Contradictory Concepts

Graham Priest
Departments of Philosophy,
Universities of Melbourne and St Andrews,
and the Graduate Center, City University of New York

June 12, 2011

For Dirk

1 Introduction

That we have concepts which are contradictory is not news. That there may be things which satisfy them, dialetheism, is, by contrast, a contentious view. My aim here is not to defend it, however;[1] and in what follows, I shall simply assume its possibility. Those who disagree are invited to assume the same for the sake of argument. The point of this essay is to think through a raft of issues that the view raises. In particular, we will be concerned with three inter-related questions:

1. Are the contradictions involved simply in our conceptual/linguistic representations, or are they in reality? And what exactly does this distinction amount to anyway?

2. Assuming that it is only in the former, can we get rid of them simply by changing these?

3. If we can, should we do so?

I will take up these issues, in that order, in the three parts of the paper. The journey will take us through a number of important issues in metaphysics, semantics, and epistemology.[2]

[1] This is done in Priest (1987), (1995), (2006). The topic is discussed by numerous people in the essays in Priest, *et al.* (2004) and the references cited therein.

[2] Much of the paper has been provoked by many years of enjoyable discussion with Diderik Batens—including his generous comments some earlier drafts of this paper. I

2 Dialetheism, Concepts, and the World

2.1 Contradiction by Fiat

A dialetheia is a pair of statements of the form A and $\neg A$ which are both true.[3] We may think of statements as (interpreted) sentences expressed in some language—a public language, a language of thought, or whatever. In this way they contrast, crucially, with whatever it is that the statements are *about*. Let us call this, for want of a better name, *the world*.

One thing that partly determines the truth value of a statement is its constituents: the meanings of the words in the sentence, or the concepts the words express. (Conceivably, one might draw a distinction here, but not one that seems relevant for present purposes.) Let us call these things, again for want of a better word, *semantic*. In certain limit cases, such as 'Red is a colour', semantic factors may completely determine the truth value of a statement. In general, however, the world is also involved in determining the truth value. Thus, the statement that Melbourne is in Australia is made true, in part, by a certain city, a certain country—literally part of this world.[4]

Given that dialetheias are linguistic, one natural way for them to arise is simply in virtue of linguistic/conceptual fiat. Thus, suppose we coin a new word/concept, 'Adult', and stipulate that it is to be used thus:[5]

- if a person is 16 years or over, they are an Adult
- if a person is 18 years or under, they are not an Adult

Now suppose there is a person, Pat, who is 17. Then we have:

(*) Pat is both an Adult and not an Adult.

Of course, one can contest the claim that the stipulation succeeds in giving the new predicate a sense. Deep issues lurk here, but I will not go into them,

thank him for all of this. The paper was orignally written for a conference in honour of his 60th birthday. The conference did not eventuate; but I'm delighted to dedicate the paper to him anyway. Diderik and I come at dialetheism from very different general perspectives. In particular, he gives much more importance to the role of context in semantics and epistemology than I do. See, e.g., Batens (1985), (1991), Batens and Meheus (1996). Some of the matters I discuss here are difficult disengaged from these differences. I do my best.

[3] Priest (1987), p. 4.

[4] Quineans would, of course, reject the distinction being made here between semantic and worldly factors. This is not the place to defend the notion of analyticity. I do so in Priest (1979) and (200+).

[5] See Priest (2001).

since my concern is with other matters. I comment only that the stipulation would seem to be just as successful as the dual kind, endorsed by a number of people,[6] which under-determine truth values—such as the following, for 'Child':

- if a person is 16 years or under, they are a Child
- if a person is 18 years or over, they are not a Child

Assuming the stipulation of the kind involved in 'Adult' to work, we have a certain sort of dialetheia here. We might call it, following Mares (2004), a semantic dialetheia. Note that, in terms of the distinction just drawn between semantic and worldly factors, the epithet is not entirely appropriate. The truth of (*) is determined only in part by semantics; some worldly factors are also required, such as Pat and Pat's age. Still, let us adopt this nomenclature.

2.2 Semantic Dialetheism

The dialetheism engendered by the definition of 'Adult' is transparent. There are other examples which are, plausibly, of the same kind, though they are less transparent. One of these concerns dialetheias apparently generated by bodies of laws, rules, or constitutions, which can also be made to hold by fiat. Thus, suppose that an appropriately legitimated constitution or statute rules that:[7]

- every property-holder shall have the right to vote
- no woman shall have the right to vote

As long as no woman holds property, all is consistent. But suppose that, for whatever reason, a woman, Pat, comes to own property, then:

- Pat both has and has not got the right to vote.

Examples that are arguably of the same kind are given by multi-criterial terms.[8] Thus, suppose that a criterion for being a male is having male genitalia; and that another criterion is the possession of a certain chromosomic structure. These criteria may come apart, perhaps as the result of surgery of some kind. Thus, suppose that Pat has female genitalia, but a male chromosomic structure. Then:

[6]E.g., Soames (1999).

[7]The example comes from Priest (1987), 13.2.

[8]See Priest (1987), 4.8, and Priest and Routley (1989), section 2.2.1.

- Pat is a male and not a male.

In this case, there is no fiat about the matter. One cannot, therefore, argue that the contradiction can be avoided by supposing that the act of fiat misfires. What one has to do, instead, is to argue that the conditions in question are not criterial. Again, I shall not pursue the matter here.

A final example that is, arguably, in the same camp, is generated by the Abstraction Principle of naive set theory:[9]

Abs Something is a member of the collection $\{x : A(x)\}$ iff it satisfies the condition $A(x)$.

This leads to contradiction in the form of Russell's paradox.[10] Again, there is no fiat here.[11] If one wishes to avoid the contradiction, what one must contest is the claim that satisfying condition $A(x)$ is criterial for being a member of the set $\{x : A(x)\}$—or, what arguably amounts to the same thing in this case, that **Abs** is true solely in the virtue of the meanings of the words involved, such as 'is a member of'.

Again, let us not go into this here. The point of the preceding discussion is not to establish that the contradictions involved are true, but to show that they may arise for reasons that are, generally speaking, linguistic/conceptual.

2.3 Contradictions in the World

Some have felt that there may be a more profound sort of contradiction, a contradiction in the world itself, independent of any linguistic/conceptual considerations. True, these are not strictly dialetheias as I have defined them, but let us call such things, following Mares again, *metaphysical dialetheias*.[12]

A major problem here is to see exactly what a metaphysical dialetheia might be. Even someone who supposes that dialetheias are solely semantic will accede to the thought that there are contradictions in the world, in one sense. None of the contradictions we considered in the previous sections, with perhaps the exception of Russell's paradox, is generated purely by semantic considerations. In each case, the world has to cooperate by producing an object of the appropriate kind, such as the much over-worked Pat. The

[9] Priest (1987), ch. 0.

[10] Take $A(x)$ to be $x \notin x$, and r to be $\{x : x \notin x\}$. Then we have $y \in r$ iff $y \notin y$. Hence, $r \in r$ iff $r \notin r$, and so $r \in r$ and $r \notin r$.

[11] An example of a similar kind, which does have an explicit element of fiat, is that of the Secretaries' Liberation League, given by Chihara (1979).

[12] Mares (2004). A number of people have taken me (mistakenly) to be committed to this kind of dialetheism. See Priest (1987), 20.6.

world, then, is such that it renders certain contradictions true. In that sense, the world is contradictory. But this is not the sense of contradiction that is of interest to metaphysical dialetheism. The contradictions in question are still semantically dependent in some way. Metaphysical dialetheias are not dependent on language at all; only the world.

But how to make sense of the idea? If the world comprises objects, events, processes, or similar things, then to say that the world is contradictory is simply a category mistake, as, then, is metaphysical dialetheism.[13] For the notion to get a grip, the world must be constituted by things of which one can say that they are true or false—or at least something ontologically similar.

Are there accounts of the nature of the world of this kind? There are. The most obvious is a Tractarian view of the world, according to which it is composed of facts. One cannot say that these are true or false, but one can say that they obtain or do not, which is the ontological equivalent. Given an ontology of facts to make sense, metaphysical dialetheism may be interpreted as the claim that there are facts of the form A and $\neg A$, say the facts that Socrates is sitting and that Socrates is not sitting. But as this makes clear, there must be facts of the form $\neg A$; and since we are supposing that this is language-independent, the negation involved must be intrinsic to the fact. That is, there must be facts that are in some sense negational, negative facts.[14] Now, negative facts have had a somewhat rocky road in metaphysics, but there are at least certain well-known ways of making sense of the notion, so I will not discuss the matter here.[15]

If one accepts an ontology of facts, fact-like structures, or something of this kind, then metaphysical dialetheism makes sense. Note, moreover, that if one accepts such an ontology, metaphysical dialetheism is a simple corollary of dialetheism. Since there are true statements of the form A and $\neg A$ then there are facts, or fact-like structures, corresponding to both of these.[16] All

[13]The point is made in Priest (1987), 11.1.

[14]This isn't quite right. Facts may not themselves be intrinsically negative: the *relation* between the facts that A and that $\neg A$ must be intrinsic. But this does not change matters much: there must still be some kind of negativity in reality. There are other ways of making sense of the idea that the world itself is contradictory. For example, it may be held that reality is composed of properties, and that objects are bundles of properties. Then a contadictory world would be one in which there are property-bundles which contain the properties P and $\neg P$, for some P. Again, there must be some kind of negativity in reality. This time, negative properties.

[15]In situation semantics, states of affairs come with an internal "polarity bit", 1 or 0. Facts with a 0 bit are negative. Alternatively, a positive fact may be a whole comprising objects and a positive property/relation; whilst a negative fact may be a whole comprising objects and a negative property/relation. For a fuller discussion of a dialetheic theory of facts, see Priest (2006), ch. 2.

[16]This assumes that all truths correspond to facts. In principle, anyway, one could

the hard work here is being done by the metaphysics; dialetheism itself is playing only an auxiliary role.

3 Conceptual Revision

3.1 Desiderata for Revision

Still, a metaphysics of facts (including negative facts) is too rich for many stomachs. Suppose that we set this view aside. If we do, all dialetheias are essentially language/concept dependent. In this way, they are, of course, no different from any other truths. But some have felt that, if this be so, contradictions are relatively superficial. They can be avoided simply by changing our concepts/language. Compare the corresponding view concerning vagueness, held, for example, by Russell (1923). All vagueness is in language. Reality itself is perfectly precise. Vague language and its problems may, therefore, be avoided by changing to a language which mirrors this precision.

Contradictions may certainly be resolved sometimes. Thus, consider the legal example concerning Pat and her rights. If and when a situation of this kind arises, the law would, presumably, be changed to straighten out the conflicting conditions for being able to vote. Note, however, that this is not to deny dialetheism. The situation before the change was dialetheic. The point of the change is to render it not so. Note, also, there is no *a priori* guarantee that making changes that resolve this particular contradiction will guarantee freedom from contradiction *in toto*. There may well be others. Indeed, making changes to resolve this contradiction may well introduce others. Laws comprise a complex of conceptual inter-connections, and the concepts apply to an unpredictable world. There is certainly no decision procedure for consistency in this sort of case; nor, therefore, any guarantee of success in avoiding dialetheism in practice.[17]

But maybe we could always succeed in principle. Consider the following conjecture:

- Whenever we have a language or set of concepts that are dialetheic, we can change to another set, at least as good, that is consistent.

endorse a view to the effect that some kinds of sentence are true in virtue of the existence of corresponding facts, whilst others may have different kinds of truth-makers.

[17]Actually, I think that the change here is not so much a change of concepts as a change of the world. Arguably, the change of the law does not affect the meanings of 'vote', 'right', etc. The statement 'Pat has the right to vote' may simply change its truth value, in virtue of a change in the legal "facts".

The suggestion is, of course, vague, since it depends on the phrase 'at least as good'. Language has many purposes: conveying information, getting people to do things, expressing emotions. Given the motley of language use, I see no reason to suppose that an inconsistent language/set of concepts can be replaced by a consistent set which is just as good for all the things that language does. I don't even know how one could go about arguing for this.

Maybe we stand more chance if we are a little more modest. It might be suggested that language has a primary function, namely, making statements (truth-apt sentences); and, at least for this function, given an inconsistent language/set of concepts, one can always replace it with a consistent one that is just as good. The claim that this is the primary function of language may, of course, be contested; but let us grant it here. We still have to face the question of what 'just as good' means now, but a natural understanding suggests itself: the replacement is just as good if it can describe every situation that the old language describes. Let us then consider the following conjecture:[18]

- Any language (set of concepts), L, that describes things in a dialetheic way, can be replaced by a consistent language (set of concepts), L', that can decribe every situation that L represents, but in a consistent way.

The conjecture is still ambiguous, depending how one understands the possibility of replacement here. Are we to suppose this to be a practical possibility, or a merely theoretical one? If the distinction is not clear, just consider the case of vagueness again. If there is no such thing as vagueness *in re*, we could, in principle, replace our language with vague predicates by one whose only predicates are crisp. But the result would not be humanly usable. We can perceive that something is red. We cannot perceive that it has a wavelength of between exactly x and y Ångstroms, where x and y are real numbers. A language with precise colour predicates would not, therefore, be humanly usable. Any language that can be used only by someone with superhuman powers of computation, perception, etc., would be useless.

To return to the case of inconsistency, we have, then, two questions:

- Can the language be replaced in theory?

[18]Batens (1999), p. 267, suggests that a denial of this conjecture is the best way to understand a claim to the effect that the world is inconsistent. '[I]f one claims that the world is consistent, one can only intend to claim that, whatever the world looks like, there is a language L and a [correspondence] relation R such that the true description of the world as determined by L and R is consistent.' He maintains an agnostic view on the matter. See also Batens (2002), p. 131.

- Would the replacement be possible in practice?

A few things I say will bear on the practical question,[19] but by and large I shall restrict my remarks to the theoretical one. This is because to address the practical question properly one has to understand what the theoretical replacement is like. In other words, not only must the answer to the theoretical question must be 'yes', the answer must provide a sufficiently clear picture of the nature of the replacement. Nothing I go on to say will succeed in doing this. I have stressed the distinction mainly to point out that even if the answer to the theoretical question is 'yes', the replaceability conjecture has another hurdle to jump if the victory for those who urge replacement is to be more than Phyrric.

So let us address the theoretical question. Is it true? Yes, but for entirely trivial reasons. L' can be the language with just one sentence, ¥. ¥ is true of any situation. Thus, every situation is describable, and consistently so. (The language does not even contain negation.) But this is not an interesting answer to the question, and the reason is obvious. We have purchased consistency at the cost of the loss of expressive power. To make the question interesting, we should require L' to have the same expressive power as L—or more. That is, everything that L is able to express, L' is able to express. The idea is vague. What, exactly is it for different languages to be able to express the same thing? But it is at least precise enough for us to be able to engage with the question in a meaningful way.

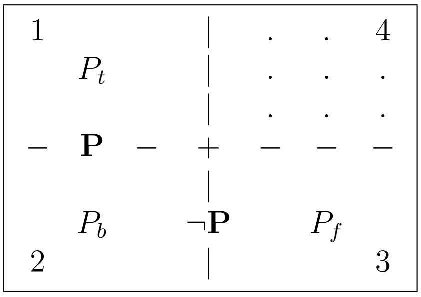
3.2 The Possibility of Revision

Return to the case of multiple criteria. A natural thought here is that we may effect an appropriate revision by replacing the predicate/concept *male* with two others, $male_1$, corresponding to the first criterion, and $male_2$, corresponding to the second. Pat is a male_2, but not a male_1, so the contradiction is resolved, and what used to be expressed by 'x is male', can now be expressed by 'x is $\text{male}_1 \vee x$ is male_2'. So far so good; but note that there is no guarantee that in this complex and unpredictable world the result will be consistent. The predicates 'male_1' and 'male_2' may themselves turn out to behave in the same inconsistent way, due to the fact that we have different criteria for 'genitalia' or 'chromosome'. More importantly, the resolution of this dialetheia depends on the fact that the old predicate falls neatly apart into two, individuated by different criteria. This will not be the case in

[19] I note that Batens (2002), p. 131, fn. 7, suggests that a consistent replacement for an inconsistent language might well be required to have a non-denumerable number of constants, which would make it humanly unusable.

general.—Just consider the case of 'Adult', for example, which is not multi-criterial in the same way.

We might attempt a more general way of resolving dialetheias as follows. Suppose we have some predicate, P, (like 'Adult) whose extension (the set of things of which it is true) and co-extension (the set of things of which it is false) overlap. Given that we are taking it that our predicates do not have to answer to anything in the world, we may simply replace P with the three new predicates, P_t, P_f, and P_b, such that the things in the extension of P_t are the things that are in the extension of P but not its co-extension; the things in the extension of P_f are the things that are in the co-extension of P but not its extension; the things in the extension of P_b are the things that are in both the extension and co-extension of P. The co-extension, in each case, is simply the complement. The situation may be depicted by the following diagram. For future reference, I call this the **Quadrant Diagram**. The numbers refer to the quadrants.



The left-hand side is the extension of P. The bottom half is the co-extension of P. Quadrant 4 comprises those things of which P is neither true nor false, and for present purposes we may take this to be empty.[20] The three new predicates have as extensions the other three quadrants. Each of the new predicates behaves consistently. Any dialetheia of the form $Pa \wedge \neg Pa$ is expressed by the quite consistent $P_b a$, and the predicate Px is now expressed, again, as a disjunction, $P_t x \vee P_b x$.[21]

So far so good. But recall that the new language must be able to express everything that the old language expressed. A necessary condition for this is that any situation described by the old language can be described by the new. To keep matters simple for the moment, let us suppose that the old language contains only the predicate P and the propositional operators of conjunction, disjunction, and negation. We have seen how any atomic sentence, A, of the

[20]Note that, if it is not, the same procedure can be used to get rid of truth value gaps.

[21]Batens (1999), p. 271 and (2002), p. 132 notes this idea. He also notes that in such a transition the theory expressed in the new language may lose its coherence and conceptual clarity, making it worse.

old language can be expressed equivalently by one, A^+, in the new. If this translation can be extended to all sentences, then any situation describable in the old language is describable in the new. The natural translation is a recursive one. For the positive connectives:

- $(A \vee B)^+$ is $A^+ \vee B^+$
- $(A \wedge B)^+$ is $A^+ \wedge B^+$

But what of $\neg A$? We certainly cannot take $(\neg A)^+$ to be $\neg(A^+)$. $\neg Px$ is true in the bottom half of the Quadrant Diagram, whilst $\neg(P_t x \vee P_b x)$ is not true in quadrant 2. In this case there is an easy fix. $\neg Px$ is equivalent to $P_b x \vee P_f x$. So we can deal with the atomic case. What of the others? There is a simple recipe that works:

- $(\neg(A \vee B))^+$ is $\neg(A^+) \wedge \neg(B^+)$
- $(\neg(A \wedge B))^+$ is $\neg(A^+) \vee \neg(B^+)$
- $(\neg\neg A)^+$ is A^+

In other words, we can drive the negations inwards using De Morgan laws and double negation until they arrive at the atoms, where they are absorbed into the predicate. In this way, every sentence of the old language is equivalent to a consistent one in the new language.

The end can therefore be achieved for this simple language. But, for the strategy to work, it must be implementable with much more complex and realistic languages. In particular, it must work for conditionals, quantifiers of all kinds, modal and other intentional operators; and is not at all clear that it can be made to do so. At the very least, then, the onus is on the proponent of the strategy to show that it can.

Moreover, there are general reasons for supposing that it cannot. Extending the translation to intentional operators would seem to provide insuperable difficulties. Take an operator such as 'John believes that', $\mathfrak{B}$. How are we to handle $\mathfrak{B}A$? The only obvious suggestion that $(\mathfrak{B}A)^+$ is $\mathfrak{B}(A^+)$, and this will clearly not work. Even logical equivalence does not guarantee equivalence of belief: one can believe $\neg\neg A$ without believing A, for example. Hence, even if A and A^+ express the same situation in some sense, one could have $\mathfrak{B}A$ without having $\mathfrak{B}A^+$. The trouble is that belief and similar mental states are intentional, directed towards propositions/sentences. These seem to be integral to the intentional state in question, and so cannot be eliminated if we are to describe the intensional state. (Indeed, the same is true of *all* conceptual revisions. If people's thoughts are individuated in terms of old concepts, one cannot describe those thoughts if the concepts are junked.)

One possible suggestion at this point is simply to take $(\mathfrak{B}A)^+$ to be $\mathfrak{B}A$ itself. Of course, if we leave it at that, we have not rid ourselves of the dialetheic concepts, since these are still occurring in the language. But we might just treat $\mathfrak{B}A$ as a new atomic sentence—a single conceptual unit. The problem with this is clear. There would be an infinite number of independent atomic sentences, and the language would not be humanly learnable. The construction would fail the practicality test. And even then, given that the language contains other standard machinery, there would still be expressive loss. For example, we would no longer have a way of expressing things such as $\exists x(Px \wedge \mathfrak{B}Px)$ or $\forall p(\mathfrak{B}p \rightarrow p)$.

Nor is this just a problem about mental states. It applies to intensional notions generally. Thus, consider the statement 'That A confirms that B'. This is not invariant under extensional equivalence. Let us make the familiar assumption that all creatures with hearts are creatures with kidneys.[22] Consider the information that a_1, ..., a_n are creatures of kind k with a heart. This confirms the claim that all creatures of kind k have a blood circulation system. The information is extensionally equivalent to the information that a_1, ..., a_n are creatures of kind k with kidneys. This does not confirm the claim that all creatures of kind k have a blood circulation system.[23]

3.3 Expressive Loss

But worse is yet to come for the conjecture that we can, in theory, always replace an inconsistent language with a consistent one. Suppose that the project of showing that every situation describable in the old language can be described in the new can be carried out, in the way just illustrated or some similar way. This is not sufficient to guarantee that there is no expressive loss.

Consider the naive notion of set again. This is characterised by the schema:

Abs $x \in \{y : A(y)\} \leftrightarrow A(x)$

which gives rise to inconsistency, as we have noted. Let us suppose that it were replaced with different notions in the way that we have just considered. Thus, we have three predicates $\in_t$, $\in_b$, and $\in_f$, where $x \in y$ is expressed by $x \in_t y \vee x \in_b y$. Let us write this as $x \in' y$. Given the above schema, we have:

[22] As a matter of fact, Diderik (an amateur beekeeper) tells me, this is false. Bees have a heart, but no kidneys.

[23] More generally, relations relevant to confirmation are well known not to be invariant under linguistic transformations. See, e.g., Miller (1974).

Abs′ $x \in' \{y : A(y)\} \leftrightarrow A(x)$

and in particular:

$$x \in' \{y : \neg y \in' y\} \leftrightarrow \neg x \in' x$$

Substituting $\{y : y \notin' y\}$ for x gives us Russell's paradox, as usual. We have not, therefore, avoided dialetheism.[24] Why is this not in conflict with the discussion of the last section? The reason is essentially that the procedure of driving negations inwards, and finally absorbing them in the predicate, produces a language in which there is no negation. The instance of Abs′ that delivers Russell's paradox cannot, therefore, even be formed in this language, since it contains negation. The procedure guarantees, at best, only those instances of Abs′ where $A(x)$ is positive (negation-free).

We face a choice, then. Either dialetheism is still with us, or we lose the general schema that we had before. But the Schema effectively characterizes the naive concept of set membership. So if we go the latter way, notwithstanding anything heretofore, there is still an expressive loss. We have lost a concept which we had before, with no equivalent replacement. We have lost the ability to express arbitrary set formation.

This provides us with an argument as to why we may not always be able to replace an inconsistent language/conceptual scheme with one that is consistent. There are cases where this can be done only with conceptual impoverishment. That one may achieve consistency by throwing away a concept is not surprising. The notion of truth gives rise to contradictions. No problem: just throw it away! But such a conceptual impoverishment will leave us the poorer.

If we were throwing away useless things, then, one might argue, this is no loss. But contradictory concepts may be useful; indeed, *highly* useful—contradictions notwithstanding. Thus, for example, the ability to think of the totality of all objects of a certain kind—closely related to our ability to quantify over all such objects, and to form them into a set—would seem to be inherent in our conceptual repertoires. It plays an essential role in certain kinds of mathematics (such as category theory), and in our ruminations about the way that language and other conceptual processes work. But abilities of this kind drive us into contradictions of the sort involved in discussions of the limits of thought.[25] If we threw away the ability to totalise in this way,[26] maybe this would restore consistency; but the cost would be

[24] This is observed by Batens (2002), p. 132. See also his (1999), p. 272.

[25] A detailed discussion of all this can be found in Priest (1995).

[26] And can we? If one has such an ability, how *can* one lose it, short of some brain trauma?

to cripple the kind of mathematical and philosophical investigations that depend on it. To do so simply in the name of consistency would be like doing so in the name of an arbitrary and repressive government *diktat*.

The situation is not to be confused with that in which the concept of phlogiston was "replaced" by that of oxygen. We did not, in fact, dispense with the concept of phlogiston. We can still talk about it now. What was rejected there was the claim that something satisfies this notion. We now think that nothing does; in consequence, the concept is of no scientific use.

4 The Norm of Revision

4.1 Methodological Consistency

As we see, one cannot always replace an inconsistent language/set of concepts with a consistent one in a satisfactory way. But if we can, should we? Inconsistency should certainly be replaced sometimes. One of the functions of law is to guide action. Contradictory laws may frustrate this purpose—should we or should we not allow Pat to vote? But as far as the purely descriptive function of language goes, there would appear to be little point. The language/concepts provide a perfectly adequate representation of reality. If it ain't broke, don't try to fix it.[27]

There is no obvious reason why we should do so, but Batens (1999), (2002), has argued that it is sound methodology to replace an inconsistent set of concepts with a consistent one if we can do so, *ceteris paribus*. He cites Earman according to whom, though we have no reason to suppose the world to be deterministic, there is methodological virtue in trying to find deterministic theories. The same, according to Batens, is true of consistency. The virtue in the case of consistency is, of course, somewhat different. Batens calls it 'precision' and illustrates as follows:[28]

> Let P be a unary predicate of the language of an inconsistent theory, and let some paraconsistent logic **PL** be the underlying logic of the theory. ... P divides the objects into three subsets: those that are P only, those that are $\neg P$ only, and those that are both P and $\neg P$. The sentence $Pa \wedge \neg Pa$ unequivocally locates a amongst the objects that are inconsistent with respect to P. There is no way, however, to locate a in the union of the first

[27] See Priest (1987), 13.6.

[28] Batens (1999), p. 271. I change his notation to bring it into line with the rest of this essay.

> and [second][29] set, not in the [third] only. Compare this situation to the one in which P belongs to a consistent theory (of which the underlying logic validates EFQ). Here P introduces two sets only; Pa unequivocally locates a in the first set, $\neg Pa$ unequivocally locates it in the second one. If there is a need for three sets, then one introduces a family of predicates (Carnap's term), say P_1, P_2, and P_3. The predicates of the family are exhaustive and mutually exclusive. So they divide the objects into three sets, P_1a unequivocally locates a is the first, P_2a in the second, and P_3a in the third. Whether you need two or three sets (this depends on 'the world') the consistent theory is more precise.

What to say about this argument?

4.2 Precision

To evaluate it, let us start by getting clear about the notion of precision in play. Note that this has nothing to do with the truth conditions of negation: these are just as precise in the paraconsistent as in the classical case. Rather, the sense of precision at issue[30] is as follows. Refer again to the Quadrant Diagram. If we want to express the claim that an object, a, is in the union of quadrants 1 and 3, and we have the consistent language at our disposal, we can say $P_t a \vee P_f a$. But if we have only the inconsistent concepts at our disposal, the best we can do is:

(1) $Pa \vee \neg Pa$

We cannot rule out a's being in quadrant 2. In particular, the following won't do:

(2) $(Pa \vee \neg Pa) \wedge \neg(Pa \wedge \neg Pa)$

Given the standard semantics of negation and conjunction, $\neg(Pa \wedge \neg Pa)$ is true in quadrants 1, 2, and 3. The precision that Batens has in mind then, is the ability to charactise a's status in a more fine-grained way. We may now ask two crucial questions. First, is precision in this sense, a virtue? Second, does an inconsistent theory lack it? Take them in that order.

Precision is not necessarily a virtue. Recall the case of vagueness again. In our ordinary colour vocabulary, we can say that something is red, or

[29]The text actually interchanges 'second' and 'third', but I take this to be a slip. The union of the first and third sets in Batens' enumeration is characterised by P.

[30]Clarified by Batens in correspondence.

some hue thereof, but we have no way of saying that it has some *precise* redness. Neither is this a problem. Our colour language is quite adequate for normal purposes. True, we can resort to the language of frequencies, but such discourse has imprecision of its own. We cannot specify a range of between x and y Ångstroms if x and y are real numbers not referred to by names or descriptions in our language. (There will always be such numbers, since the totality of real numbers is uncountable.) Nor, generally speaking, does this matter. Indeed, precision may not just fail to be a virtue; it may be a vice. As already observed, our colour language works only because its vagueness matches the limitations of our perceptual apparatus: a precise colour vocabulary would be unworkable. Another example: you do not know how to play cricket, and ask a friend. Reading out the rule book would provide a very precise answer, but it would not be very helpful. One needs the main points; details obfuscate. Sufficient to the occasion is the precision therefor.

Let us turn now to the second point. Are paraconsistent theories imprecise in the way suggested? Note, at the start, that there is nothing about paraconsistency, or even dialetheism as such, that prevents the language containing an operator that behaves as does classical negation. It is just that the operator isn't negation. Of course, this possibility is ruled out if one wishes to run a dialetheic or paraconsistent line on the paradoxes of self-reference, since such an operator gives rise to triviality-producing contradictions.

However, assuming that there is no operator with the powers of classical negation in the language, is it the case that using a paraconsistent logic we cannot express the consistent parts of the Quadrant Diagram? As noted, $\neg(Pa \wedge \neg Pa)$ will not do. But it can be expressed if there is a truth predicate, T, and some naming device for sentences, $\langle . \rangle$, in the language. The four quadrants can be expressed by the following four conditions:

1. $T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle$
2. $T\langle Pa\rangle \wedge T\langle \neg Pa\rangle$
3. $\neg T\langle Pa\rangle \wedge T\langle \neg Pa\rangle$
4. $\neg T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle$

In particular, the union quadrants 1 and 3 can be specified by:

(*) $(T\langle Pa\rangle \vee T\langle \neg Pa\rangle) \wedge \neg(T\langle Pa\rangle \wedge T\langle \neg Pa\rangle)$.

Note that $\neg T\langle Pa\rangle$ is not equivalent to $T\langle \neg Pa\rangle$;[31] if it were, and given the T-schema, $\neg(T\langle Pa\rangle \wedge \neg T\langle \neg Pa\rangle)$ would be equivalent to $\neg(Pa \wedge \neg Pa)$.

[31] See Priest (1987), 4.9.

Batens would no doubt object at this point. If the negation used in (*) is paraconsistent (which I take it to be), the sentence could be true even though $T\langle Pa\rangle \wedge T\langle\neg Pa\rangle$ (second quadrant) holds as well. The diagram itself might be inconsistent. If one objects in this way, the point is no longer that the facts of the diagram cannot be represented, but that they cannot be represented in a way that guarantees consistency. This is true: there is nothing a paraconsistent logician can say that enforces consistency. But this is no objection, since exactly the same holds of one who subscribes to classical logic! Such a person can, of course, assert $\neg(Pa \wedge \neg Pa)$ where $\neg$ is, or is taken to be, classical negation; this does prevent them endorsing $Pa \wedge \neg Pa$ as well. If they do, then they will be committed to everything. This, I take it, is the relevance of the reference to EFQ (*ex falso quodlibet*, $\{A, \neg A\} \vdash B$) in Batens' words. As he says elsewhere:[32]

> To adopt the ex falso quodlibet has dramatic consequences. Someone who asserts $\neg A$ is truly committed to the rejection of A: asserting A would commit one to triviality. The dramatic character lies in the fact that triviality constitutes the end of all thinking...

However, enforcing collapse into triviality can be secured by perfectly legitimate paraconsistent means as well. A paraconsistent logician may endorse a claim of the form:

(**) $(Pa \wedge \neg Pa) \to \perp$

where $\to$ is a detachable conditional, and $\perp$ is a logical constant that implies everything. (It may be defined as $\forall x Tx$.) A subsequent endorsement of $Pa \wedge \neg Pa$ will then commit them to everything (and so, presumably, force then to give up something to which they are committed).[33] The classical logician is, in the end, then, no better off than the paraconsistent logician.

Batens addresses essentially this matter explicitly in a later article.[34] He claims that, at least without Boolean negation, there is no way to express the thought that two claims, A and B, are incompatible, or not jointly possible. You can't simply say $\neg\Diamond(A \wedge B)$. For that could be the case, even though

[32]Batens (1990), p. 222.

[33]The conditional (**) is not a logical truth; one can think of it as part of the theory of P. (Classical logic, in effect, promotes this contingent truth into a necessay (logical) one.) This is beside the point, though; what is at issue is whether triviality can be triggered in a paraconsistent context.

[34]Batens (2002), pp. 142-4. The objection is given under the rubric 'objections to dialetheism'. This is misleading, for he himself is a dialetheist. He holds that there are inconsistent concepts, and so dialetheias.

$A \wedge B$ is true. Much the same considerations apply. Ruling out in the pertinent sense is a function of EFQ,[35] and this can be done by a paraconsistent logician using $\perp$. Batens points out that even the trio of claims A, B, and $(A \wedge B) \rightarrow \perp$ can be endorsed by someone who is prepared to accept that everything is true—trivialism. But classical logic is no defence against trivialism: trivialists *are* classical logicians! They endorse, and reason in accord with, all the principles of classical logic, including EFQ.

Batens goes on to argue for a further claim: the fact that it is *logically* possible to accept everything in a paraconsistent logic is a shortcoming in the context of theory-revision: to handle such revision in the case of inconsistent data requires an adaptive logic. Now, for a start, dialetheists can use adaptive logics. One is endorsed in Priest (1987), ch. 16. But the point that theory-revision goes beyond logic is correct. Theory-revision uses norms of rationality that go beyond those of mere logic—adaptive or otherwise. It is rational to replace an old theory with a new one if that theory performs better on the aggregate of positive criteria for theory-choice, such as simplicity, unifying power, etc.[36] And because the mechanism is broader than that of logic, it can account for change in the received logical theory too.[37]

4.3 Boolean Negation—Again

There is another, and harder, point here.[38] Batens is, in fact, advocating not just replacing inconsistent concepts with consistent ones, but replacing concepts employing a paraconsistent logic with concepts employing classical logic. The possibility of this presupposes that the notions of classical logic make sense, and, in particular, that Boolean negation does so. It seems to me that it does not.

The idea may seem absurd. Can't we simply recognise the meaning of classical negation? Unfortunately, no. Things do not wear their meaning—or lack thereof—on their face. Whether something is meaningful can be determined only by the articulation and application of a theory of meaning. A classical theory of meaning may deliver the result that Boolean negation is meaningful. But the adequacy of a classical theory of meaning is, in part,

[35]There are other senses to do with denial. On these, see Priest (2006), ch. 6. The claim about the expressive limitations of paraconsistent logic has been pressed most strongly by Shapiro (2004). See the discussion in Priest (1987), 20.4.

[36]One of these criteria may well be consistency. See Priest (2006), Part 3, for a full account of the details.

[37]See Priest (2006), p. 151, and for more detail, Priest (200+).

[38]Many of the arguments in this section are given in more detail in Priest (1987) and (2006). Hence, the treatment here can be reasonably terse.

what the debate at hand is all about. And as far as Boolean negation goes, a dialetheic theory of meaning can side with an intuitionistic theory of meaning in holding that it does not. Nor need a classical logician feel smug about the matter. *No one*, on pain of triviality, can endorse both a classical notion of negation and an unrestricted truth predicate.[39] Hence a classical logician must deny the meaningfulness of the latter notion, which seems just as bad, if not worse.[40]

Why should we suppose that classical negation does not make sense? In a nutshell, the argument goes as follows.[41] A connective that satisfies the rules of Boolean negation appears to be in the same camp as Prior's connective *tonk* (a connective, †, satisfying the rules $A \vdash A \dagger B$, $A \dagger B \vdash B$). If such a connective is in the language, then any sentence entails any sentence. Similarly, if a connective obeying the rules of Boolean negation is in the language, then any sentence entails any sentence (in the context of self-reference and the T-schema). Since *tonk* is meaningless; so is Boolean negation.[42]

But may we not show that Boolean negation is legitimate by giving it truth conditions in the standard way? Say:

- $\neg A$ is true in (a world of) an interpretation if A is not true (there).

The truth conditions may determine a perfectly legitimate notion, but to establish that they deliver a notion underwriting EFQ we need to do more than state truth conditions; we need to reason about what follows from them. And—to cut a long story short—we have no reason to suppose that the conditions do so unless we reason classically—in particular, using Boolean negation—in the metalanguage, and so presuppose the meaningfulness of the very notion whose meaningfulness we are supposed to be establishing.

This argument has been contested by Batens, who raises a number of objections.[43] One is the following. Negation may not actually be needed to give the truth conditions of negation. Thus, assume for the sake of illustration

[39] See Priest (2006), ch. 5.

[40] Of course, in consistent contexts, a paraconsistent negation may behave indistinguisably from classical negation. That does not mean that it is classical negation that is being used.

[41] See Priest (2006), ch. 5.

[42] A number of people have suggested to me that *tonk* is perfectly meaningful. Its meaning is just defective. I have no objection to this if one can give a satisfactory account of defective meaning (which I don't know how to do). The point is that, whatever one says about *tonk*, the same applies to Boolean negation.

[43] Batens (2002), p. 141 ff. Batens' own views about meaning depend heavily on his contextualism, and he would dispute my whole approach to meaning, but that is far too big an issue to take on here. In what follows, his objections may be taken to be *ad hominem*.

that we have a three-valued semantics with the truth values $\{t\}$, $\{t, f\}$, and $\{f\}$, the first two being designated. We may give the truth conditions of negation without using negation, and thereby presupposing its properties, as follows:

- $\neg A$ is $\{t\}$ if the value of A is $\{f\}$
- $\neg A$ is $\{f\}$ if the value of A it is $\{t\}$ or $\{t, f\}$.

or as:

- $\neg A$ has the value $\{t\}$ if the value of A is designated, and $\{f\}$ otherwise.

Indeed we may; but in either case we have to reason using negation to show that the $\neg$ so defined grounds EFQ. Thus, we may establish that A and $\neg A$ never take a designated value together. But to establish that $A, \neg A \vDash B$, we need to reason that, since the premises are never both designated, whenever the premises are designated, so is the conclusion. This is an argument of the form $\neg A \vdash A \rightarrow B$, which a paraconsistent logician is not going to accept if $\rightarrow$ is a detachable conditional. Alternatively, if validity is defined in terms of truth preservation using the material conditional, the inference to the validity of EFQ is of the form $\neg A \vdash A \supset B$. The inference is perfectly correct; but now we cannot get to B from A and $\neg A$ because the material conditional does not detach. We still do not have the force of explosion.

Batens' next point concerns classical logic and metatheoretic reasoning. It might appear that paraconsistent logicians are committed to the meaningfulness of Boolean negation, since they use it themselves when doing the metatheory of paraconsistent logic. Certainly, it is often assumed that the metatheory of a logic, classical or non-classical, must be undertaken in classical logic; but this is false. There is no reason why, in principle, the metatheory of a logic cannot be undertaken in a non-classical logic. In practice, metatheory is done in informal set theory; the question is how to regiment this formally. Standardly, classical logic and ZF set theory suffice; this does not show that they are necessary. Batens points out that paraconsistent logicians have not given a great deal of thought to the paraconsistent regimentation of metatheory. This is a fair point. How best to turn the trick is still moot. One way of doing it is explained in Priest (1987), ch. 18. I will not reproduce the details here, but the idea is that a certain understanding of paraconsistent set theory allows a paraconsistent logician simply to appropriate classical metatheoretic arguments.

Batens' third point is swift. He says:[44]

[44] Batens (2002), p. 141.

> Once a ... dialetheist theory for handling functions will be around, things will get serious. The question will not any more be whether some metalinguistic negation is paraconsistent rather than classical, but whether $t \in \{f\}$ and hence $t = f$.

The point, I take it, is that if triviality looms anyway, the blame cannot be laid at the door of Boolean negation. Why, exactly, triviality looms, Batens does not spell out here, but in Batens (1990), § 5, when he considers the matter, he notes that if ν is an evaluation function, and we can show that $\{t\} = \nu(A) \neq \nu(B) = \{t, f\}$, this does not '*rule out*' (his italics) $\nu(A) = \nu(B)$, and so $\{t\} = \{t, f\}$. Indeed, in a sense, it does not. The question, though, is whether there is any reason to believe what is not ruled out to be true true. Without this, the point has no bite.[45]

Batens' final major objections is to the effect that Boolean negation must be meaningful because people can reason in accord with the rules of classical logic—or any other logic. However, this may be explained without resort to an appeal to the meaningfulness of Boolean negation. Using a non-monotonic adaptive logic—of the kind pioneered by Batens—a paraconsistent logician may reason in exactly the same way as a classical logician in consistent situations, though the negation employed has exactly the same meaning as it does in the underlying monotonic paraconsistent logic.[46] One cannot say the same thing about reasoning in accord with other logics, or (fortunately!) about reasoning using classical logic in inconsistent situations; but there is a much more general point here. Given *any* set of putatively logical rules—at least as long as they are not too complex—a person may follow them and know that they are doing so. Nothing follows about meaning at all, however. One can just as well follow the rules for reasoning with *tonk*. And just as with *tonk*, following the rules may lead to disaster.[47]

5 Conclusion

This has been an essay about contradictory concepts, concepts which generate dialetheias. Assuming there to be such things, three further claims are

[45] There is, in fact, a small industry of people (not including Batens) who have attempted to produce arguments to this effect, based on "strengthened paradoxes". The arguments are discussed and rejected in Priest (1987), 20.3.

[46] See Priest (1987), 8.6 and esp. ch. 16.

[47] Batens has one more objection, not about Boolean negation, but about the classical material conditional. The claim is that this is needed to deal with restricted quantification. This objection is answered in Priest (1987), 18.3. A more general discussion can be found in Beall *et al*, (2006).

tempting. 1: Dialetheias are merely in our concepts; there are no such things as contradictions *in re*. 2: Dialetheias may always be removed by revising our concepts. 3: Even if this is not the case, if they can be, they should be, *ceteris paribus*. We have seen that there are ways in which one may resist all of these suggestions. I think that Hegel would have been delighted; but that is another matter.[48]

References

[1] Batens, D. (1985), 'Meaning, Acceptance and Dialectics', pp. 333-60 of J. C. Pitt (ed.), *Change and Progress in Modern Science* (Dordrecht: Reidel).

[2] Batens, D. (1990), 'Against Global Paraconsistency', *Studies in Soviet Thought* 39: 209–29.

[3] Batens, D. (1991), 'Do we Need a Hierarchical Model of Science?', pp. 199-215 of J. Earman (ed.), *Inference, Explanation, and Other Frustrations. Essays in the Philosophy of Science* (Berkeley, CA: University of California Press).

[4] Batens, D. (1999), 'Paraconsistency and its Relation to Worldviews', *Foundations of Science* 3: 259-83.

[5] Batens, D. (2002), 'In Defence of a Programme for Handling Inconsistencies', pp. 129-50 of J. Meheus (ed.), *Inconsistency in Science* (Dordrecht: Kluwer Academic Publishers).

[6] Batens, D., and Meheus, J. (1996), 'Steering Problem Solving between Cliff Incoherence and Cliff Solitude', *Philosophica* 58: 153-87.

[7] Beall, JC, Brady, R., Hazen, A., Priest, G., and Restall, G. (2006), 'Restricted Quantification in Relevant Logics', *Journal of Philosophical Logic* 35: 587-98.

[48]Versions of this paper, or parts of it, have been given under various titles at a number of philosophy departments and conferences over the last few years: the University of Melbourne, the University of Queensland, the *Australasian Association of Philosophy* (Australian National University), the University of Chapel Hill (North Carolina), the University of Connecticut, the Massachusetts Institute for Technology, *Logic and Reality* (Universities of Namur and Louvain la Neuve), the University of Gent, the City University of New York (Graduate Center), the *Fourth Cambridge Graduate Conference on the Philosophy of Logic and Mathematics*. I thank the participants for many lively discussions and helpful comments. Thanks for comments go also to two referees for the volume in which this essay appears.

[8] Chihara, C. (1979), 'The Semantic Paradoxes: a Diagnostic Investigation, *Philosophical Review* 13: 117-24.

[9] Mares, E. (2004), 'Semantic Dialetheism', ch. 16 of Priest, *et al.* (2004).

[10] Miller, D. (1974), 'Popper's Qualitative Theory of Verisimilitude', *British Journal for the Philosophy of Science* 25: 160-77.

[11] Priest, G. (1979), 'Two Dogmas of Quineanism', *Philosophical Quarterly* 29: 289-30l.

[12] Priest, G. (1987), *In Contradiction: a Study of the Transconsistent*, (Dordrecht: Martinus Nijhoff); second edition, Oxford: Oxford University Press, 2006. (References are to the second edition.)

[13] Priest, G. (1995), *Beyond the Limits of Thought* (Cambridge: Cambridge University Press); second edition, Oxford: Oxford University Press, 2002.

[14] Priest, G. (2001), Review of Soames (1999), *British Journal for the Philosophy of Science* 52: 211-5.

[15] Priest, G. (2006), *Doubt Truth to be a Liar* (Oxford: Oxford University Press).

[16] Priest, G. (200+), 'Logical Disputes and the *a Priori*', to appear.

[17] Priest, G., Beall, JC, and Armour-Garb, B. (eds.) (2004), *The Law of Non-Contradiction: New Philosophical Essays* (Oxford: Oxford University Press).

[18] Priest, G., and Routley, R. (1989), 'The Philosophical Significance of Paraconsistent Logic', ch. 18 of G. Priest, R. Routley, and J. Norman (eds.), *Paraconsistent Logic: Essays on the Inconsistent* (Munich: Philosophia Verlag).

[19] Russell, B. (1923), 'Vagueness', *Australasian Journal Philosophy*, 1: 84-92; reprinted as ch. 3 of R. Keefe and P. Smith (eds.), *Vagueness: a Reader* (Cambridge, MA: MIT Press, 1999).

[20] Shapiro, S. (2004), 'Simple Truth, Contradiction, and Consistency', ch. 20 of Priest, *et al.* (2004).

[21] Soames, S. (1999), *Understanding Truth* (Oxford: Oxford University Press).