Paradox and Paraconsistency

Graham Priest

June 26, 2021

Departments of Philosophy, the CUNY Graduate Center, the University of Melbourne, and the Ruhr University of Bochum

1 Introduction

Paradoxes can amuse, frustrate, puzzle, fascinate—but above all, teach us things. They take us out of our intellectual comfort zone, and show us that our thinking needs to be rejigged in some way or other. Hence it is that paradoxes have played a significant role in advances in science and philosophy. Unsurprisingly, the result can be surprising.

One surprise to have arisen in the last forty years or so is that we may need to revise out thinking about one of the oldest, most established and entrenched principles in Western logic: the Principle of Non-Contradiction.

As one might expect, discussions involved in matters soon become contentious, complex, and, because they involve issues in modern logic, technical. However, in what follows, I will try to explain the whys and hows of the matter in as simple and non-technical terms as I can.

Further reading relevant to each section, where discussion of and references to all the matters discussed there, can be found at the end of the essay. (References cited in connection with some sections may contain material relevant to other sections too.)

2 What is a Paradox?

Let us start by getting clear on what a paradox (*para/doxa*—beyond belief) is. The word is commonly used to refer to many sorts of thing, but for our purposes, we may understand it as follows.

A paradox is an argument where the premises seem to be true, the argument seems to be valid (that is the conclusions drawn really seem to follow), but yet the final conclusion of the argument seems not to be true. As is clear, given a paradox, there are only three responses:

[1] Some of the premises are not, after all, true.

[2] Some of the inferences involved are not, after all, valid.

[3] The final conclusion is, after all, true.

In any of these cases, some of our seemings turn out not to be so, and we need to know how and why not—and preferably why we were mislead into supposing otherwise. Responses of all three kinds are well known. Let us work backwards.

[3] Consider the standard counting numbers, $0, 1, 2, ...$ (natural numbers as they are called). These can be put into a one to one correspondence with the even numbers in an obvious way:

$$\begin{array}{ccc} 0 & \rightarrow & 0 \\ 1 & \rightarrow & 2 \\ 2 & \rightarrow & 4 \\ 3 & \rightarrow & 6 \\ \vdots & & \vdots \end{array}$$

This appears to show that there are just as many even numbers as numbers, which cannot be correct since all the odd ones have been thrown away. The paradox has been known since the Middle Ages, and was taken to show that the notion of infinity is incoherent. This changed with the work of the mathematician Georg Cantor in the late 19th century. The conclusion came to be accepted as true. There are exactly as many even numbers as numbers. Indeed, the phenomenon became a definition of the infinite. A set is infinite, just if you can throw away some of its members and still have a set of the same size. The reason we were wont to reject the conclusion, it is natural to suppose, is that our intuitions about size are drawn from the familiar finite, and no longer hold for the infinite.

[2] The distinction between [2] and [1] is, to a certain extent, flexible, since an inference from A to B ($A \vdash B$) may be replaced by a conditional premise of the form 'if A then B' ($A \to B$) with an application of *modus ponens* ($A, A \to B \vdash B$). However, the following is a paradox which is naturally seen as of type [2]. This is called Simpson's paradox. Suppose there is a group of men and women, all of whom have a certain illness. The men are more likely to recover if they receive a certain drug. The women are more likely to recover if they receive the drug. It seems to follow that all people in the group are more likely to recover if they receive the drug. However, this is an invalid inference, as the following statistics show:

Men		Recover		
		Y	N	Success Rate
Treatment	Y	8	5	$8/13 \simeq 62\%$
	N	4	3	$4/7 \simeq 57\%$

Women		Recover		
		Y	N	Success Rate
Treatment	Y	12	15	$12/25 \simeq 48\%$
	N	2	3	$2/5 \simeq 40\%$

Total		Recover		
		Y	N	Success Rate
Treatment	Y	20	20	$20/40 \simeq 50\%$
	N	6	6	$6/12 \simeq 50\%$

The data shows that the inference is invalid. Why we are inclined to believe otherwise is still a matter of debate.

[1] For the last group, consider one of Zeno's paradoxes, the Dichotomy. An object is moving from x to y with constant velocity. Let the distance to be covered be d. Before the object gets to y it must cover a distance of $d/2$. Then, before it gets to y, it must cover further distance of $d/4$. Then, to get to y, it must cover further distance of $d/8$, etc. In other words, it must do an infinite number of things. But one can't do an infinite number of things in a finite time, so the object can never get to y. The solution is that one *can* do an infinite number of things in a finite time, if the time required for

each gets smaller and smaller (specifically, if the infinite sum of the times converges to a limit). We are inclined to think otherwise since we assume that each of the things takes at least some minimum time.

3 Paradoxes of Self-Reference

Let us now turn to the paradoxes of self-reference. These form a family of paradoxes which have played a central role in the history of logic—though how, exactly, to characterise self-reference is a thorny issue. One thing that is distinctive about such paradoxes is that they are arguments which conclude in a contradiction.

The oldest, most famous, and simplest of these is the Liar paradox, invented by the Ancient Greek paradoxer, Eubulides (fl. 4c, BCE). To fix some terminology, let us take *false* to mean *has a true negation*. The argument involved uses a principle called the 'T-Schema', which we may think of as the inference from A to 'A' is true, and vice versa—an apparently anodyne principle, which no one would ever have doubted but for its involvement in paradox.

The Liar paradox concerns a sentence which says of itself that it is false. Let us refer to this sentence as L. Then L is of the form:

- L is false

Is L true or false? If it is true then, by the T-Schema, it is false. If it is false, then, by the T-Schema in the other direction, it is true. Either way, it is true and false.

The solutions that have been offered to the paradox over the last couple of millennia are legion. Since—it has been assumed—no contradiction can be true, this has to be a paradox of kinds [1] or [2]. It is a mark of how unsuccessful such solutions have been that there is still no consensus on the matter, over two thousand years after Eubulides. To see the sort of problems that arise, let us consider one of the most popular solutions to have been offered. As a little thought shows, the above argument assumes that L is either true or false. The claim that every sentence is either true or false is called the Principle of Excluded Middle (PEM). The paradox may therefore be solved by rejecting the PEM. Some sentences—notably L—are neither true nor false.

A major problem with this kind of approach is that the paradox can be reformulated to take the proffered solution into account. Let L' be sentence:

- L' is either false or (neither true or false).

If it is true, it is either false or neither true nor false. If it is false it is *either* false *or* neither true or false, so it is true. And now suppose that it is neither true nor false. Then again, it is *either* false *or* neither true nor false, so it is true. Any way you look at it, we have a contradiction.

Paradoxes of this kind are sometimes called *Revenge Paradoxes*, and they plague consistent solutions. The paradox is generated by a division of the sentences into the *bona fide* truths and the Rest—a set that contains all and only the others. How to characterise the Rest may depend on whatever machinery is available: false, false or neither true nor false, false or meaningless. It doesn't matter. We may simply formulate a sentence L'' of the form:

- L'' is in the Rest

If it is true, it is in the Rest; and if it is in the Rest, it is true. The self-referential construction is simply a mechanism that tears through any boundary of this kind.

In fact, there are paradoxes in the same family which do not use this PEM at all. One of these is a paradox discovered in the 20th Century, called after its discoverer, Berry. Consider the counting numbers again. There is an infinitude of these. They can be referred to by noun phrases of, say, English. In principle, we can make these noun phrases as long as we like, and so refer to every number. However, once the available vocabulary is fixed (say, to that of my idiolect now), there is only a finite number of noun phrases of any given length—say less than 100 words. Hence, there must be numbers (an infinite number of such) that cannot be referred to by phrases of less than 100 words. Given any set of numbers, one of them must be the least. So consider the phrase:

- the least number that cannot be referred to with a noun phrase of less than 100 words.

Let this number be n. Then by construction, n cannot be referred to by a noun phrase of less than 100 words. But the displayed noun-phrase clearly has less than 100 words, and it refers to n. So n both can and cannot be referred to in this way. There is no use of the PEM in this argument.

Considerations of this kind suggest that we are just barking up the wrong tree. The Liar is not a paradox of kinds [1] or [2], but of kind [3]. It shows us that some contradictions are true. L is both true and false. Some have averred that such an account of the paradox faces a revenge paradox. The Rest in this case now comprises those things that are false and not both true and false. Come back to the sentence L''. As before, if L'' is true it is in the Rest; and if it is in the Rest it is true. So it is in both truths and the Rest. That's a contradiction.

The supposed objection is clearly short-sighted. The whole point of a type [3] solution to the paradox is *not* to get rid of contradiction, but to accept it and cognate contradictions. As we saw, self-reference is precisely a mechanism which tears through any boundary of this kind. So contradiction is exactly what one should expect in this case.

Indeed, once the machinery of an appropriate underlying logic is in place (on which, more in a moment), one may show that there are formal theories which include the T-Schema and the Liar contradiction, but for which one may prove that the contradictions delivered are very limited. Specifically, only some sentences involving the truth predicate are contradictory. The details of the proof are more complex than I can go into here, however.

And, finally, why, did we assume that the contradictory conclusion could not be true? Well, one shouldn't underestimate the power or orthodoxy. The PNC really has been very firmly so for a long time. But one may be a bit more generous. In the case of our paradox of the infinite, it was natural to take the conclusion not to be true since our intuitions are drawn from normal situations, the finite; and the usual facts about size hold there. Similarly, one may hold that our intuitions about the PNC are draw from normal situations, where things seem consistent. Paradoxical sentences like the Liar are rare birds. In other words, the intuition depends on what Wittgenstein called in the *Philosophical Investigations* (§593) a one-sided diet of examples.

4 Infinity

Another class of paradoxes ending in contradiction (or maybe part of the same class, depending on how one reckons this) are paradoxes of infinity. As we noted in Section 2, our thinking about infinity was revolutionized by Cantor, and the older paradoxes then disappeared. However, in the process, new, more subtle and sophisticated paradoxes appeared. Perhaps the most famous of these is Russell's paradox, but let us look at one where the connection with infinity is much more obvious: Burali-Forti's paradox.

In modern set theory the natural numbers, finite ordinals as they are called in this context, are just the beginning. There is one coming immediately after all them, the first transfinite ordinal, ω:

- $0, 1, 2, 3, \ldots \omega$

Then there is a next, and a next...:

- $0, 1, 2, 3, \ldots \omega, \omega + 1, \omega + 2, \omega + 3, \ldots$

Then $\omega + \omega$, that is 2ω, after which we go on the same way:

- $0, 1, 2, 3, \ldots \omega, \omega + 1, \omega + 2, \omega + 3, \ldots 2\omega, 2\omega + 1, 2\omega + 2, \ldots$

And so on with 3ω,4ω,5ω..., till $\omega.\omega$, that is ω^2:

- $0, 1, 2, 3, \ldots \omega, \omega + 1, \omega + 2, \omega + 3, \ldots 2\omega \ldots 3\omega \ldots \omega^2$

And so on again with ω^3,ω^4, ... ω^ω.

- $0, 1, 2, 3, \ldots \omega, \omega + 1, \omega + 2, \omega + 3, \ldots 2\omega \ldots 3\omega \ldots \omega^2 \ldots \omega^3 \ldots \omega^\omega$

And these are just the beginning of the sequence: it never terminates. After any sequence of ordinals, whether there is a last member or not, there is a least ordinal greater than all of them.

The transfinite ordinals mentioned so far are, in fact, very small ones: there are much, much, larger ones. How far the sequence of transfinite ordinals goes on—indeed, even how to phrase this thought in a sensible way—is a moot point, both philosophically and mathematically, but the whole sequence may be said to be *absolutely* infinite.

There are many ways in which the ordinals can be defined, but the simplest is to use a construction due to von Neumann. Each ordinal is thought of as the set of all prior ordinals. Since there are no ordinals less than 0, 0 is the empty set, $\emptyset$. $1 = \{0\}$, $2 = \{0, 1\}$, $\omega = \{0, 1, 2, \ldots\}$, etc. On this construction, given any sequence of ordinals, the least ordinal greater than all of them is just the set of them.

Now Burali-Forti's paradox: Consider the set of all ordinals, On. Every ordinal is in this set, so there cannot be an ordinal greater than all of them. But On is the least ordinal greater than all of these. So there is a greater.

Unlike the case with the Liar paradox and its ilk, there is a standard solution to the Burali-Forti paradox and similar paradoxes of absolute infinity. This is enshrined in Zermelo Frankel set theory (ZF). According to this, there is no set of all ordinals: On does not exist; it is too "large". In other words, this is a paradox of type [1].

The solution is not without its problems however. It is highly counter-intuitive. We can talk of all ordinals, and make claims about them that are determinately true or false. So there appears to be a determinate totality over which our quantifiers range. Worse, there appear to be mathematical constructions which presume that large sets of this kind do exist. (For example, in category theory there is a category of all sets.) These problems are well known, well discussed, and too technical to go into here.

We can, however, treat it as a paradox of type [3]. On exists; there is no ordinal greater than all its members, yet On is an ordinal greater than all its members. To fill out such a solution, it is necessary to have an inconsistent theory of sets which does the job required. But unlike the case with the theories of truth, there is (as yet) no standard way of doing this. The problem is that set theory plays many important roles in mathematics other than theorising the infinite. (In fact, the theory of the infinite is of no particular interest to most mathematicians.) And the theory must be able to perform all the other tasks required of a theory of sets. It is not so easy to construct a set-theory which accommodates the contradiction about On, and does all the other things required of a set-theory. How best to do this is still a matter of debate and investigation; and the details are too technical to go into here.

However, assuming that a type [3] solution is correct, why did we take the conclusion of the paradox to be untrue? For exactly the same reason as with the semantic paradoxes of self-reference: our intuitions are drawn from more mundane matters, where things seem to be consistent. Dealing with transfinite situations is not a commonplace.

5 Dialetheism

What we have so far seen is that one may well take certain paradoxes that end in contradiction to be type [3] paradoxes. And why not? The first and most obvious answer concerns the Principle of Non-Contradiction (PNC): no contradiction can be true.

To discuss this, first some terminology. A *dialetheia*—two (way) truth—is a true contradiction. *Dialetheism* is the view that some contradictions are dialetheias. The terms were coined about 1980 by Richard Routley (Sylvan, as he later became) and myself to distinguish the view from paraconsistency (more of which in a moment). The neologisms were motivated by Remark 59 of Wittgenstein's *Remarks on the Foundations of Mathematic*s, where he likens the Liar sentence to a Janus-headed creature facing both truth and falsity.

Now, taking a paradox which ends in a contradiction to be a type [3] paradox requires dialetheism; dialetheism clearly flies in the face of the PNC; and the PNC has been highly orthodox in the history of Western philosophy, with only a few dissenting voices (such as that of Hegel).

The PNC was set into orthodoxy by Aristotle in his *Metaphysics*—so much so, that there is hardly a defence of it of any substance since. Given that, one might have expected the defence to be decisive. This is hardly the case. The relevant passage contains one long argument and six or seven very swift ones (depending how you break up the text). The long argument is tangled and convoluted, so much so that modern commentators cannot agree on *how* the argument is supposed to work, let alone *that* it works. However, the core of the argument seems to be as follows.

Take some predicate. Aristotle uses *man*, but this is meant to be typical. *Man* can mean several things. (We do not have a dialetheia simply because Hypatia was a man (person) and not a man (female).) Fix on one meaning. Aristotle chooses *two-footed animal*. Then necessarily, if something is a man it is a two-footed animal. Hence it is impossible for something to be a man and not a two-footed animal. Hence it is impossible for something to be a man and not a man (for to be a man is to be a two-footed animal). Using the symbolism of modal logic, what, then, has been established is that $\neg\Diamond(Ma \wedge \neg Ma)$.

Given that this generalises, the point seems to have been established. It has not. For if one has $\neg\Diamond(Ma \wedge \neg Ma)$, one has $\Box\neg(Ma \wedge \neg Ma)$, and so certainly $\neg(Ma \wedge \neg Ma)$. But one may still have $Ma \wedge \neg Ma$! If one does, then one has $(Ma \wedge \neg Ma) \wedge \neg(Ma \wedge \neg Ma)$. We might call this a secondary contradiction; and one cannot rule that out without begging the question, since the argument was supposed to show that no contradictions are acceptable.

The bevy of short arguments are even worse than the long argument, since they are clearly beside the point. What they establish—if they establish anything—is that it is not the case that *all* contradictions are true. That, of course, is quite compatible with some contradictions being true. (Why Aristotle makes the slide from *not any* to *not all* has no explanation in the text.) Indeed some of the arguments are even worse, since they establish (if anything) only that no one can *believe* that all contradictions are true—which is even weaker.

In fact, Aristotle's arguments against dialetheism were already demolished by the Polish logician Jan Łukasiewicz in his book of 1910. (Though why it took over 2000 years for this last bastion of Aristotelianism to be subjected to such acute scrutiny is an interesting question, which I will not pursue here.)

6 Paraconsistency

Unsurprisingly, then, if one asks a contemporary logician why no contradiction is true, they are likely to give a quite different answer.

First, some terminology again. The principle of inference that everything follows from a contradiction ($A, \neg A \vdash B$, for all A and B) is called *Explosion*—or to give it its Medieval name, *ex falso quodlibet sequitur* (from a falsehood follows anything you like). A consequence relation in which Explosion is not valid is called *paraconsistent*—beyond (*para*) the consistent. (The term was coined by Peruvian philosopher Miró Quesada in 1975 in response to work by the Brazilian logician, Newton da Costa.)

Dialetheism and paraconsistency are quite different. (One might think that a correct consequence relation should be paraconsistent for reasons having nothing to do with dialetheism.) Indeed, Aristotle's system of inference, syllogistic, is paraconsistent, as he himself points out the the *Analytics*. Awareness of the principle of Explosion appears to come into Western logic in the 12th Century, with the work of William of Soissons. But it becomes securely entrenched in logic only around the turn of the 20th Century, with the work of Frege and Russell.

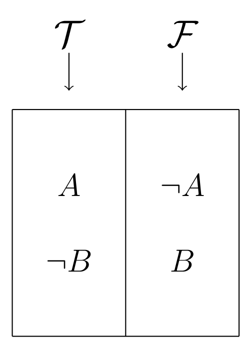
That the principle is presently orthodox, may come as a surprise to those who have never studied much logic. Consider the inference: Melbourne is in Australia and Melbourne is not in Australia. So Caesar died in 44 BCE. This hardly seems to be valid. The premise has nothing whatsoever to do with the conclusion. Notwithstanding, Explosion *is* currently orthodox, and it is the reason most likely to be given against dialetheism by modern logicians.

If Explosion is valid and dialetheism is true, then everything is true. And that is too much.

Let us see why it is held to be valid. Take a simple language for propositional logic, with connectives $\neg$,$\vee$, and $\wedge$. ($A \supset B$ may be defined as $\neg A \vee B$.) An interpretation of the language divides the sentences up into those that are true in the interpretation, $\mathcal{T}$, and those that are false in the interpretation, $\mathcal{F}$. The two sets are exclusive and exhaustive. The truth and falsity conditions for the connectives (as enshrined in the familiar truth tables) are:

- $\neg A$ is in $\mathcal{T}$ iff A is in $\mathcal{F}$
- $\neg A$ is in $\mathcal{F}$ iff A is in $\mathcal{T}$
- $A \wedge B$ is in $\mathcal{T}$ iff A is in $\mathcal{T}$ and B is in $\mathcal{T}$
- $A \wedge B$ is in $\mathcal{F}$ iff A is in $\mathcal{F}$ or B is in $\mathcal{F}$
- $A \vee B$ is in $\mathcal{T}$ iff A is in $\mathcal{T}$ or B is in $\mathcal{T}$
- $A \vee B$ is in $\mathcal{F}$ iff A is in $\mathcal{F}$ and B is in $\mathcal{F}$

Normally, one would not bother to give the second of each pair, since it follows from the first, given that $\mathcal{T}$ and $\mathcal{F}$ are exclusive and exhaustive; but bear with me. Negation, then, flips a sentence from $\mathcal{T}$ to $\mathcal{F}$, and vice versa. We may depict this in the following diagram:

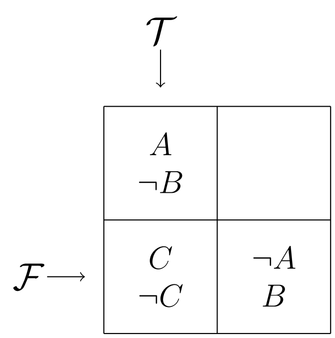


An inference is *invalid* iff there is an interpretation in which all the premises are in $\mathcal{T}$ and the conclusion is not in $\mathcal{T}$. It is valid otherwise.

Clearly, for any sentence, A, there is no interpretation in which A and $\neg A$ are both in $\mathcal{T}$. Hence, for any B one likes, there is no interpretation where A and $\neg A$ are in $\mathcal{T}$ and B is not. Hence Explosion is valid—vacuously, one might say.

Let us turn to paraconsistent logic. Starting about the 1960s many of these were developed. There are now many paraconsistent logics, which operate in quite different ways. However, let me describe one of the simplest and most natural. This is called *First Degree Entailment* (FDE)—don't ask.

FDE is *exactly* the same as classical logic, except that we do not insist that in an interpretation $\mathcal{T}$ and $\mathcal{F}$ are exclusive and exhaustive. In some interpretations they are. But in some they overlap; in some they underlap; in some they do both. So the situation concerning negation now looks like this:



In the bottom left quadrant, things are true and false. In the top lright quadrant, things are neither true nor false. If, in an interpretation, C is in the bottom left quadrant, it is true and false, so its negation is false and true, that is, in the bottom left quadrant as well. Beware: one must now distinguish clearly between being false (bottom half) and not being true (right hand half).

Now, consider the inference $C, \neg C \vdash B$. In the interpretation depicted, the premises are true (and false as well, but that is irrelevant), and B is not. So Explosion is invalid.

To return to our objection: When classical logic is explained, there is virtually never any attempt to justify the claim that truth and falsity must always be exclusive and exhaustive. It is simply assumed. What we have seen is that if truth and falsity may overlap, Explosion is invalid. So to appeal to

Explosion in an argument against dialetheism is to beg the question entirely. Someone who holds that there are dialetheias holds that truth and falsity may overlap, and, in virtue of that, reject Explosion.

Let me end this section by noting a few other matters. If, in an interpretation $C \wedge \neg C$ is in $\mathcal{T}$, then so is $(C \wedge \neg C) \wedge \neg(C \wedge \neg C)$. This is exactly a secondary contradiction of a kind that arose in the discussion of Aristotle's argument.

Secondly, if we define $\supset$ in the usual way, then it does not satisfy *modus ponens*, $C, \neg C \vee B \vdash B$, as the interpretation in the diagram shows. In virtue of this, the language should be augmented with a new conditional connective, $\rightarrow$, which does. (In truth, the material condition was always a hopeless account of a genuine conditional.) This can certainly be done, though we do not need to go into the details here.

Finally, the same diagram shows that the conjunctive form of *modus ponens* for the material conditional does not hold either. The inference $C, \neg(C \wedge \neg B) \vdash B$ is not valid.

7 The Sorites

So far, we have been dealing with paradoxes which are arguments for explicit contradiction, but not all paradoxes are like that. This does not mean that dialetheism and paraconsistency are irrelevant to their solution, however. To see why, let us consider Sorites paradoxes.

Sorites paradoxes were discovered by Eubulides—the same Eubulides who discovered the Liar—and concern predicates that are vague in a certain sense. Specifically, a predicate is vague in this sense if it is such that, if it applies to an object then making a small change of an appropriate kind to the object does not affect that applicability. Thus, for example, if someone is a (biological) child, they are still so after one nanosecond; if someone is sober, they are still sober after consuming 1 cc of alcohol; if a building is tall, it is tall if 1 cm is added to its height. This behaviour concerning the predicate is sometimes called *tolerance*.

Despite their antiquity, Sorites paradoxes have not been the subject of much investigation historically, unlike the paradoxes of self-reference. However, they have come under intense scrutiny by logicians in the last 40 years or so. I might say that the topic is a vexed one—if anything, even more vexed than that concerning the paradoxes of self-reference; there is certainly no current consensus amongst logicians on a solution.

To fix the discussion, let us consider a typical example: a colour sorites. Take a long coloured strip of paper—say a kilometer long. The colour changes from clear red at one end to clear blue at the other. The change, however, is very gradual. Slice up the strip into segments of 1 cm in length. Call these segments $a_0, a_1, a_2, ..., a_n$. Because of the gradual nature of the change, if one compares two consecutive segments, a_i and a_{i+1}, one can detect no difference in colour.

Now, consider the predicate 'is red'—or even 'appears red'. These are tolerant predicates. a_0 is red. By tolerance, so it a_1. By tolerance, so is a_2. And so on, till we infer that a_n is red—which it clearly isn't. The conclusion of the argument, that a_n is red, is unacceptable, though it's not a contradiction. Of course a_n is not red, and 'a_n is red and not red' *is* a contradiction. It is unacceptable; but its unacceptability is not due to the fact that it is a contradiction; it due to the fact that one of its conjuncts in unacceptable.

At any rate, this is clearly not a type [3] paradox. However, paraconsistency and dialetheism are still relevant. Let us write 'x is red' as Rx. Ra_0 is clearly and simply true; Ra_n is clearly and simply false. Let a_i be some segment in the middle. This is symmetrically poised between the two extremes. There are only two symmetrical statuses: *both* and *neither*. So Ra_i is either both true and false or neither true nor false. Note that both possibilities may be accommodated in the semantics of FDE.

Now, one might argue about how to formulate tolerance, but a natural way is as:

Tolerance: $\neg(Ra_i \wedge \neg Ra_{i+1})$ (for all i)

(It can't be the case that a_i is red, and a_{i+1} isn't.) If Ra_i is neither true nor false, **Tolerance** is not true. (It's, neither true nor false itself.) Since it is true to say that R is tolerant, this rules out the *neither* case. In the *both* case, the inference that takes us along the segments is of the form $Ra_i, \neg(Ra_i \wedge \neg Ra_{i+1}) \vdash Ra_{i+1}$; and as we noted in the last section, this is invalid. If Ra_i is both true and false, and Ra_{i+1} is simply false, both premises are true, but the conclusion isn't. Hence, the paradox is solved, and it turns out to be a type [2] paradox.

Vagueness is a very tricky phenomenon, and there are many more complexities—for example concerning so called "higher order vagueness". However this is not the place to go into them.

But finally, if this is the correct account of the paradox, why did our intuitions lead us astray? It cannot be that the situations we are reasoning about are recondite. Vague predicates are ubiquitous in natural languages, and borderline areas of these are not uncommon either. Here is one plausible suggestion. Studies in cognitive psychology, such as those concerning the Wason card test, have shown that when people reason they often use "quick and dirty" inferences that strike them as correct and give the right answer in usual contexts, though they may not do so in more unusual ones. Now, when we apply a tolerance-type inference, we usually do so only once or twice, and it gives the right answer. Things may go wrong only when we make a long chain of such inferences; but this we rarely do. The suggestion, then, is that the tolerance inference is exactly of this "quick and dirty" kind.

8 How do You Know?

Let us turn to another important topic. We may start by putting on an objection to using dialetheism to solve paradoxes on the table.

Solving the paradoxes by accepting a contradiction is cheap (too easy). If one can accept contradictions, one can get out of any problem. If you accept some view, and someone comes up with an argument for the negation of the view, you can just accept that too!

The objection is lame. The fact that something is logically possible does not mean that it is rationally possible. It is logically possible that Donald Trump is a frog. That is, the claim violates no laws of logic. It is not rationally possible, however. To believe it true would be crazy.

The objection does raise a much more important issue, however. When is it rational to accept a contradiction as true? Well, when is it rational to accept anything as true? The answer is essentially as given by Hume, when he said that a wise person apportions their beliefs according to the evidence (*Enquiry Concerning Human Understanding*, sect. 10, part i). One believes something to be true if the evidence shows it to be so, and not otherwise. The answer is quite general. It applies to a claim of any syntactic form, even one of contradiction.

Of course, this raises a further question: what counts as evidence? Doubtless different kinds of evidence are relevant in different cases (mathematical, scientific, aesthetic). But in the case of a paradox with a contradictory conclusion, part of the answer is obvious. The very paradoxical argument itself provides a certain kind of evidence. But this hardly gets to the heart of the matter. Sometimes, as we have seen, dialetheism may be invoked to solve a paradox which is not of kind [3]; and even when it is, there is more to the matter than this.

A paradox will involve a topic of some kind. Thus the semantic paradoxes of self-reference involve truth and other semantic notions; Sorites paradoxes involve vagueness, etc. How such notions behave is not at all obvious—or we would not find the arguments paradoxical. So we have to formulate an appropriate theory, and the rational thing to accept is the best theory.

There is a very general methodology for evaluating theories. Many factors are relevant: adequacy to the data, simplicity (maybe of different kinds), compatibility with received theories of other things, power, and so on. The exact list is contentious, but the general picture is not. Moreover, different considerations may pull in different directions. Thus, after Copernicus proposed his heliocentric theory of the cosmos in the 16th Century, there was a choice between this and the geocentric theory. Both theories were roughly adequate to the empirical data. The Copernican theory was simpler because it did not use the equant. (It *did* use epicycles.) But the motion of the Earth was at odds with the prevailing dynamic theory of the day, that of Aristotle.

Given the fact that the relevant *desiderata* concerning a choice of theories may pull in different directions, which is the rational theory to accept? The one that performs overall best. How, exactly, to cash out this thought is an interesting question, though we do not need to pursue it for present purposes.

Let us apply these considerations to our paradoxes. Let us take the Liar as an example. Proffered solutions to the Liar are many. Most are consistent, but a dialetheic account is not. Let us agree (at least for the sake of argument) that inconsistency is a black mark against a dialetheic account. There are many other factors that are relevant. The dialetheic theory may do greater justice to the phenomena. (It endorses the T-Schema, with some other theories do not.) It is simple. (There are no transfinite hierarchies of the kind involved in a number of solutions.) The account is not plagued by revenge paradoxes. And so on. It might well be the case that overall a dialetheic account is better. In fact, I think it is, though again this is not the place to go into a detailed discussion of matters. But even if this is not the case, one can see how a dialetheic theory *could* turn out to provide the best solution to the Liar paradox.

More generally, we see two things. First, a dialetheic solution to a paradox may turn out be the most rational theory to accept. Second, that endorsing such a theory is not an easy option. Indeed, no theory is an easy option. A lot of hard work has to be done to establish the properties of a relevant theory and compare these with those of others.

9 Negation

Let us now turn to a couple of other objections to dialetheism, and so to solutions to paradoxes that are dialetheic. These concern negation, and are to the effect that an account of negation that admits dialetheias cannot be right.

For a start, it should be noted that in the history of philosophy accounts of negation are many. Broadly speaking, one can divide these up into three kinds, depending on how, according to them, contradictions behave. Specifically, contradictions may entail:

1. everything
2. nothing
3. some things but not others

Accounts of the first kind are to be found in "classical" logic, intuitionist logic, and a number of others. Since these validate Explosion, they rule out (a sensible) dialetheism. Accounts of the second kind, are to be found in the connexivist logical tradition, which featured (amongst others) Boethius, Abelard, and Berkeley. Though such accounts invalidate Explosion, they tend do so by holding that contradiction have no content ($\neg A$ simply cancels out A), and so are unsuitable for dialetheic purposes. Accounts of the third kind are to be found in more usual paraconsistent logics, such as FDE.

Why might one suppose that accounts of the third kind are incorrect? A first objection is to the effect that to be meaningful, a statement must rule something out. Now, in the accounts of negation in question, $\neg A$ does not rule out A—or *a fortiori* anything else.

The claim that meaningfulness requires some ruling out is clearly problematic. In classical logic, tautologies are perfectly meaningful. Yet they hold in all interpretations (possible worlds, if you like) and so rule out none of them.

But in any case, there is a more important point here. It is true that in FDE there are interpretations where A and $\neg A$ both hold. Indeed, there are interpretations where everything holds (namely the one where every propositional parameter, and so every statement, is both true and false. However, for any A, there are interpretations where $\neg A$ holds, and interpretations where it does not. $\neg A$ therefore rules out *those* interpretations. In general, we may think of the content of a statement, A, as a pair, $[A]^+$ and $[A]^-$, the first comprising those interpretations where A is true, and the second comprising those where it is false. All statements will then have determinate content. And generally speaking, different contradictions will have different contents.

A more interesting objection concerns the *use* of negated sentences. The use of a negated sentence, $\neg A$, it is claimed, is to deny A. Since a dialetheist who asserts $\neg A$ may yet accept A, asserting $\neg A$ does not deny A, which it must.

Now, to assert $\neg A$ is not necessarily to deny A. An easy way to see this is to note that most of us have inconsistent beliefs. These are often exposed by a process of Socratic questioning. The questioner will get us to assert A, but then, with some more questions, get us to assert $\neg A$—without there having been a change of mind. This may well cause us to revise our beliefs, and maybe quite correctly so. But the point is that the assertion of $\neg A$ is not a denial of A. A and $\neg A$ have both been endorsed. That is *why* we may feel compelled to revise our beliefs.

To understand what is going on here, one needs to be very clear about the difference between negation and denial. Negation is an *operation on sentences*. It applies to one sentence to generate another. Denial is a *speech act*, something we do by uttering sentences in a certain way.

There are many kinds of speech act: asserting, denying, questioning, commanding, and so on. Defining kinds of speech act is a sensitive business; but at least as a first cut, to assert A is to utter something with the intention that the hearer thereby comes to accept A—or at least believe that the utterer accepts A; to deny A is to utter something with the intention that the hearer thereby comes to reject (refuse to accept) A—or at least believe that the utterer rejects A.

Quite different speech acts may be performed by uttering one and the same sentence. Thus, if I say 'The door is open', this could be an assertion, a command, a question. Determining what speech act is being performed is a matter of decoding the intentions of the utterer. To do this, we take into account context, stress and intonation, power relations, and so on. Doubtless this is a fallible and sometimes tricky business; but we do it successfully all the time.

Now, uttering $\neg A$ could indeed be a denial of A. You (a non-dialetheist) say 'the Liar sentence is consistent'. I (a dialetheist) say 'It most certainly is *not* consistent'. That would surely be a denial. But if I next say 'The Liar sentence is true' and add 'Moreover, it is not true (as well)', the utterance of the second sentence would surely be an assertion. I am telling you *more* about the status of the Liar sentence, not rejecting what I have just said.

So in sum, negation can be used to make denials, but uttering a negation is not necessarily a denial. It all depends.

10 The Boundaries of Thought

To finish, let us return to the subject of paradox one final time. This book is called *Borderlands of Thinking*. Borderlands are boundaries between two regions; in this case, presumably, between thinking and not thinking. Before we turn to the boundaries of thought, let us pause for a moment to consider boundaries in general.

Boundaries are curious objects. They are contradictory almost by definition, since they both separate and join their flanks. One might say that they cleave the two. (In English, the world *cleave* has contradictory meanings. It can mean *separat*e, as in 'With one stroke she will cleave the rock in two'. It can also mean *adhere to*, as in 'We cleave to these fundamental truths'.)

Now, boundaries can be extended. We have already met ones—contradictory ones—of this kind in our discussion of the sorites paradox. The boundary between segments that are red and segments that are not is not sharp, but extends over several segments. But apparently contradictory boundaries can be sharp too, as we are about to see.

Let us now come to the boundaries of thought. Suppose that there are some things that can be thought and some things that cannot be thought. There is, then, a division, and so a boundary, between them. However, there is an issue here—as was pointed out by both Hegel in his *Logic* (bk. 1, ch. 2, sec. B(c)(β)) and Wittgenstein, in his introduction to the *Tractatus*. To think a boundary is to think of things on each side of it. But in the present case this is something one obviously cannot to—or better, to do so requires one to think the literally unthinkable, and so engender contradiction.

Of course, one might just take that as a proof that there are no things that are unthinkable. But let us now return to set theory. Many paradoxes of set theory do not occur at absolute infinity. Some occur much "lower down".

One such is König's paradox, an infinitary version of Berry's paradox, and concerns the ordinals.

To think about something one must be able to refer to it; and if one can refer to it, one can think about it. (If one quantifies over all ordinals, then there is a sense in which one is thinking about all of them; but this is not the sense of *thinking about* which at issue here, which we might call *individual thought.*) So the thinkable ordinals are exactly the referable ordinals.

How far the ordinals go on (up?) is, as I have said, a vexed subject. However, it is beyond contention that there are many more than can be referred to by noun phrases of standard English (in, say, my current idiolect). This is established by a simple cardinality argument. So there are many ordinals that cannot be referred to (thought about). Now, the ordinals share with the natural numbers the property that any set of them has a least member. Hence, consider the set of ordinals that cannot be referred to. This has a least member, namely:

- the least ordinal that cannot be referred to.

(Or, if you like, consider the set of all those ordinals, x, such that all the ordinals less than or equal to x can be referred to. This is a sequence of ordinals. Hence there must be a least ordinal greater than all the things in it. This must be the least ordinal that cannot be referred to.) Now, by construction, the least ordinal that cannot be referred to cannot be referred to. But it can be referred to. It is referred to by the displayed noun phrase.

This is König's paradox. The argument shows that there are indeed things that cannot be thought (referred to). Moreover, it shows that there is a number which is on *both* sides of the boundary between the thinkable and the unthinkable. We may even take the contradictory number to be the boundary (cut off point) between those things that are thinkable and those things that are not.

11 Conclusion

So much for our whistle-stop tour of paradox, paraconsistency, and the relationship between them. What we have seen is that despite the long-standing orthodoxy concerning the PNC, dialetheism is relevant to the solution of certain equally long-standing paradoxes. Contemporary developments in logic and metaphysics have opened up a whole new vista on some of these.

Our contemporary ability to go beyond the consistent (para-consistent) has made it possible to explore things that were at one time beyond belief (para-doxical). No longer so.

12 Further Reading

- *Section 2:* Oppy, Hájek, Eswaran, and Mancosu (2021). Malinas (2021). Huggett (2018).
- *Section 3*: Bolander (2017). Beall, Glanzberg, and Ripley (2016). Priest (2006a), ch. 1 ('Semantic Paradoxes').
- *Section 4*: Bagaria (2019). Priest (2006a), chs. 2 ('Set Theoretic Paradoxes'), and 18 ('Paraconsistent Set Theory'). Weber (2021).
- *Section 5*: Priest, Berto, and Weber (2018). Priest (2006b), ch. 1 ('Aristotle on the Law of Non-Contradiction'). Łukasiewicz and Heine (2021).
- *Section 6*: Priest, Tanaka, and Weber (2018). Priest (2008), ch. 8 ('First Degree Entailment'). Priest (2002).
- *Section 7*: Hyde and Raffman (2018). Priest (2010). Priest (2019).
- *Section 8*: Priest (2006a), ch. 7 ('Pragmatics'). Priest (2006b), Part 3 ('Rationality').
- *Section 9*: Routley and Routley (1985). Priest (1998). Priest (2007). Sainsbury (2009), ch. 7 ('Are Any Contradictions Acceptable?').
- *Section 10*: Cantini and Bruni (2021). Priest (1995), part 3 ('Limits and the Paradoxes of Self-Reference').

References

[1] Bagaria, J. (2019), 'Set Theory', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/set-theory/.

[2] Beall, J., Glanzberg, M., and Ripley, D. (2016), 'The Liar Paradox', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/liar-paradox/.

[3] Bolander, T. (2017), 'Self-Reference', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/self-reference/.

[4] Cantini, A., and Bruni, R. (2021), 'Paradoxes and Contemporary Logic', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/paradoxes-contemporary-logic/.

[5] Huggett, N. (2018), 'Zeno's Paradoxes', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/paradox-zeno/.

[6] Hyde, D., and Raffman, D. (2018), 'Sorites Paradox', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/sorites-paradox/.

[7] Łukasiewicz, J., and Heine, H. (tr.) (2021), *The Principle of Contradiction in Aristotle—a Critical Study*, Honolulu, HI: Topos Books.

[8] Malinas, G. (2021), 'Simpson's Paradox', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/paradox-simpson/.

[9] Oppy, G., Hájek, A., Eswaran, K., and Mancosu, P. (2021), 'Infinity', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/infinity/.

[10] Priest, G. (1995), *Beyond the Limits of Thought*, Cambridge: Cambridge University Press; 2nd edn, Oxford: Oxford University Press, 2002.

[11] Priest, G. (1998), 'What's so Bad about Contradictions?', *Journal of Philosophy* 95: 410–26; reprinted as ch. 1 of Priest, Beall, and Armour-Garb (2004).

[12] Priest, G. (2002), 'Paraconsistent Logic', pp. 287–393, Vol. 6, of D. Gabbay and F. Guenthner (eds.), *Handbook of Philosophical Logic*, 2nd edn, Dordrecht: Kluwer Academic Publishers.

[13] Priest, G (2006a), *In Contradiction*, 2nd edn, Oxford: Oxford University Press.

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

[15] Priest, G. (2007), 'Paraconsistency and Dialetheism', pp. 129–204, Vol. 8 of D. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, Amsterdam: North Holland.

[16] Priest, G. (2008), *Introduction to Non-Classical Logic*, Cambridge: Cambridge University Press.

[17] Priest, G. (2010), 'Inclosures, Vagueness and Self-Reference', *Notre Dame Journal of Formal Logic* 51: 69–84.

[18] Priest, G. (2019), 'Dialetheism and the Sorites Paradox', ch. 7 of S. Oms and E. Zardini (eds.), *The Sorites Paradox*, Cambridge: Cambridge University Press.

[19] Priest, G., Beall, J., and Armour-Garb, B. (eds.) (2004), *New Essays on the Law of Non-Contradiction*, Oxford: Oxford University Press.

[20] Priest, G., Berto, F., and Weber, Z. (2018), 'Dialetheism', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/dialetheism/.

[21] Priest, G., Tanaka, K., and Weber, Z. (2018), 'Paraconsistent Logic', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, https://plato.stanford.edu/entries/logic-paraconsistent/.

[22] Routley, R., and Routley, V. (1985), 'Negation and Contradiction', *Rivista Colombiana de Matemáticas* 19: 201–231.

[23] Sainsbury, M. (2009), *Paradoxes*, 3rd edn, Cambridge: Cambridge University Press.

[24] Weber, Z. (2021), *Paradoxes and Inconsistent Mathematics*, Cambridge: Cambridge University Press.