{x | x \in d\}}(2)(4)}{c = \{x | x \in d\}}}{c = d}(3)}{\forall z(z \in c \leftrightarrow z \in d) \rightarrow c = d}(4)}{\forall y(\forall z(z \in c \leftrightarrow z \in y) \rightarrow c = y)}(5)}{\forall x \forall y(\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y)}$$

That Zermelo's axiom of extensionality is now a derived result, rather than an axiomatic stipulation, testifies to the deeper analysis achieved (of abstractive set-formation in terms of set-membership) via the introduction and elimination rules just specified.

I submit that in so far as a philosophical case for or against dialetheism in set theory needs to rest on a selection of principles claimed to be analytic of the notion of set, the introduction and elimination rules given above have a more convincing claim to be genuinely analytic of the notion of set than do the principles that Priest favors, namely naive abstraction

and extensionality. For the introduction and elimination rules above are obtained as one instance of a general method for generating such rules for abstraction operators.[14] Moreover, these rules provide the deep reason why extensionality holds. The rules bring out the consistent analytic connections among the notions of set, membership and predication in general. Moreover, those connections are brought out with the introduction-elimination form of analysis that is the hallmark of the anti-realist's theory of meaning: a theory stressing how the justificatory obligations that a speaker must discharge before making an assertion are matched by the inferential entitlements that the audience enjoys upon hearing it.

9.3 Gödel's Theorem for naive provability is not a dialetheia

We turn finally to the outstanding alleged example of a dialetheia in mathematical and logical foundations, namely Gödel's Theorem for naive provability. Priest's argument runs as follows (*IC*, p. 56):

> ... the consistency of our naive proof procedures entails a contradiction. For let T be (the formalisation of) our naive proof procedures. Then since T satisfies the conditions of Goedel's theorem, if T is consistent there is a sentence φ which is not provable in T, but which we can establish as true by a naive proof, and

> hence *is* provable in T. The only way out of the problem, other than to accept the contradiction, and thus dialetheism anyway, is to accept the inconsistency of naive proof.[fn] So we are forced to admit that our naive proof procedures are inconsistent. But our naive proof procedures just are those methods of deductive argument by which things are established as true. It follows that some contradictions are true, that is, dialetheism is correct.

Priest derives his contradiction from the assumption (among others) that T is a formalization of our naive proof procedures. In the metalogical reasoning, T is thus a parameter for an existential elimination on the premiss that *there is* a formalization of our naive proof procedures.

Extraordinarily, Priest does not consider the obvious point that it is this existential assumption that the Gödelian reasoning reveals to be in error. It may well first come as a surprise; but on further reflection the falsity of this assumption sinks in. Even in order to make sense of the informal notion of effective decidability of proofhood (not: recursiveness of the proof predicate, courtesy only subsequently of Church's Thesis), we have to have a rigorous conception of the symbolic resources and the combinatorial, algorithmic methods that must be involved in 'proof recognition', in order for this process

to be effective. What Gödel's theorem shows is that we can never once-and-for-all delimit, in this required rigorous manner, the resources of 'naive provability'. They are open-textured and indefinitely extensible.

To realize that this is so, and to disbelieve the claim that the notion of naive provability can be formalized, is a more rational reaction to Gödel's theorem than to believe that one has discovered a true contradiction. This anti-realist resolution of the 'paradox' is exactly that of the earliest intuitionists, for whom Gödel's theorem came as no great surprise, but as a kind of rigorous confirmation of their view that mathematical thought could not be constrained within a single formal system of proof.

Notes

[1]When reading this acronym, the reader should be aware that in fact I don't see at all! (how dialetheism could possibly be correct). This essay in a sense extends the critique that I undertook in Tennant (1998).

[2]In order to aid the exposition, the dialetheist will be picked up by anaphoric pronouns in the masculine, the anti-realist by ones in the feminine. (I make no provision for any hermaphroditic philosophical position.)

This device achieves gender-neutrality without resort to pronominal hybrids or plurals. But of course *she* prevails argumentatively over *him*; so a suit against sexism might still lie. I shall just have to ask my reader to live with that. The rationale is simple: there will be frequent references to Priest as the main proponent of dialetheism, so the masculine pronouns will be strongly associated with that position; and, for the sake of political correctness, I am willing, myself, to be a philosophical drag-queen—a Priscilla of the desert landscape. This, surely, should be a mitigating factor for the court of literary opinion.

[3]I note here that Priest's own 'Logic of Paradox' does not use the absurdity symbol, and does not have the usual rule of $\neg$-Elimination. Since, however, I take absurdity to be essential to our grasp of negation, and take the standard rule of $\neg$-Elimination to be essential to $\neg$'s being a sign of *negation*, I am allowing myself the expository liberty, at this stage, of 'thinking out loud' on behalf of the would-be dialetheist. See Tennant (1999).

[4]It is worth noting that the dialetheist's reason for rejecting the absurdity rule is that one can have both A and $\neg A$ *true*; but one does not wish to acknowledge, on that basis alone, that an arbitrary, topically unrelated proposition B is thereby true. So the dialetheist's relevantism is really very self-serving. A consistentist, by contrast, who is also a relevantist (such

as the present author) raises her objections to the absurdity rule, and to Lewis's first paradox $A, \neg A : B$, on considerations of relevance in reasoning. A consistentist need never fear that the premisses of Lewis's first paradox really could both be true.

[5]I am indebted here to Joshua Smith.

[6]'Negation, Absurdity and Contrariety'.

[7]Psychologists appear not to have a scientific term for this widespread phenomenon. Thinking that there might be a term for it, I put out a query on the email list of The Ohio State University's very large and disciplinarily diverse community of cognitive scientists. I received nine different answers. Two of the respondents independently suggested that the extreme, pathological form of the phenomenon I had in mind is Fregoli's Syndrome. Others offered terms such as 'hindsight bias', 'selective processing', 'motivational priming' and so on. One who suffers from Fregoli's Syndrome believes that a persecutor is adopting many different disguises in order to 'get him'. I hesitate to say that your average dialetheist is Fregolian. But it does make a nice contrast with 'Fregean'.

[8]See Tennant (2000).

[9]See Tennant (1987), ch. 10.

[10]See Tennant (1997).

[11]See fn. 2, if you have not yet done so.

[12]In a monograph under preparation, I try to give an account of this rational procedure, covering logical, epistemological and computational issues.

[13]Details of these proof-theoretic considerations can be found in Tennant (1982), (1995).

[14]Space is insufficient to allow a more detailed discussion of this claim. The reader is referred to Tennant (2004). We confine ourselves to pointing out here that rules for the definite description operator are exactly analogous to those for the set-term-forming operator; one simply replaces the membership predicate by the identity predicate in the rules stated in the text. See also Tennant (1978), §7.10.

References

Priest, Graham. 1982. 'To be and not to be: dialectical tense logic', *Studia Logica* XLI, 2/3, pp. 248–68.

Priest, Graham. 1987. *In Contradiction: A Study of the Transconsistent.* Nijhoff.

Tennant, Neil. 1978. *Natural Logic*, Edinburgh University Press, 1978 (2nd, revised edn. 1990).

Tennant, Neil. 1982. 'Proof and Paradox', *Dialectica* 36, pp. 265-296.

Tennant, Neil. 1998. 'Beyond the Limits of Thought', a critical study of Priest's book of that title (Cambridge University Press, 1994), in *Philosophical Books*, 39, pp. 20-38.

Tennant, Neil. 1995. 'On Paradox without Self-Reference', *Analysis* 55, pp. 199-207.

Tennant, Neil. 1997. *The Taming of The True*, Clarendon Press, Oxford.

Tennant, Neil. 1999. 'Negation, Absurdity and Contrariety', in D. Gabbay and H. Wansing (eds.), *What is Negation?*, Kluwer, Dordrecht, pp. 199–222.

Tennant, Neil. 2000. 'Anti-Realist Aporias', *Mind* 109, 436, pp. 831-860.

Tennant, Neil. 2004. 'A General Theory of Abstraction Operators', *The Philosophical Quarterly*, forthcoming.