thanks is due Bernard Katz for his invariably sound advice and philosophical insight; I have benefited greatly from both. I also want to thank Steven Gerrard of Acumen Publishing for his understanding and patience. To the family and dear friends who have provided unstinting support, encouragement, and dinners out, I am more grateful than I can say.

This book is not finished. I make this admission not for the sake of creating a new paradox of the preface, but to acknowledge that each time I review the manuscript, there are further considerations and responses I feel inclined to add. Practical considerations, however, dictate that it is time to stop.

1 The nature of paradox

Paradoxes can be fun. They can also be instructive, for the unravelling of a paradox may lead to increased philosophical knowledge and understanding. The paradoxes studied in this work offer promise of both these features. But paradoxes may be also disturbing; their study may reveal inadequacies, confusion or incoherence in some of our most deeply entrenched principles and beliefs. The reader is forewarned: some of the material that follows may prove unsettling.

It seems wise to begin at the beginning, with the questions "What is a paradox?" and "How does one resolve a paradox?" But first we need some examples of paradoxes at our disposal.

The Monty Hall paradox

You are invited to be a contestant on a fabulous game show. The host of the show, Monty Hall, explains how the game works. After some initial banter and scintillating chat, you will be presented with three doors, *A*, *B* and *C*. Behind one of the doors will be the car of your dreams – a Porsche, a Jaguar, whatever you wish. Behind each of the other two doors is a worthless goat. Which door conceals the car is decided randomly. You will first be asked to pick a door; then Monty, who knows what is behind each door, will pick, from one of the other two doors, a door that has a goat behind it, open that door, and show you the goat.

At that point, you will be offered a second option. You may stay with your original choice, and keep whatever is behind that door. Or, you may switch to another door, and keep whatever lies behind it.

Naturally, you are delighted to accept the invitation. With a week to go before the show, you feel there is nothing to deliberate about other than what you will wear. It seems clear that it is all a matter of luck. The first choice is entirely arbitrary. It is equally likely that the car is behind any one of the three doors; the probability that the car is behind any given door is 1/3. Similarly, the second choice is a matter of whim; there is no reason to prefer either switching to another door or staying with your original choice. Suppose, for example, you first pick door *A*, and Monty then shows you the goat behind door *C*. That means the car is either behind door *A* or behind door *B*. But it is equally likely that it is behind either door; there is no reason to prefer one to the other. So the probability that the car is behind either door is now 1/2, and there is nothing to gain either by switching or by not switching.

It's all a matter of common sense, you tell yourself. How could being shown that one of the doors I did not choose has a goat behind it give me any reason to prefer one of the two remaining doors?

The night before your television appearance, a mathematician friend appears at your door, seemingly agitated. "Do whatever you want on the first choice", he says. "But on the second choice, you *must* switch! It's just become clear to me", he continues. "Look at it this way. Suppose you pick door *A* on the first round, and that Monty then shows you that door *C* has a goat behind it. Monty had to choose between doors *B* and *C*, and he wanted to pick a door that concealed a goat. He might have been in a position where he could pick either door (both were 'goat doors'); or he might have *had* to pick door *C*. Initially, the probability that the car was behind either door *B* or door *C* was 2/3. So the probability that his choice was *forced* was 2/3. But his choice was forced only if the car is behind door *B*. So if you switch to door *B*, your chance of winning is 2/3. You can't do better than that!"

Panicked, you start to protest, but he interrupts. "Let me put it another way. Suppose you get to play the game many times and you are going to pick a strategy. If you consistently pick a door (say door *A*) and stay with it, you will win 1/3 of the time (the 'car door' is determined randomly). But 2/3 of the time the car will be behind either door *B* or door *C*, in which case Monty will, in effect, show you which of the two it is not behind. So you will win 2/3 of the time if you follow the strategy of switching – twice as often as if your strategy were not switching."

What should you do?

Other paradoxes

The barber paradox

Imagine a charming village, as yet untouched by the tourist trade, in which there is only one barber. He is extremely busy, for he cuts the hair of all and only those villagers who do not cut their own hair. But who, we may wonder, cuts the barber's hair? Suppose he cuts his own hair. If he does, then, since he is a villager, he does not cut his own hair. Suppose, alternatively, that he does not cut his own hair. If he does not, then, since he is a villager, it follows that he does cut his own hair. So the barber in this village cuts his own hair if and only if he does not cut his own hair.

The Achilles and the tortoise paradox

The tortoise and Achilles are to have a race. Of course, the tortoise is much slower than Achilles; Achilles, at his best, can run ten times faster than the tortoise. To make the contest interesting, the tortoise is given a head start of 10 metres; the racetrack is 100 metres. Can Achilles overtake the tortoise? Consider. By the time Achilles reaches the tortoise's starting point (point 1, which is 10 metres ahead of Achilles' starting-point), the tortoise will have travelled another metre to reach point 2 (since Achilles runs ten times as fast as the tortoise). But once Achilles has reached point 2, the tortoise will have travelled another tenth of a metre to reach point 3. By the time Achilles has reached point 3, the tortoise will still be one hundredth of a metre further ahead at point 4. And so on. It seems that whenever Achilles has caught up to where the tortoise *was*, the tortoise is still some tiny distance ahead. Thus, Achilles cannot pass the tortoise and cannot win the race.

The ship of Theseus paradox

Theseus, an experienced sailor well aware of the hazards of the sea, has a ship that he decides needs complete renovation. The ship – call it "*T*" – consists of 1,000 old planks. When the renovation begins, Theseus' ship is placed in dock *A*. The crew is ordered to work as follows. In the first hour of renovation, they are to remove one plank from *T*, replace it with a new one, and carry the old plank to dock *B*. In the second hour of renovation, they are to remove an adjoining plank, replace it with a new one, and carry the old plank to dock *B*, where it is appropriately fastened to the plank that has been removed in the previous hour. They are to remove a third plank in the third hour. And so on. After 1,000 hours, a ship has been assembled in dock *A*, call it "*X*", that consists of 1,000 new planks; there is also a ship in dock *B* – call it "*Y*" – that consists of the 1,000 old planks removed from Theseus' ship and then reassembled in exactly the same way they had been arranged prior to the renovation. Which ship is Theseus' ship? Which ship is *T*?

If you methodically took apart the ship, and then reassembled it exactly as it was, surely you would say that it was the same ship.[1] But that is exactly what has happened here. *T* was first taken apart, then reassembled and is now in dock *B*. So *Y* is *T*. Note that *Y* is made out of exactly the same materials, arranged in exactly the same fashion, as *T* was when Theseus brought it into port.

On the other hand, if you remove one plank from a ship and replace it with a new one, you still have the same ship. Such a slight change cannot affect the identity of the object. So after one hour, the ship in dock *A* is still *T*. But again, removing one plank from a ship and replacing it does not affect the identity of the ship. Thus, after two hours the ship in dock *A* is *T*. And so on. Finally, after 1,000 hours, the ship in dock *A* is *T*. Thus, *X* must be *T*.

The taxi-cab paradox

In the town of Greenville there are exactly 100 taxis, of which 85 are green and 15 are blue. A prominent citizen witnesses a hit-and-run accident that involves a taxi, and testifies that the taxi was blue. The witness is subjected to tests that determine that, in similar circumstances, he is 80 per cent reliable in his colour reports. Is it likely that the taxi in the accident was blue?

First, it seems clear that we are entitled to accept what the witness says as highly likely. He has proved 80 per cent reliable in similar circumstances, and there is no reason to think there is any relevant difference in this situation. Surely what he says can be considered highly probable, and should be regarded as such in a court of law.

On second thought, if we take the long view, it seems unlikely that the witness was correct in his colour identification. To see this, consider 100 randomly selected taxi accidents in Greenville. About 85 of these accidents will involve a green taxi and about 15 will involve a blue taxi. If the witness were to report on the 85 green taxi accidents, he would report correctly in about 80 per cent of the cases and incorrectly in 20 per cent. This means that of the 85 green taxi accidents, he would report about 17 as involving a blue taxi. The 15 blue taxi accidents would presumably also yield 80 per cent correct reports, or 12 reports of blue taxis involved in accidents. Were a witness of 80 per cent reliability to report on 100 randomly selected taxi accidents in Greenville, then, there would be about 29 (= 17 + 12) blue taxi reports, only 12 of which would be accurate; that is, only 41 per cent of the blue taxi reports would be correct. So it seems more likely than not that the witness in our original case was mistaken in his report of a blue taxi.

What is a paradox?

In order to appreciate why these scenarios seem baffling, confusing and yet absorbing, it is necessary to have a better understanding of the sort of problem they pose. Using these few paradoxes as background, let us consider the question: what is a paradox?

One striking feature of these problems is that they present a conflict of reasons. There is, in each, an apparently impeccable use of reason to show that a certain statement is true; and yet reason also seems to tell us that the very same statement is utterly absurd. Apparently letter-perfect operations of reason lead to a statement that reason is apparently compelled to reject.

Let us unpack what this means. It should first be noted that each paradox presented above contains an argument; this feature is central to the philosophical notion of paradox. The popular use of the term "paradox", by contrast, is undoubtedly broader. A recent newspaper report, for instance, says that the rosier health picture for those with HIV-AIDS has "sparked a paradoxical response, a disturbing trend to unprotected sex among young gay men".[2] Here "paradoxical" seems to have the force of "irrational" or "unfitting". Statements that seem absurd at first sight, but on closer examination are seen to be true, are also referred to as "paradoxical" in popular usage. In Gilbert and Sullivan's *The Pirates of Penzance*, for instance, the following is taken to be paradoxical: Frederic is 21 years old, but has had only five birthdays. (The clue is that Frederic was born in a leap year on February 29.)

But we are pursuing the philosophical notion of paradox. We might say with Quine that "a paradox is just any conclusion that at first seems absurd, but that has an argument to sustain it".[3] This seems to be essentially in line with the traditional definitions in the literature. It should be made explicit, however, that the argument in question must seem strong or compelling; arguments that are clearly fallacious do not yield paradox. Revising Quine's definition, we can say: a paradox is an argument that appears flawless, but whose conclusion nevertheless appears to be false.

But what is meant by speaking of an argument as flawless? Evaluating an argument normally requires assessing two components: the premises, and the reasoning from the premises. For an argument to be without fault, the premises must be true and the reasoning correct. So we have:

> A paradox is an argument in which there appears to be correct reasoning from true premises to a false conclusion.

This is to be understood as saying that the appearance of each of three elements is required: correct reasoning, true premises and a false conclusion.

Is this an adequate account of the notion of paradox? It is easy enough to see how this traditional definition fits the example of Achilles and the tortoise. There we have what seems to be a meticulous argument leading to the obviously false conclusion that Achilles can never pass the tortoise. However, some of the other paradoxes considered above do not fit quite so neatly into this mould. In the ship of Theseus paradox, for instance, there are seemingly compelling arguments for two different conclusions.

And while neither conclusion may appear clearly false, the two conclusions (*X* is *T*, *Y* is *T*) certainly appear to be inconsistent. The Monty Hall paradox and the taxi-cab paradox also seem to share this feature: two apparently faultless arguments lead to two apparently inconsistent conclusions.

This points to the need to distinguish two types of paradox. A type I paradox, such as Achilles and the tortoise, has one argument and one conclusion; a type II paradox, such as the ship of Theseus, involves two arguments and two conclusions. The definition just given may do for type I paradoxes, but type II paradoxes require a more complex account as follows:

> A type II paradox occurs when there is one argument in which there appears to be correct reasoning leading from true premises to a conclusion *A*, and another argument in which there appears to be correct reasoning leading from true premises to a conclusion *B*, and *A* and *B* appear to be inconsistent.

Since it would be tedious to express every point made in the following discussion as it applies both to type I and to type II paradoxes, I shall sometimes illustrate just with one type, and allow the reader to work out the corresponding point for the other type.

The discussion began by noting that a paradox presents us with a conflict of reasons; the successive accounts of "paradox" just offered are to be understood as attempts to unpack or spell out just what is meant by this initial characterization. It remains to be considered whether *any* conflict of reasons constitutes a paradox. Is it always the case, for instance, that when there are two apparently strong arguments leading to apparently inconsistent conclusions, there is an intellectual problem to be solved, or a conflict about what to believe?

Consider the following two arguments:

99% of Texans are rich.	99% of philosophers are poor.
Jones is a Texan.	Jones is a philosopher.
∴ Jones is rich.	∴ Jones is poor.[4]

Suppose that the premises of the two arguments appear to be true. Certainly, the conclusions are inconsistent. So each of the two arguments appears to be an instance of good inductive reasoning, and (let us assume) to have true premises, but the conclusions are clearly incompatible.

Nonetheless, it is it is quite clear that this sort of conflict of reasons does not constitute a paradox.[5] Situations of this sort, situations in which there is strong evidence for two competing claims, are commonly encountered. But there is nothing mysterious or baffling here, nothing to puzzle or confound. The example trades on a well-known feature of inductive arguments: they are defeasible. That is, certain premises may *in themselves* provide strong inductive support for a conclusion, and yet further evidence may "defeat" the argument; even if the premises are justifiably believed to be true, they may not, *given further information*, warrant belief that the conclusion is true. So in the Texan/philosopher case, there is no difficulty in granting that, given only this evidence, we are not entitled to regard either conclusion as true. The evidence provided by one set of premises defeats the evidence provided by the other. The rational course is clear: suspend belief in each conclusion. There is no sense here, as there is in a paradox, that we are *compelled* to accept the conclusion of each argument, despite recognition of the inconsistency. Thus there is nothing troubling or problematic about this sort of conflict of reasons; there is no intellectual problem to be solved.

Note that the same sort of difficulty may arise with regard to the account of a type I paradox. Consider an argument of the form:

> 99% of *A*s are *B*.
> x is an *A*.
> $\therefore$ x is a *B*.

Suppose that the premises seem to be true, and yet you apparently can see that x is not a *B*. Again, there is a conflict of reasons. In such a case, it may be that the perceptual evidence takes priority; or, in some extreme cases, it may be that the inductive evidence is taken to be stronger. Or, finally, it may be that the two sorts of evidence are considered roughly equal in strength, in which case you must suspend belief concerning whether or not x is a *B*. But in none of these cases is there anything paradoxical.

The account of the notion of paradox thus requires further refinement. Happily, there is a simple revision that suffices to handle

cases of this sort. Up to now, it has been implicitly allowed that the argument(s) in a paradox might be either inductive or deductive. But what seems essential now is that the paradox-generating argument appears to be deductively correct, that is, that the premises appear to logically imply the conclusion. So we have:

> A type I paradox is an argument in which there appears to be valid reasoning from true premises to a false conclusion.

The same sort of qualification is to be understood in the account of a type II paradox. Valid arguments, it should be noted, do not share with inductive arguments the feature of defeasibility. Given premises we are entitled to accept as true, valid reasoning will yield a conclusion we are entitled to accept as true, *no matter what further information we have*. Clearly, if there is an apparently valid argument from apparently true prèmises to an apparently false conclusion, then we do indeed have an affront to reason; we are entitled to feel baffled and confounded. In a paradox, what appears to be cannot possibly be.

One matter remains. What may seem puzzling, in this series of definitions, is a presupposition that is made here, and in the philosophical literature generally, that the premises of a paradox-generating argument can be, or appear to be, true. The paradoxes presented thus far all involve a description of *a situation that does not exist*; they all involve a story that is not factual. How then can statements about these situations be, strictly speaking, true? In the taxi-cab paradox, for example, a premise of the second argument is: in 100 randomly selected accidents in Greenville involving one taxi, about 85 will involve a green taxi and about 15 will involve a blue taxi. But how can we regard this premise as true when, as far as we know, Greenville does not exist?

Most paradoxes start with a story (although not all do, as we shall see later). This suggests that the notion of truth we use in evaluating the premises of a paradoxical argument, in such cases, is not the everyday, straightforward sense, but is more analogous to fictional truth. As in the taxi-cab paradox, we can discuss what is true in Tolstoy's *Anna Karenina* even though we are clear that the events and situations described never occurred. Moreover, when we speak of the character of Vronsky, our comments can be understood as prefaced by "In the novel *Anna Karenina* . . .".[6] Similarly, our statements about Greenville, it could be argued, are to be taken as prefaced by "In the taxi-cab paradox . . .". Of course, there are significant disanalogies between fictional truth and truth in a paradox; nonetheless, the comparison is suggestive.

"Narrative-paradox" is the term I shall use for those paradoxes based on descriptions of nonexistent situations, or stories. To say that a statement is true in a narrative-paradox is to say that it is true in the story or narrative *N* that is the basis of the paradox. But how is this to be understood? We need an account of "true in *N*" that is appropriate for the context of paradoxes.

To begin with the obvious, truth in the story of the taxi-cab paradox is clearly a function of the description of the story that gives rise to the paradox. Let *D* be the conjunction of statements that describe the narrative *N*. Then we might suggest:

(I) A statement *S* is true in *N* just in case *S* is logically implied by *D*.

Put differently, this says that any possible world in which *D* is true is a world in which *S* is true.

It might be objected, however, that this first analysis is too restrictive. The condition that *S* be implied by *D* is sufficient for truth in *N*, it might be argued, but it is not also necessary. Recall that in the taxi-cab paradox, we want to say that it is true that in 100 randomly selected accidents involving one taxi, about 85 will involve a green taxi and about 15 will involve a blue taxi. But this is not implied just by the data concerning the colour of taxis in Greenville. At the least, we also need a statement that is a matter of common knowledge: the colour of a car is irrelevant to how accident-prone it is.

There are two possible responses here. The first is to grant the force of the objection, and attempt to revise the account so as to get around the difficulty. For instance, it might be proposed that:

(II) A statement *S* is true in *N* just in case *S* is logically implied by *D*, or by the conjunction of *D* and *C*, where *C* is a contingent true statement.

("True" used without a qualifier here means "true *simpliciter*"). The chief drawback of this proposal is that it is hard to see

exactly what restrictions should be placed on *C*. Clearly, *C* must at least be consistent with *D*. But more than this seems to be required. Otherwise, we will have to say that in the taxi-cab paradox, it is true that, for instance, Glendon College has a bilingual curriculum. This surely does not conform exactly to our intuitions concerning truth in a narrative-paradox.[7] A still further qualification might be suggested in response, to the effect that *C* be an item of common knowledge. But this is not a very precise notion, and, in any case, the analysis still appears too weak. The statement "George Bush is president of the United States" seems to qualify as common knowledge yet, again, is not true in the paradox.

The alternative, which I favour, is to maintain that the objection to the first proposal does not go through. To do so, we must insist that the statement "The colour of a car is irrelevant to how accident-prone it is" is properly part of the description of the situation envisaged in the taxi-cab paradox, although it is not normally made explicit. It is by no means surprising that this should be so, for it is often no trivial matter to say exactly how the story should be specified, to see precisely what is necessary for the argument.[8] Part of the work demanded by a paradox is to determine how the description must be filled out, sharpened or refined, and this may emerge gradually as the result of ongoing analysis. So there is nothing implausible in the suggestion that a statement not normally made explicit is nonetheless intended as part of the story.

Thus far, the truth of a statement *S* in *N* is straightforwardly a function of *D*. However, one further consideration remains. Paradoxical arguments sometimes appeal to abstract general principles. In evaluating the argument for switching in the Monty Hall paradox, for instance, we grant that if there are two possible outcomes of your doing either *A* or *B*, *O* is the preferred outcome, and *B* is more likely to yield *O* than *A* is, then it is rational to do *B* rather than *A*. If this principle is a necessary truth (true in all possible worlds), then it need not be added explicitly to the premises of the argument. If the argument is valid with it, then it is also valid without it. But such principles are often included in the paradoxical argument, and this may make it easier to assess the validity of the argument. So we may say:

(III) A statement S is true in N just in case S is logically implied by D, or by the conjunction of D and $T_1, \ldots, T_n$, where each T_i is a necessary truth.

Assuming that such general principles, if true, are necessarily true, (III) seems the most promising account thus far.[9]

Let us recast the definition of "paradox" in light of the above. We now have a distinction between "true" ("true *simpliciter*") and "*N*-true". Thus we may say:

> A type I paradox is an argument in which either (i) the premises appear to be true, the conclusion false and the argument valid or (ii) the premises appear to be *N*-true, the conclusion *N*-false and the argument valid.

Finally, a word on terminology. When we say that a particular statement is true in a narrative *N*, we are not using the ordinary, everyday notion of truth. Still, there seems to be no danger of confusion if we use the word "true" with no further qualification in discussing paradoxes based on stories. In contexts where it seems wise to remind the reader that we are using a special sense of "true", the term "*N*-true" will be used. For paradoxes that do not begin with a story, of course, we need only the ordinary sense of "true". The reader should easily adapt to supplying the appropriate sense of "true" in a given context.

Types of paradox

We have already seen that paradoxes may be classified as type I or type II. Before going on to consider how a paradox may be resolved, further distinctions will be helpful. First, the notion of a veridical paradox:[10]

> A type I paradox is veridical just in case its conclusion is true (*N*-true in the case of a narrative-paradox, true *simpliciter* otherwise).

> A type II paradox is veridical just in case the conclusions of the two arguments are both true.

Of course, the truth of the conclusion(s) does not provide a logical guarantee that the reasoning is impeccable, nor that the premises are true (although one might search in vain for a veridical paradox in which there is a flaw in either). Still, what is deceptive in a veridical paradox, at the very least, is the appearance of falsity in the conclusion (or the appearance that at least one of the conclusions is false).

The notion of a falsidical paradox is understood analogously. A type I paradox is falsidical provided that its conclusion is false; a type II paradox is falsidical if at least one of its conclusions is false. So in a falsidical paradox, the fault lies in the paradox-generating argument (or in at least one of the arguments), either in the premises, or in the reasoning.

Of the paradoxes we have to draw on, only the barber paradox can be considered to be veridical, for it is generally conceded that there is no flaw in the paradoxical argument, either in the premises or in the reasoning. But the conclusion of the barber paradox is:

> The barber cuts his own hair if and only if he does not cut his own hair.

How can we say that this conclusion is true, as required by the definition of "veridical"? After all, any statement of the form "*P* if and only if $\sim P$" is necessarily false.

Here it is essential to remember that in classifying the barber paradox as veridical, we are committed only to saying that the conclusion of the paradoxical argument is *N*-true, true in the barber story. This is guaranteed if the conclusion is implied by the statements in the description of the paradox. But it is now generally recognized that the story that gives rise to the paradox (there is a village in which there is a barber who cuts the hair of all and only those villagers who do not cut their own hair) is incoherent, that it is impossible for there to be such a village. Hence, there is no problem in granting that the description implies an impossible conclusion, and thus no problem in granting that the conclusion is *N*-true. A necessarily false premise may imply a necessarily false conclusion.

The same sort of thing may occur in a type II veridical narrative-paradox. Given that the paradox is veridical, the conclusions of both arguments will be *N*-true. One might naturally assume, in such a case, that the conclusions must be consistent. But if the description of the paradox is itself necessarily false, then the conclusions may both be implied by the description, and yet be inconsistent.

A clear example of a type I falsidical paradox is provided by the Achilles and the tortoise paradox. Of the type II paradoxes already introduced, all seem to be falsidical: the two conclusions, in each case, are unquestionably inconsistent and yet the descriptions seem logically coherent.

The final distinction to be drawn here is between controversial and uncontroversial paradoxes:

> A type I paradox is uncontroversial if either there is general agreement that its conclusion is true or general agreement that its conclusion is false.

Note that to say that a paradox is uncontroversial does not mean that there is *no* controversy surrounding it. There may be broad agreement that the conclusion of an argument is false, even though there is no consensus concerning the diagnosis of the flaw in the argument. Both the barber and Achilles and the tortoise provide examples of uncontroversial paradoxes. The corresponding definition for type II paradoxes is:

> A type II paradox is uncontroversial if either there is general agreement that both conclusions are true or there is general agreement that a particular conclusion is false.

Looking at the type II paradoxes, the Monty Hall paradox can fairly be regarded as uncontroversial, since it is generally recognized that it is rational to switch one's choice of door. But the taxi-cab and the ship of Theseus are both controversial.

To sum up, the distinctions drawn is this section can be illustrated as follows:

Barber	type I, veridical, uncontroversial
Achilles and the tortoise	type I, falsidical, uncontroversial
Monty Hall	type II, falsidical, uncontroversial
Ship of Theseus	type II, falsidical, controversial
Taxi-cab	type II, falsidical, controversial

How to resolve a paradox

At this point, we have an account of what constitutes a paradox, and an understanding of the different types of paradox. What remains to be dealt with is the question of how to respond to a paradox, how to provide a resolution.

Paradoxes present us with apparently impeccable operations of reason that nonetheless lead to apparent absurdity. They are upsetting because, while the illusion persists, we have a challenge to the supposed veracity and reliability of reason. If this is where logic can lead, then why would we recommend logic or respect its dictates? The threat to reason can be overcome only by puncturing the illusion created by the paradox.

To resolve a paradox it is necessary to show that the paradoxical argument does not in fact present us with an impeccable use of reason leading to a patent absurdity. There are thus two principal options in providing a resolution for a type I paradox: (i) we may dispel the illusion that the argument is air-tight by isolating and diagnosing a flaw or fallacy in the argument; or (ii) we may *explain away* the appearance of falsity in the conclusion. This is accomplished by explaining why the conclusion appears to be false even though it is in fact true. In pursuing alternative (i), attempting to find a flaw in the argument, there are two further options: (a) show that at least one of the premises is not true; or (b) show that the argument is invalid.

Briefly, the options for resolving a type II paradox are as follows. One sort of resolution will consist in finding a flaw in one of the two paradox-generating arguments, either in the premises or in the reasoning. Alternatively, we may explain why it *appears* that the conclusions cannot both be true even though, in fact, both are.

Later chapters will study instances of the different types of resolutions in detail. For now, let us look at some of the paradoxes already introduced in order to illustrate these distinctions. Of necessity, the discussion will be limited to the uncontroversial paradoxes.

The barber paradox has been cited as an instance of a veridical paradox. This despite that the fact that its conclusion, that the barber cuts his own hair if and only if he does not, is a contradiction. To make sense of this, it is essential to keep in mind, as was pointed out earlier, that it is strictly *N*-truth and *N*-falsity that are relevant to the assessment of the argument of a paradox based on a story. The real issue is whether the conclusion is *N*-true, that is, whether the description of the village implies the premises of the argument, which, in turn, imply the conclusion. It may at first seem that this could not be, since the description of the village seems perfectly consistent, and a consistent set of statements does not imply a contradiction. But it is not difficult to convince oneself that the description is, in fact, contradictory. After all, it refers to, among others, the barber himself, and says of him that he cuts his own hair if and only if he does not. So there is no problem in granting that the conclusion of the paradoxical argument is *N*-true, that it is implied by the description of the village. It appears to be *N*-false only because it is contradictory and the description of the village appears, at first, perfectly consistent. Thus, we can explain the *appearance* of falsity while granting that the conclusion is true.

This treatment of the barber paradox exemplifies one basic approach to veridical paradox resolution: showing that the description of the paradox is inconsistent. If it is, then there need be no surprise or shock at what the description implies, since anything follows from an inconsistency.

The Monty Hall paradox provides an example of an uncontroversial falsidical paradox, for it is generally agreed that the correct strategy is to switch your choice of door after Monty shows you a goat door. This means that there must be a flaw in the argument that there is no good reason to switch – in the "no switching argument". But what, precisely, is wrong with it? Suppose you pick door *A*, and then Monty shows you that door *C* has a goat behind it. A key premise in the no switching argument is that after door *C* is revealed as a goat door, it is *as likely* that the car is behind door *A* as it is that it is behind door *B*; the two possibilities are equally likely. To find a fallacy in the no switching argument and thereby resolve the paradox, it is this premise, I believe, that must be successfully rebutted.

Consider. If Monty's intent had been simply to open one of the three doors at random, and this intent had resulted in his opening door *C*, and revealing a goat behind it, then it would be equally likely that the car was behind door *A* and that it was behind door *B*. But Monty's choice was in fact restricted to door *B* or door *C*, and his intent was to choose a goat door. So you know something about

door *B* that you do not know about door *A*; there is an asymmetry in your knowledge of the two doors. In a choice between door *B* and another door, where the intent was to choose a goat door, door *B* was not chosen. How is this knowledge relevant to the assignment of probabilities? Well, there was a 2/3 chance that Monty's choice was forced – that only one of the two doors was a goat door. Since you knew that Monty's intent was to open a goat door, and that he could do this, *his opening door* C *and revealing it to be a goat door does not change this probability*. But if Monty's choice was forced, then the car is behind door *B*. Thus, there is a likelihood of 2/3 that the car is behind door *B*; and, accordingly, there is only a 1/3 likelihood that it is behind door *A*.

If this analysis is correct, then we have dissolved the paradox, dispelled the illusion, by showing that one of the premises in one of the paradox-generating arguments is false. Of course, the reasoning is subtle; the flaw in the no switching argument is not easy to discern. Indeed, some mathematicians and probability theorists have been initially taken in by the no switching argument, although none has persisted in defending it.

Further illustrations of the resolution of a paradox must await the more detailed treatment of individual paradoxes. But, at this point, some words of caution are in order. First, to be satisfactory, the resolution of a paradox should be robust: it should stand up to strengthened versions of the paradox. For instance, the paradox-generating argument may initially be presented with an extremely strong premise, a premise that makes a very broad, sweeping claim. If so, it may take no great acumen to point out counter-examples to the premise. However, before declaring the paradox vanquished, we need to be sure that it cannot simply be reinstated when a suitably weakened version of the critical assumption is provided. To be robust, the attack on a paradoxical argument should be focused on the strongest, most impregnable version of the argument available.

A related point concerns different versions of the same paradox. The Achilles and the tortoise paradox, for example, seems to be essentially the same as the racetrack paradox (see the Appendix). If so, then any solution to the one should also be applicable, with the appropriate changes, to the other. An attempted resolution of a paradox that cannot be applied successfully to *every* version of the paradox must be off the mark in that it focuses on some inessential feature of the paradox. This is not to say that it is always evident whether one paradox is a variant of another. In fact, the criteria for two arguments being versions of the same paradox are far from obvious. One can even imagine cases in which the fact that a solution applies to one argument, but not to the other, would be cited as reason for denying that these are just two versions of the same paradox. Nonetheless, as we shall also see, there are many cases in which it is entirely clear that different scenarios are all versions of the same paradox, and thus require a unified solution.

What does *not* count as a resolution of a paradox? The negative may be almost as significant as the positive here. One very common and natural response to a stated paradox is *to present another argument.* More specifically, the response is an attempt to present an even more compelling or persuasive argument for (or against) the conclusion (one of the conclusions) involved in the original paradox. Consider the Monty Hall paradox, for example. Suppose a mathematician friend, having announced that she has a solution to the paradox, proceeds to give a very clear, very explicit, and very powerful argument for the conclusion that one ought to switch doors. Whatever the merits of her argument and the worth of her contribution, they do not constitute a *solution* to the paradox, for the argument in favour of not switching is left untouched. So there is still an apparent conflict of reasons: two ostensibly strong arguments for inconsistent conclusions. Consequently, there is still a sense of confusion, of being befuddled, which can be cleared up only by an analysis of an error or flaw in one of the arguments. A paradox is not unravelled by attempting, however successfully, to prove that one "side" in the conflict is correct. Later chapters will provide examples of philosophers responding to the challenge of a paradox in this way.

The ideal, in treating a paradox, is to puncture the illusion of letter-perfect reasoning leading to clear absurdity. But short of achieving this ideal, there are still worthwhile contributions one can make. The mathematician's argument alluded to in the previous paragraph may convince us that the rational response in the Monty Hall scenario is indeed to switch, when previously we had been uncertain. While such an argument does not suffice to dissolve the paradox, it may convert a previously controversial paradox to the status of uncontroversial. Assuming the argument to be correct, this constitutes an advance in the understanding of the problem, and progress in the search for a solution. If it is known that the correct strategy is to switch doors in the Monty Hall game, then the focus must be squarely on the no switching argument, and the attempt to locate a flaw in it. The range of possible solutions has been narrowed.

One other way to move a controversial type I paradox into the uncontroversial category is worth mentioning here (and will be illustrated in Chapter 4). Suppose we construct an argument that is strongly analogous to the original paradoxical argument, but that leads to a conclusion even more preposterous or bizarre than that of the paradoxical argument; so bizarre, in fact, that it is *completely clear* that the conclusion, and therefore the argument, have to be rejected. Since the new argument is strongly analogous to the original paradoxical argument, it is now apparent that the original argument must also be rejected. Again, the set of possible options for a solution has contracted.

There is, finally, one other way to make progress on a paradox, short of resolving it or narrowing the range of possible solutions: progress can be made by clarifying one of the central arguments. This may be achieved in a variety of ways. Among the more significant are: analysing one of the key concepts; setting out, fully, rigorously and explicitly, the premises necessary for the argument; ensuring that the premises are just as strong as needed, but no stronger, so that the argument is as immune to criticism as possible; and making clear exactly what the inferential steps are that take us from the premises to the conclusion, so that any logical gaps in reasoning will be more apparent.

This section has considered how to resolve a paradox, how not to resolve a paradox and how to make progress on a paradox short of resolution. To conclude, let us consider the question of why we feel a pressing need to untangle a paradox, why we care about paradoxes. One answer has already been suggested. An unresolved paradox is a threat to the trustworthiness of reason. How can reason command our respect if it leads to absurdities? But another motivation stems from the fact that the proper resolution of a paradox may give us greater philosophical knowledge. Faulty assumptions concerning, for instance, justified belief, or rational action, may be uncovered in the unravelling of the paradox.

Of course, some paradoxes, of which the barber is one, have little, or no, philosophical punch. The ship of Theseus, on the other hand, may reveal a good deal about the principles governing our concept of the identity of a physical object. The depth of a paradox is generally considered to be a function of the sort of philosophical impact its resolution will have. At one extreme of the spectrum, a paradox may reveal an incoherence that necessitates a fundamental revision of our conceptual scheme; at the other end of the spectrum there is the barber. Unfortunately, a proper appreciation of the depth of a paradox frequently must await its resolution.

2 Paradox and contradiction

Dialetheism

Paradoxes are baffling. Faced with an apparently impeccable argument that leads to an apparently outrageous conclusion, we are confused and confounded. On the one hand, the conclusion appears false; on the other hand, it apparently must be true. What appears to be cannot be, we assume. This is the source of our fascination; this is why there is a problem.

Recently, impressive arguments have been advanced that this underlying assumption is mistaken. A statement can be *both true and false*, it is maintained; further, it can be rational to believe that a given statement and its negation are both true. Contradictions (statements of the form "$A \& {\sim}A$") can be true, and can be rationally believed.[1] If this view, known as "dialetheism", prevails, there are clear consequences for the account of paradox. In Chapter 1, three strategies for dealing with a paradox were distinguished: show that the argument is invalid; show that a premise is false; and explain away the appearance of falsity in the conclusion. But if contradictions can be true, and can be rationally believed, then there is another legitimate response to a paradox: accept everything that appears to be the case, that is, grant that the conclusion of the paradoxical argument is both true and false.

This is an apparently fantastic proposal. Until recently, dialetheism would have been dismissed out of hand as a simple conceptual confusion. Largely because of the work of philosopher Graham Priest, however, it has come to be regarded as at least deserving of serious consideration and response.