The Language of Dialetheism

Graham Priest

Contents

1 Introduction

Language is a pervasive phenomenon. It pokes its nose into many issues where one might not expect it to. Dialetheism is a relatively novel phenomenon on the current philosophical scene; its ambit and ramifications are not yet at all clear. However, there is no doubt that language gets in on the act. In this essay I want to look at some of the scenes.

I will start by looking at the role that language plays in dialetheism itself. We will see that it carries a lot of the can. Languages are human creations, and so can be revised. It is therefore natural to ask how the revision of

language bears on dialetheism. That is the topic of the next part of the paper. This will raise the issue of what language can and cannot do, as far as dialetheism is concerned; and that will be the focus of the last part of the paper.

The important relationship between language and dialetheism has not been lost on a number of the philosophers and logicians who have commented on dialetheism. We will take up some of their views along the way. Of these, the most important is Diderik Batens. He, perhaps more than anyone else, has been has been sensitive to the issues. His views will therefore come in for especial scrutiny.

2 Dialetheism, Language, and the World

2.1 Contradiction by Fiat

Let us start with a definition of dialetheism. A dialetheia is a true contradiction, that is, a pair of truth-bearers of the form A and $\neg A$ which are both true.[1] I use the term 'truth-bearer'since people disagree about what sorts of things are the (primary) bearers of truth—sentences, statements, propositions, beliefs. Fortunately, this is a debate we may largely bypass here. Its resoloution is largely irrelevant to most of the considerations in what follows. Whatever truth-bearers are, however, they are language-like. They are expressed in a public language, a language of thought, or whatever. In this way they contrast, crucially, with whatever it is that the truth bearers are *about*. Let us call this, for want of a better name, *the world*.

One thing that partly determines the truth value of a truth-bearer is its constituents: the meanings of the words in the sentence, the concepts involved in the proposition, or whatever. 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 truth-bearer. 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.

Given that dialetheias are sentential/propositional, one natural way for them to arise is simply in virtue of linguistic/conceptual fiat. Thus, suppose we coin a new word (or concept), 'Adult', and stipulate that it is to be used thus:[2]

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

[2] See Priest (2001). In the book reviewed there, Soames argues for the legitimacy of the

- 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, Hilary, who is 17. Then we have:

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

Of course, one can contest the claim that the stipulation succeeds in giving the predicate a sense. Deep issues lurk here, but I will not go into them, since my concern is with other matters. I comment only that the stipulation would seem to be just as successful as stipulative definitions that under-determine truth values, such as that 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,[3] 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 Hilary and Hilary'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:[4]

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

dual under-determined case.

[3]Mares (2004).

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

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

Hilary both has and has not got the right to vote.

Examples that are arguably of the same kind are given by multi-criterial terms.[5] 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 fall may apart, perhaps as the result of surgery of some kind. Thus, suppose that Hilary has female genitalia, but a male chromosomic structure. Then:

Hilary 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 claim that:[6]

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.[7] Again, there is no fiat here.[8] 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. Let us agree, at least for the sake of argument, that the examples we have just reviewed are dialetheias. 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.

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

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

[7]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 $r \in r \wedge r \notin r$.

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

2.3 Contradictions in the World

Some have felt that there may be a more profound sort of dialetheia, a contradiction in the world itself, independent of any linguistic/conceptual considerations. Let us call such dialetheias, following Mares again,[9] *metaphysical dialetheias*.

A major problem here is to see exactly what a metaphysical dialetheia might be. Even someone who supposes that all dialetheias are 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 Hilary. The 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.[10] 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 equivalent.

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 exist 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 existent facts of the form A and $\neg A$, say the fact that Socrates is sitting and 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.[11] Now, negative facts have had a somewhat rocky road in

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

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

[11] This isn't quite right. Facts may not themselves be intrinsically negative, but the *relation* between the facts that A and that $\neg A$ must be intrinsic. But this does not seem to help the friends of negative facts much.

metaphysics, but there are at least certain well-known ways of making sense of the notion, so I will not discuss the matter here.[12]

If one accepts an ontology of facts or fact-like structures, 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.[13] All the hard work here is being done by the metaphysics; dialetheism is playing only an auxilliary role.

3 Linguistic 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, dialetheias are relatively superficial. They can be avoided simply by changing our concepts/language. Compare the corresponding view concerning vagueness. 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 by changing language/concepts sometimes. Thus, consider the legal example concerning Hilary and her rights. If and when a situation of this kind arises, the law would, presumably, be changed to straighten out the conflicting criteria for being able to vote. Note, however, that this is not to deny dialetheism. The language/concepts, as they were, were dialetheic. The point of the change is to find a language that is not dialetheic. 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

[12]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 (2000).

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

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.

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.

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 representation; 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 representation 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 represent everything that the old language represents. Let us then consider the following conjecture:[14]

- Any language (set of concepts), L, that represents the world in a dialetheic way, can be replaced by a consistent language (set of concepts), L', that can represent everything that L represents, but in a consistent way.

The question is still ambiguous, depending how one interprets the modality in question. Are we to suppose that the replacement is a practical possibility, or a merely theoretical one? If the distinction is not clear, just consider

[14]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.

the parallel question, not for inconsistent language, but for vague language. Natural languages contain many terms that are vague. Could we use a language with no vague terms to describe the world with no representational loss? Some have argued that the world itself is vague. What, exactly, to make of this claim is less than clear. However, if, in some sense, it is right, then there is no way of replacing vague terms that describe it with precise ones that are as adequate. So if this is the case, the answer to the theoretical question is 'no'. If, on the other hand, as many people have argued, there is no vaguess in reality, but only in our representations thereof, then we could change our vague language for a precise one preserving, or even improving representational ability. In this case, the answer to the theoretical question is 'yes'.

But even if the answer to this question is 'yes', such a replacement would seem to be entirely impractical. The resulting language is not humanly usuable. 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?
- Would the replacement be possible in practice?

A few things I say will bear on the practical question, 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, α. α means 'something is the case here'. α is always assertable, and consistently so. 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.

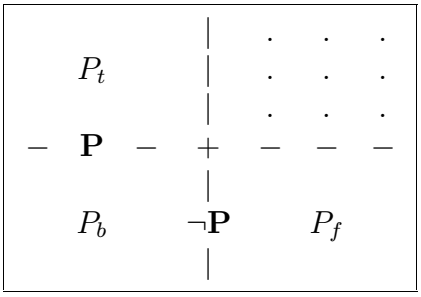
3.2 The Possibility of Revision

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. Hilary 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 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 general. (Just consider the case of 'Adult'.)[15]

We might attempt a more general way of resolving dialetheias as follows. Suppose we have some predicate, P, 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

[15]Even for this simple resolution, there are further problems. Some philosophers of science, such as Bridgeman, wanted to operationalise all our concepts. Were we to do so, our concepts would fragment in exactly the same way. (Think how many operational criteria there are for having a temperature of $5°C$.) A result of this would be that important theoretical inter-connections would thereby be lost. For example, suppose that we endorse the claims that $\forall x(Px \rightarrow Qx)$ and $\forall x(Qx \rightarrow Rx)$. The obvious connection between P and R follows. But if Q is now fragmented into Q_1 and Q_2, and, on examination, these laws now become $\forall x(Px \rightarrow Q_1x)$ and $\forall x(Q_2x \rightarrow Rx)$, this connection is lost. Perhaps all the connections lost in a wholesale fragmentation can be reinstituted somehow, but only at the cost of a major increase in complexity.

each case, is simply the complement. The situation may be depicted in the following diagram.



The left-hand side is the extension of P. The bottom half is the co-extension of P. The top right quadrant comprises those things of which P is neither true nor false, and for present purposes we may take this to be empty. The three new predicates have as extension each of 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, $P_t x \vee P_b x$.[16]

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, 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 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 above diagram, whilst $\neg(P_t x \vee P_b x)$ is true in the bottom right quadrant. 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:

[16]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.

$(\neg(A \vee B))^+$ 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. 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. 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 can have $\mathfrak{B}A$ without having $\mathfrak{B}A^+$. The trouble is that belief and its kind are intentional states, directed towards propositions in a certain language. That language seems to be integral to the intentional state in question, and cannot be eliminated.[17]

Nor is this just a problem about intentional 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 suppose that all creatures with hearts are, as a matter of fact, creatures with kidneys. 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

[17]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 ourselve 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 practical 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 $\forall p(\mathfrak{B}p \rightarrow p)$.

a blood circulation system. The information is extensionally equivalent to the information that a_1, ..., a_n are creatures of kind k with a kidneys. This does not confirm the claim that all creatures of kind k have a blood circulation system.[18]

3.3 Expressive Loss

But worse is yet to come for the conjecture that we can always theoretically 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. This is characterised by the schema:

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

The naive notion of set membership is inconsistant, as we have already had occasion to note. 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:

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.[19] Why is this not in conflict with the discussion of the last section? The reason is essentially that the policy 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. We are guaranteed, 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 such a schema effectively

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

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

characterizes the naive concept of set membership. So if we go the latter way, even if every sentence of the old language has an equivalent in the new, there is still an expressive loss. We have lost a concept which we had before. We have lost the ability to express arbitrary set formation. We have purchased consistency only at the expense of expressive impoverishment. Not everything that could be expressed before can still be expressed.

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, this would be no loss; but we are not. It is granted that we already had a functioning language/set of concepts. The concepts therefore had a use, and so were useful.

Indeed, they may be *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 structures. It plays an essential role in certain kinds of mathematics (such as category theory), in our ruminations about the way that language and other conceptual processes work. But abilities of this kind drive us into contradictions of the kind involved in discussions of the limits of thought.[20] We could throw away the ability to totalise in this way. Maybe this would restore consistency, but the cost would be 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*.

Actually, it is not even the case that one *can* do this. If we have the conceptual ability to totalise, in what sense can this be given up? One can refuse to exercise the ability, but this would seem to get us nowhere. (It would be like solving the liar paradox as follows. A: 'Suppose I say that I am lying'. B: 'Don't.') If you have the ability to think certain thoughts, you cannot, it would seem, lose this without some kind of trauma to the brain, caused by accident or senility. And if this is the case, the recommendation to change our language/concepts fails the practicality test in this most fundamental way.

[20]See Priest (1995).

4 The Powers of Language

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? Sometimes there might be good reason. One of the functions of law is to guide action. Contradictory laws may frustrate this purpose – should we or should we not allow Hilary to vote? But as far as the purely representational 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.[21]

However, Batens (1999), (2002) has argued that it is sound methodology to replace an inconsistent language with a consistent one if we can. 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 explains it as follows:[22]

> 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 $\neg P$, 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 and third set, not in the second 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.

Calling this precision is, I think, a little misleading. The truth conditions for the operators involved are as precise and accurate in the paraconsistent case as in the classical case. The complaint is, rather, one concerning expressability: the paraconsistent language has no way to express the thought

[21] On these matters, see, further, Priest (1987), 13.6.

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

that Pa is consistently true or false.[23] Now, two questions here: Is this true? Is the ability to do such a thing a virtue?

Take the questions in reverse order. Batens does not explain why he takes it to be a virtue, but he cites with approval Popper's *Logic of Scientific Discovery*,[24] which suggests that what is at issue here is refutability. Suppose that our theory is a paraconsistent one, and that it predicts $\neg Pa$. Suppose that Pa is observed. If we have no way of expressing the thought that Pa is consistent, we are not forced to jettison the theory, but may simply accept the resultant contradiction.

Now, as methodologists such as Lakatos have pointed out, even if the logic of the theory is classical, the proponent of the theory is not at all forced to jettison it when a contradiction rears its head. They may blame the contradiction on some auxiliary hypothesis; or they may reject the observation, since observation is soft and theory-laden. Refutation, in the only sense in which one may speak of it, is a long and drawn out process. We have to weigh up all the possibilities and, if we reject the theory, we do so because some rival appears to be better according to some methodological criteria. The ability to accept Pa and $\neg Pa$ adds nothing to this picture. We have an extra option, it is true—accept the inconsistent theory—but a final decision on what to do still depends on the same methodological considerations.[25] So, if refutability is a virtue, the ability to accept an inconsistent theory does nothing to threaten this.

Let us turn to the first question: is it the case that in a paraconsistent logic one cannot express the claim that something is consistent? Let us 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 a paraconsistent logic cannot express the thought that something is consistent? Well, $\neg(Pa \wedge \neg Pa)$ won't do, since this is a logical truth (in a 3-valued paraconsistent logic). But as long as there is a truth predicate, T, in the language, we can express this

[23] What Batens actually says that there is no way of expressing the fact that an object is in the union of the first and third sets. As he enumerates them, this is just the set of objects that have the property P, and can therefore be expressed by that predicate. He means the union of the first and second, I take it.

[24] Batens (1999), p. 272, fn. 22.

[25] The matter is discussed at length in Priest (2006), Part 3.

by $\neg T\langle Pa \wedge \neg Pa\rangle$. Batens will of course point out at this juncture that if the negation used in this expression is paraconsistent, it might, for all this says, be the case that $T\langle Pa \wedge \neg Pa\rangle$ as well. So though one may be able to express the thought that Pa is consistent, one cannot do so in a way that guarantees consistency (as would be required for a knock-down refutation, assuming such to be possible).

But neither can a classical logician. They can assert $\neg(Pa \wedge \neg Pa)$ where $\neg$ is, or is taken to be, classical negation. But this does prevent them endorsing $Pa \wedge \neg Pa$ as well. Of course, 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*) in Batens' words.) But if enforcing collapse into triviality is what we require, then a paraconsistent logician can do exactly the same thing. They can endorse $T\langle Pa \wedge \neg Pa\rangle \rightarrow \bot$ or just $(Pa \wedge \neg Pa) \rightarrow \bot$, where $\rightarrow$ is a detachable conditional, and $\bot$ is a logical constant that entails everything.

We see that the supposed virtue of consistent theories over inconsistent ones has evaporated.

4.2 Boolean Negation—Again

But there is a more profound, and much harder point here. The assumption that we can replace an inconsistent language with one with an explosive negation, such as classical (Boolean) negation, presupposes that such a notion makes sense. If a notion is a nonsense, so is the suggestion that we replace a theory with one using it. It seems to me that Boolean negation does not make sense.

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 of it—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, 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.[26]

Perhaps, more persuasively, it might be argued that classical negation is meaningful since we use it perfectly sensibly in reasoning about consistent

[26]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. (See Priest (2006), ch. 5.) Hence a classical logician must deny the meaningfulness of the latter notion, which seems just as bad, if not worse.

situations or within consistent theories. But this point is not clear either. For in consistent interpretations, classical and paraconsistent negation behave in exactly the same way. Moreover, using an adaptive paraconsistent logic, a dialetheist may also reason in exactly the same way as a classical logician in a consistent theory.[27] Whilst one remains in a consistent context, then, there is no way one can tell whether one is employing classical negation or a paraconsistent negation.[28]

Why should we suppose that classical negation does not make sense, however? In a nutshell, the argument goes as follows.[29] How are we supposed to characterise Boolean negation? We may do so proof-theoretically: it is the notion that satisfies certain rules of inference. But the mere specification of such a set is not guaranteed to characterise a meaningful notion. We have Arthur Prior and *ton*k to remind us of this fact. And just like *tonk*, a notion characterised with the rules of classical negation effects a collapse into triviality (in the context of self-reference, etc.). It would seem, then, to be of the same kind.

Alternatively, we may characterise the notion by giving its truth conditions in standard ways. Say, $\neg A$ is true (at a world of an interpretation) if A is not true (there). This may determine a perfectly legitimate notion, but to establish that it has the proof-theoretic properties of Boolean negation, we need to do more than state the 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 generate a notion with respect to which the rules of proof for Boolean negation are sound, unless we reason classically 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 (2002), Section 5. He proceeds by giving the truth conditions of negation in the classical way in a metatheory that he says is classical. As is clear from the preceding summary, this is simply question-begging. A paraconsistent logician will insist that the metatheory be paraconsistent. Against this, Batens raises four considerations.

1. Negation may not actually be needed to give the truth conditions of negation. Thus, assume for the sake of illustration that we have a three-valued semantics with the truth values $\{t\}$, $\{t, f\}$, and $\{f\}$, the first two being designated. We may say that the value of $\neg A$ is $\{t\}$ if the value of A

[27] See Priest (1987), 8.6, and especially ch. 16 of the second edition.

[28] This does not mean, contra Batens (2002), p. 140, that paraconsistent negation and classical negation mean the same thing. Indeed, classical negation is not a meaningful notion at all.

[29] See Priest (2006), esp. ch. 6.

is $\{f\}$ and it is $\{f\}$ if the value of A it is $\{t\}$ or $\{t, f\}$. Indeed we may; but it remains the case that we have to reason using negation to show that $\neg$ has its classical properties. 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 then whenever the premises are designated, so is the conclusion. This is an argument of the form $\neg\alpha \vdash \alpha \rightarrow \beta$, which a paraconsistent logician is not going to accept, if $\rightarrow$ is a detachable conditional.

It might fairly be pointed out at this point that I have claimed that in consistent contexts, paraconsistent reasoning and classical reasoning are indistinguishable. So if the argument goes through classically it goes through paraconsistently. The rub here is that we are not in a consistent context. Metatheory is couched in set-theory, and a paraconsistent set theory is, notoriously, inconsistent.

2. This leads to Batens' second point. There is currently no extant metatheory for paraconsistent logics based on a paraconsistent set theory. This is a fair point. How best to turn the trick is still moot, but one way of doing it is explained in the second edition of Priest (1987), ch. 18. The details are too complex to reproduce here, but the idea is that a certain understanding of paraconsistent set theory allows a paraconsistent logician simply to appropriate certain classical metatheoretic arguments.

3. Batens' third point is to suggest defining a classical-style notion of negation by the following truth conditions:

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

Such truth conditions have exactly the same problem as the others. Namely, we cannot use these truth conditions to establish that $\neg$ satisfies inferences such as EFQ without reasoning classically.

4. Batens' final point is to raise the issue that we may be able, in the context of the above semantics, to construct an extended paradox which establishes that $t = f$; hence things collapse into triviality. He does not himself give such an argument, but arguments of this kind have certainly been attempted by others. As far as I can see, none of them works. (The matter is discussed at length in the second edition of Priest (1987), 20.3.)

4.3 Further Objections

Batens goes on ((2002), pp. 142-4) to cite further problems for dialetheism. In fact, these are mis-targeted; for he himself is a dialetheist. He certainly holds that there are inconsistent languages, and so dialetheias. What these

objections are objections to is a global dispensing with Boolean negation. Since I have dealt with most of the issues at length elsewhere, the discussion here may be relatively brief.

Batens' first point is a familiar objection. Without classical negation there is no way to express the thought that two claims, A and B, are incompatible, or that you can't have them both—'with the required force'. Of course, there are many ways to express this fact: Batens himself gives several, including $\neg\Diamond(A \wedge B)$. But even if $\neg\Diamond(A \wedge B)$ is true, as Batens' would point out, A and B *may* still both hold. That is why he uses the phrase 'with the required force'. But what is the required force and who requires is? The required force would seem to be that the notion behave consistently; and the person who is supposed to require it is me. No doubt I have used words of this kind in my writing. What needs to be demonstrated, however, is that I require the words to behave consistently—something that Batens does not attempt. Nor is this an easy matter. In many cases where words of this kind are used, they can be understood not as assertions at all, but as denials—where denial is a *sui generis* speech act, quite distinct from asserting a negation.[30] The best arguments I know as to why one needs to be able to express an appropriate notion of incompatibility consistently are given by Shapiro (2004), and I have explained why they do not work in the second edition of Priest (1987), 20.4.

Batens' second objection appeals to his notion of adaptive logic. In general, reasoning is adaptive. In particular, if we have a theory which is consistent, the adaptive logic gives the full force of classical logic. If an inconsistency in the theory turns up, the adaptive logic does not deliver classical logic, but a logic midway between classical logic and a base paraconsistent logic. This is Batens' view, and he claims that it is unavailable to a dialetheist (that is, someone who does not accept the legitimacy of classical negation). I find this objection rather puzzling. As I noted in the last section, this view is exactly my view! Indeed, I thought Batens' idea such a good one that I adopted it as my own.[31]

The third objection is that dialetheists are able to reason in accord with classical and intuitionist logic, understand classical proofs, etc. Hence (I assume to be Batens' point), the notions of such logic must be meaningful. However, one can explain how dialetheists are able do this in a perfectly natural way. How they can reason in accord with classical logic is entirely obvious. As I have already observed, classical and paraconsistent reasoning

[30]See Priest (2006), ch. 6.

[31]In Batens' next paragraph, the objection seems to get mixed up with the claim that paraconsistency makes theories unrefutable. I have dealt with that matter above.

come to the same thing in consistent situations. Reasoning in the one way *is* reasoning in the other. The case with intuitionist logic (or with classical logic in inconsistent situations) is different. This is not a special case of dialetheic logic (at least, not without a lot more fast footwork). Given *any* set of rules that are not too complex, be they those of intuitionist logic or anything else, 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 in certain—notably, inconsistent—contexts.

Batens' final objection is that without classical negation, and so the classical material conditional, we have no adequate understanding of 'All As are Bs', assuming this to be of the form $\forall x(Ax \to Bx)$. A paraconsistent material conditional does not detach, so if $\to$ here is a material conditional, the inference:

> x is an A;
> all As are Bs;
> hence x is a B

is invalid. And the $\to$ cannot be a relevant conditional since the categorical form may be true even if there is no connection of relevance between being an A and being a B. The point is well made. It would seem, however, that an appropriate enthymematic conditional will do the job. Choosing this wisely will give us a conditional which allows for the preservation of most of the sensible properties of the categorical forms.[32]

5 Conclusion

We have seen the extent to which language is implicated in dialetheism, and to what extent dialetheism may transcend language. Assuming that it does not, we have also seen what (limited) scope linguistic revision has for eliminating dialetheism from our conceptual schemes. Finally, we have seen how dialetheism bears on the issue of what language itself can and cannot do.

Language is a pervasive phenomenon. It pokes its nose into many issues where one might not expect it to. Dialetheism is a relatively novel phenomenon on the current philosophical scene; its ambit and ramifications are not yet at all clear. However, there is no doubt that language gets in

[32]See the second edition of Priest (1987), 18.3, where the connection between the categorical form and subsethood is also pursued. For a more general discussion of the shape of the categorical forms in the context of a relevant logic, see Beall *et al*, (2006).

on the act. As we have now seen, though, it would be wrong to suppose that language plays a leading role; it is just one of the supporting cast.

References

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

[2] 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.

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

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

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

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

[7] Priest, G. (1987), *In Contradiction*, Dordrecht: Kluwer Academic Publishers; second edition, Oxford: Oxford University Press, 2006.

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

[9] Priest, G. (2000), 'Truth and Contradiction', *Philosophical Quarterly* 50, 189-95; reprinted as ch. 2 of (2006).

[10] Priest, G. (2001), Review of S. Soames, *Understanding Truth*, *British Journal for the Philosophy of Science* 52, 211-5.

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

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

[13] 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.

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