Chapter

7

Impossibilities

By rights, nothing in this chapter should exist. But a chapter on impossibilities is a necessity in a book on paradoxes. We begin by looking at pretend impossibilities, illusory paradoxes that are readily resolved by careful clarity. We next look at genuine impossible objects, or rather actual drawings of impossible objects. We also pay attention to some impossible intentions. Impossibility is sometimes due to language, sometimes to our conceptual powers; it is never grounded in the impossible object itself, which necessarily fails to exist. We end by considering a mathematical necessity so strange it seems it must be impossible.

VARIETIES OF IMPOSSIBILITY

To be impossible is almost a contradiction in terms. If something is impossible, it cannot be. So how can we speak of being impossible at all? Impossibility is rather a form of necessary non-being. The impossible does not exist because it cannot exist. The impossible is a species of nothing. And if the impossible is nothing, surely nothing is impossible.

On What There Isn't

There is an old philosophical question about whether holes exist. A hole is not so much a something as a nothing—a lack, an absence, something missing. But you can count holes, which means that many exist. If many holes exist, why shouldn't one exist? If there are more holes in a piece of Swiss cheese than there are in this argument, then that means there must exist at least one extra hole in that cheese that is not in my argument. And if that hole exists, so does every other hole.

The impossible is very much like a hole. Impossibilities may literally be nothing, but they can be counted and set out in a bewildering array of varieties. One is tempted to say that impossible objects come in many shapes and sizes; the trouble is they don't "come" at all.

And yet the impossible can not only be imagined, it can be discovered. For example, in the field of physics there is the law of conservation of energy (energy cannot be created or destroyed) and the law of special relativity (it is impossible to exceed the speed of light); also, in the field of mathematics, we saw earlier that Gödel proved a complete theory of arithmetic is impossible (see p. 59).

So it is possible to discern various types of impossibility and, as we shall see, there are innumerable species of impossible objects. Impossibility is necessarily diverse. One might almost say that many forms of the impossible are possible, and here are some.

Practically versus Inherently Impossible

A useful distinction can be made between what is practically impossible and what is inherently impossible. What is practically impossible today may be routine tomorrow, for instance if technology develops. Travel to the moon was once impossible, but ingenuity, determination, lots of money, and a competitive spirit changed that. It has, however, always been logically possible.

Our concern in this chapter is the more difficult impossibilities that are not due to practical obstacles. We are looking for the inherently impossible—that which cannot be achieved or realized in any way, and never can be.

Simple versus Composite Impossibility

Most impossibilities are made up of inherently possible parts that simply cannot be put together in the way proposed. For example, a self-contradiction is a proposition conjoined to its negation. Each is consistent by itself, but they mutually exclude each other so can't both be true. A married bachelor offends by being both unmarried, as are all possible bachelors, and yet married. Satyrs, centaurs, and sphinxes are all impossible monsters composed of possible parts. In this way composite impossibilities seem to be impossible wholes composed of possible parts.

Could there be an inherently impossible object not composed of possible parts? Or does it take at least two possibilities to be put together incorrectly in order to make an impossibility? One must wonder whether simple impossibility is even possible. But if a simple impossibility is simply impossible, that would surely prove to be the case in point. And, with that mind-bending thought, you may be relieved to read that simple impossibilities are too complex to be further considered here.

The Meaningless versus the Impossible

The attempt should be made to distinguish between the meaningless and the impossible. The meaningless (also called the nonsensical) is defined in some regions of philosophy as any claim that can be neither true nor false. The impossible, in contrast, is whatever is necessarily false.

Decide for yourself:

"Virtue is triangular."

"Breakfast is the first meal to become president."

These statements, you will no doubt agree, are not true. But would you call them false? To call them false has seemed to some philosophers to be giving them more credit than they deserve. These sentences commit category errors; they are too outlandish even to be called false. For instance, the first sentence applies a concept from one domain (geometry) to a concept from a totally different domain (ethics). This sentence is in fact a stock example of a meaningless or nonsensical claim (meaning it cannot be significantly said to be either true or false).

Against this pro-nonsense position it may be argued that, if the above claims were false, they would not be false by accident. Indeed, if they were false, they would be necessarily false, and so impossible. Moreover, those who try to distinguish the meaningless from the impossible in this way will find themselves uncomfortably committed to the claim that it is impossible that it should be true that virtue is triangular. The distinction itself becomes meaningless, and the nonsensical is seen as simply the impossibility of sense.

A QUESTION FROM WITTGENSTEIN

What time is it on the sun? Is it even possible for this question to have an answer? Can it be morning on the sun? Is it not 4:00 am on the sun once a day? If it is (merely) false that it is prior to 4:00 am on the sun, then is it not 4:00 am or later there?

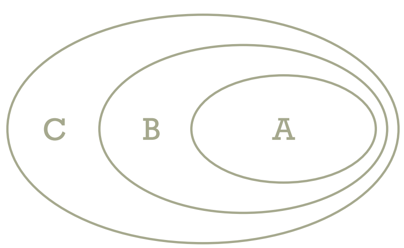
1081_BBo-Paradoxes_132-147.indd 134-135 04/02/2013 15:00

TRANSGRESSIONS THROUGH TRANSITIVITY

Imagine a businesswoman who is detained on business in Boston, and misses home and family in Cleveland. Homesick, she sighs: "If I were not in Boston, I would be in Cleveland." We may assume she is speaking the truth. Now surely it would also be true for that person to affirm the following sentence: "If I were in Alaska, I would not be in Boston."

This innocuous conjunction leads by an apparently indubitable principle to an impossible conclusion. The principle, known as "transitivity," takes the form of valid argument: if the two premises are true, the conclusion must also be true. Check for yourself:

If A, then B.
If B, then C.
Therefore, if A, then C.



However, by an obvious process of substitution, it follows that our lonely traveler is in a position to assert: "If I were in Alaska, I would be in Cleveland." How can such geographical nonsense follow logically from true premises in a valid argument form? Could the principle of transitivity be wrong?

Well, transitivity is safe, but it is misapplied here. Transitivity assumes that the B clause (not in Boston) means the same thing in both premises; however, the vagaries of language mean it doesn't in this case. Intuitively we understand that a more complex analysis is required than simply forcing these sentences to fit the principle of transitivity.

A Biological Impossibility?

A further example of transitivity going awry can be found in the following example.

We start with an, admittedly crude, definition of a species involving the ability of two individuals belonging to it to interbreed. (This is based on biological principle, so irrelevant factors like age, opportunity, and inclination are ignored.)

Now imagine that a particular species of fish populates a lake. A catastrophic climate change dries the lake until it is no longer whole, but five smaller, separate lakes each containing roughly one fifth of the original population. Over time, each separate group evolves independently, and eventually the fish grow so different that, in some cases, interbreeding becomes biologically impossible. According to our definition this is the point at which we can say that distinct species have evolved.

Now imagine a biologist comes along and takes a number of specimens from each lake. He returns to his lab and places the different fish in the same tank. In the course of his experiments he observes the following:

Group A can interbreed with Group B.
Group B can interbreed with Group C.
Group C can interbreed with Group D.
Group D can interbreed with Group E.
Group E cannot interbreed with Group A.

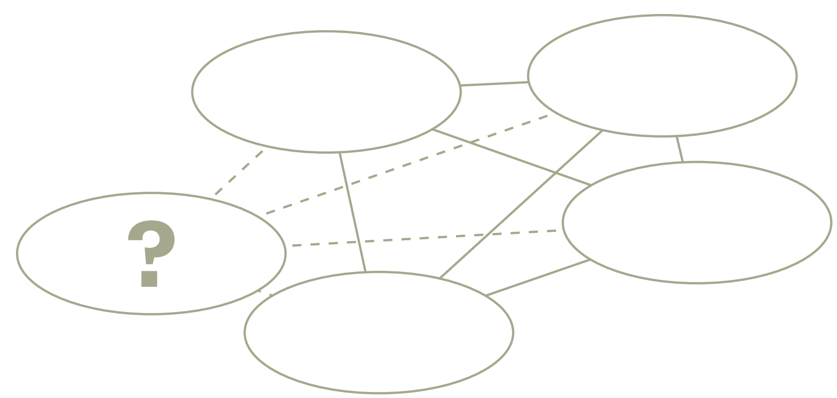
We can apply transitivity to this to say that if Group A can interbreed with Group B, and Group B can interbreed with Group C, then Group A can interbreed with Group C. By applying transitivity a few more times you can reach the conclusion that Group A can interbreed with Group E—but this contradicts the observed fact that Group E cannot interbreed with Group A

The same conundrum can be put in another way, considering groups separated not by physical barriers but by the passage of time. People of today could in principle mate with people of the previous millennium, who in turn could mate with people who lived in a millennium previous to that. Assuming transitivity, one can appear to prove that there was never a distinct species from which human beings arose.

Intuitively we know that this can't be the case, but what has gone wrong? Is our definition of a species (based on interbreeding) at fault; is the set of observed facts a biological impossibility; or is the principle of transitivity worthless?

Is the case laid out above an a priori argument against the evolution of species? Well, in a word, yes—albeit a very bad one.

In reality, what it shows is that our chosen definition is inconsistent with evolution. Despite its intuitive appeal, the definition is an abstraction with at best a local truth. Species in reality are more like individuals (like particular twigs on the evolutionary tree) rather than discrete classes.



IMPOSSIBLE OBJECTS

One might be excused for thinking the impossible is inconceivable. But being unable to exist is different from being unable to be thought or imagined. The impossible cannot be, but in some cases it is imaginable. Our ability to conceive the impossible may be limited, but it is greatly assisted by depictions of impossible objects. It even turns out that the impossible can be represented by a variety of means, which are startling in what they reveal about us.

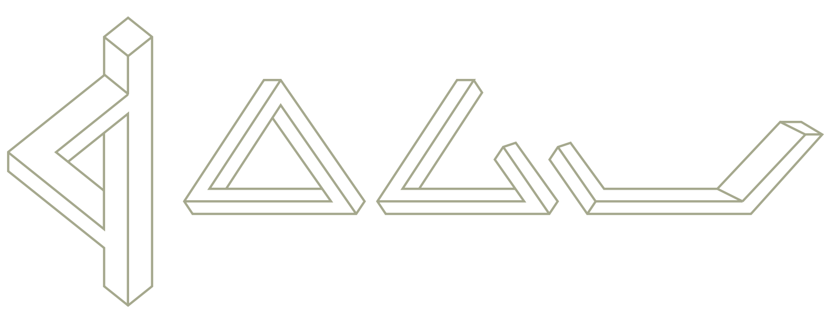
Some Possible Figures

The fact that the following constructions cannot be physically created does not prevent them from making attractive images. The impossible can be depicted, and proves to be pleasantly jarring. The impossible object shown here was devised by Swedish artist, Oscar Reutersvärd, a variation on his more famous tribar—also called the Penrose triangle after Lionel and Roger Penrose, who later independently rediscovered it.

Each corner in this figure is spatially coherent, but the figure as a whole is not. Each corner is shown as if from a different point of view; perspective on the whole is not unified. The result is a depiction of a structurally impossible object. No such object could exist in the three dimensions we occupy.

The distinguished scientist Richard L. Gregory has urged a distinction be made between impossible figures and impossible objects. The Penrose triangle, he insists, is an impossible figure, not an impossible object. The grounds for his distinction are that there exists after all a three-dimensional object which, when viewed from a particular perspective, looks just like a tribar.

This is demonstrated by the illustration below, which initially appears to be a Penrose triangle, but the illusion falls apart when the perspective is rotated.



Examples of the objects depicted here have actually been built and displayed as public art, as if to defiantly affirm the existence of the impossible. Often, the illusory effect is all the more robust, because it is not a mere drawing. When you see such an object from a particular perspective, it seems to violate the very space in which it exists.

Contrary to Gregory, it ought to be said that the Penrose triangle is precisely a possible figure, not an impossible one, as its invention clearly demonstrates. The figure was invented; its possibility was discovered. What the figure represents, however, is an impossible object, an object that cannot exist in three dimensions. This is true, even though by way of trickery of perspective it can appear otherwise.

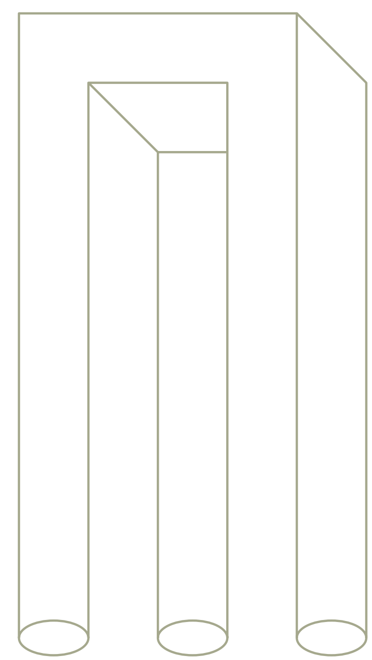
The existence of three-dimensional objects that look (from one perspective) like a tribar does not make the tribar a possible object. At most, hitherto unexpected ambiguities or alternative interpretations of the figures are introduced. What these three-dimensional objects show is that the illusion of the Penrose triangle does not depend entirely on a two-dimensional representation, but rather on perspective.

A Possible Tribar?

All the cleverness of illusion demonstrated by a tribar sculpture does not make a tribar possible. The tribar is a possible figure, but an impossible object after all. However, it is worth noting a slight caveat: While a tribar, or similarly impossible object, cannot exist in three-dimensional space, it is possible to construct animated models of them. To animate such a figure it is necessary not only to construct a three-dimensional model on a computer, but also to alter that model as the figure rotates or the viewpoint changes. Only in this way can the illusion be retained.

The Devil's Trident

On Gregory's view, the illustration below—variously known as a poiuyt, blivet, or devil's trident—is a genuine impossible object. One ought to say on the contrary that it is only a depiction of a genuine impossible object. Cover up the top or the bottom of the image to enhance the effect.



1081_BBo-Paradoxes_132-147.indd 138-139 04/02/2013 15:00

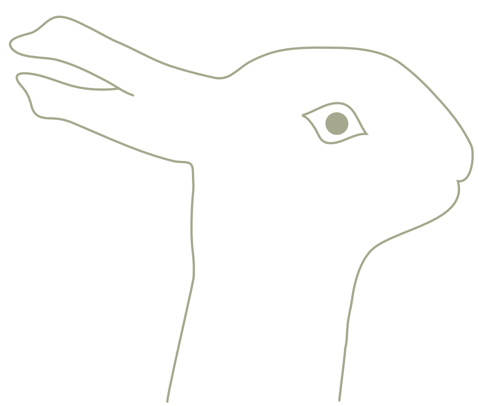
DOUBLE-TAKE

Perhaps you are familiar with stereoscopes, devices for viewing three-dimensional images. If the images loaded in the stereoscope represent slightly different perspectives on the same object, the two-dimensional pictures are fused by the brain to create a three-dimensional experience. This fusion is understood to be an unconscious inference based on information derived from the two retinal images. The visual experience of three dimensions is an inference.

Here's a fun impossibility: an attempt to fuse the unfusable! Just take two hollow cardboard tubes (cut an empty paper-towel roll in two) and fasten them together, as if they were binoculars. Then attach a very different image to the end of each tube. Look through them and relax your eyes.

What happens when the images are totally distinct and cannot be fused? A remarkable phenomenon known as "binocular rivalry" sets in. At first, both images jostle with each other—instead of fusing they may unstably overlap, each partially obliterating the other for a brief time. But in time one image will come to predominate, while the other disappears completely. Both eyes are still open, but only one image is seen. What has happened to the other image? Just wait a little longer, and the lost image will reappear and replace the first one, which now fades away. As you keep watching, the images slowly alternate, as if the two images were competing for your attention—and they are!

How does this happen? Well, each eye receives a totally distinct image, and both can be seen. Each of these images also activates the associated regions in the visual cortex of the brain; however, the level of processing varies. Since the two images cannot be fused, they are like competing hypotheses about what you are looking at. When one hypothesis wins out, the processing of its image is enhanced, while the disconfirming evidence from the other image is suppressed. But it persists, making the original hypothesis shakier, until a different inference is made. The lost image now has its moment in the sun and, in this way, a competitive rivalry continues.



A PRACTICAL EXPERIMENT

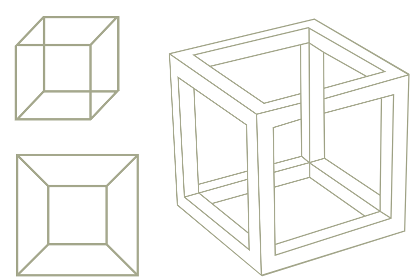
Part 1: Make a cube out of stiff wire, attached at a corner to a rod. In a dark room, shine a flashlight at a wall, casting a shadow from the wire-frame cube. Rotate the cube. What do you see?

Part 2: Paint your wire cube fluorescent so it will glow in the dark. In a dark room, while it can be seen, hold it in your hands. You may experience a visual reversal despite tactile evidence to the contrary from your hands, which is a freaky experience.

It Can't Be Both!

A similar rivalry occurs with bi-stable figures. These are ambiguous images that are capable of two incompatible interpretations, and below left is one of the most famous—the duck–rabbit. It can't be both a duck and a rabbit, so your view of it alternates between one and the other and will not remain stable.

A yet simpler example of a bi-stable image is the Necker Cube—the two left-hand images at the bottom of the page—which spontaneously flits back and forth between appearing as a cube projecting into the page and one standing out of it. The image cannot be interpreted in both ways simultaneously, so alternates—which suggests representational rivalry. As a check against its own errors, the brain looks for alternative interpretations, and in this instance finds a plausible alternative, which in its turn comes to dominate.



The two images directly below are bi-stable figures, which offer incompatible and competing interpretations. They are both variations on the Necker Cube, and can be seen as either projecting out from the page or cutting into it. The second image from the left can be seen either as a cube in linear perspective or as a truncated pyramid. By contrast, the two depictions below right represent genuinely impossible objects, objects that cannot exist in real space. These illusions suggest that impossible objects exist, even though we know they can't. No coherent structural description is possible for these objects, and spatial anomalies and violations of space are perceived.

Seeing Is Disbelieving

Even four-month-old babies have been shown to be sensitive to the inconsistent relationship in the depth cues and to the structural irregularities that make the cube depicted below impossible. Meanwhile, in adults, a certain brain signal, associated with previous exposure to items in a recognition test, has been found in the case of possible objects, but not for impossible objects of comparable complexity. It is hypothesized that the brain region generating this signal is responsible for representation of the global three-dimensional structure of objects.

Short of space here, can we get away with just 3 cubes?

1081_BBo-Paradoxes_132-147.indd 140-141 04/02/2013 15:00

DIVINE IMPOSSIBILITIES

Suppose an omniscient being to be one who knows every truth. Now consider: "This statement is not known to be true by any being." If this statement is false, it cannot be known, even by an omniscient being. But also, if it is false that the statement is not known, it must be known, hence true. So the statement cannot be false.

Any being omniscient by the above definition would therefore know it to be true—but this contradicts the statement itself! It follows that no being knows all truths. Either there is no omniscient being, or omniscience is not the knowledge of every truth!

The Paradox of the Stone

Could God make a stone so heavy that even He could not lift it? To answer yes is to imply that there is something God could not do (lift this stone). To answer no is also to imply that there is something God could not do (create this stone). Either way there is something God cannot do. So something is impossible for God to do, and this seems to threaten the presumptive omnipotence of God. To assume God is all-powerful, together with the consideration of this stone, seems to result in a logical contradiction.

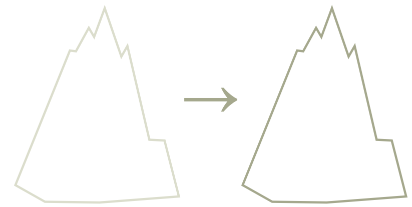
So which is it? Is the idea of omnipotence self-contradictory? Or does God's power embrace even the inherently impossible?

There are several lines of response to this paradox open to believers. A believer might accept the conclusion that God could not create such a stone, but deny that this constitutes any genuine limitation on infinite power. For instance, one might argue that, if God could create this stone, it would entail a logical contradiction; so God can't. But not being able to achieve the impossible is no skin off His nose. For, if the impossible were achievable, it would not really be impossible.

Another approach similarly concedes that God cannot create a stone He cannot lift, but uses a very different argument to show that, despite this, there is not an act that God cannot do. God's omnipotence in this regard requires only that: for any weight, God could create a stone of that weight; and for any weight, God could lift a stone of that weight.

Under these two conditions, there is no limit on God's creating or lifting of stones; yet there will not be a stone God could create but not lift. Indeed it follows from these two conditions that God can lift every stone He can create.

But this, expressed differently, says that God cannot create a stone He cannot lift. Granted, if God could create a stone He could not lift, it would be bad for His



omnipotent reputation. But if He cannot create a stone He cannot lift, this is thanks to His boundless abilities, not despite them.

This nice solution suffers from the problem that "indefinitely large weights" is a radically incoherent notion. Infinite mass and gravity suffer from a physical incoherence God should not have to answer for. In any case, perhaps there is a more radical option for the believer.

Miracle or Contradiction?

"Jesus said 'I tell you, it is easier for a camel to go through the eye of a needle than for a rich man to enter the kingdom of God.' When the disciples heard this, they were greatly astonished and asked, 'Who then can be saved?' Jesus looked at them and said, "With man this is impossible, but with God all things are possible."' (Matt. 19.24–26)

Now it is physically impossible to put a camel through the eye of a needle. Of course, all that means is that it would take a miracle, a contravention of physical law, to make it happen. It would be astonishing, but it does not yet amount to a logical impossibility, or a contradiction in terms. (In fact, if it were pure geometry, it might even be a theorem; see The Pea and the Sun—pp.132–3).

But maybe self-contradiction is simply the most extreme miracle. Perhaps the most extreme test of faith requires an attempt to believe the absurd.

So a believer may go further and ask: Is God limited by logical possibility, or does even the realm of logical self-contradiction fall within God's ambit? If all things are possible, then even the impossible should be possible. If nothing is impossible, again even the impossible should be possible.

The Belief in Impossibility

As author of physical laws, God may violate them at will. As author of the laws of logic, why couldn't God suspend them as well? This seems to have been the view of René Descartes:

"The truths of mathematics … were established by God and entirely depend on Him, as much as do all the rest of His creatures. … You will be told that if God established these truths He would be able to change them, as a king does his laws; to which it is necessary to reply that this is correct. … In general we can be quite certain that God can do whatever we are able to understand, but not that He cannot do what we are unable to understand. For it would be presumptuous to think that our imagination extends as far as His Power." Descartes, Letter to Mersenne, April 15, 1630.

At a crucial moment, Jesus rebukes his followers whose lack of faith leaves them without the power to perform miracles: "If you have faith as small as a mustard seed, you can say to this mountain, 'Move from here to there' and it will move. Nothing will be impossible for you." Matt. 17.20–21. So even for human beings, provided they have sufficient faith, nothing is impossible. Perhaps the true test of belief is precisely to believe the unbelievable.

THE PEA AND THE SUN

Believe it or not, you can take a solid ball, dissect it into five disjoint pieces (one of which can be a single point), and then rearrange these by rigid motions such as translation and rotation—in other words motions that don't stretch or distort the pieces—so as to form two spheres the same size as the original one.

Alternatively, if you wish, you can dissect a ball into as few as five non-overlapping parts that can be reassembled, using rigid motions, into another ball of any volume you please. A pea-sized ball can be split and then reconstituted into a solid sphere the size of the sun.

Can such a counterintuitive idea really be true? Well, it is, albeit within the field of theoretical geometry. This very strange idea is known as the Banach–Tarski paradox, after Stefan Banach and Alfred Tarski, the two great Polish mathematicians who proved it in the 1920s.

It is known as a paradox because it is counterintuitive, but it is not a genuine paradox at all. Far from being necessarily false, it is necessarily true, deducible from some widely assumed mathematical principles.

Illusions of Solidity

The appearance of a paradox depends in part upon prevalent confusions regarding solidity. Even as great a philosopher as John Locke found it difficult to remove the notion of solidity from the notion of body itself. Space could be empty, he thought, but bodies in it were solid, so that where one body was no other could be at the same time. In more recent times, it has become commonplace to muse about the conundrum propounded by the physicist James Jeans. He pointed out that since everyday solid objects like walls and tables are made up of atoms, and atoms in turn are mostly empty space, then everyday solid objects are mostly empty space. This may be subverted by mentioning the obvious opposing point that objects are not the space they occupy.

Ironically, it is only a mathematical solid that is well and truly solid. It, however, being theoretical, has no mass. The mathematical ball is a solid sphere of geometrical space. That is to say, it consists in a set of points making up a spherical region of space. Geometrical space is understood to be composed of points that may be identified by their real-number coordinates. Just as the real-number line is "made up" of points corresponding to real numbers, so three-dimensional space is represented by ordered triplets of real numbers (which give the coordinates of points in space). A ball is therefore an infinite set of points, and parts of the ball are subsets of this set of points.

The requisite subsets are discrete—in other words non-overlapping—subsets that, together, make up the original sphere. These subsets are unusual, not like the parts into which you might slice a real pea, a real baseball, or a real sun. These are not divisions of mass at all, because our mathematical ball has no mass. As scatterings of points, some of these subsets do not even have volume, which is how they can be rearranged and reassembled into a solid with a different volume. Technically, subsets of the ball involved in the Banach–Tarski dissection are said to have no measure.

How can this be shown? It is impossible to prove it here, but a rough idea of the proof can be given. First, Banach and Tarski proved their theorem by generalizing a similar result by Felix Hausdorff—a Jewish German mathematician who committed suicide in 1942 in the face of Nazi persecution. Hausdorff's theorem asserted that a sphere (that is, the surface of a mathematical ball) is equidecomposable into two copies of itself—in other words the one hollow sphere can be cut up and reordered to make two. Banach and Tarski applied this finding to a solid ball, as opposed to simply the surface, by conceiving the latter to consist of nested spheres (much like an onion is made up of layers). This brief summary leaves the doubling of the sphere wholly unexplained, but it does give at least some hint of the approach taken.

Some Lexical Lunacy

However, the crucial idea of doubling can be suggested by considering the case of a special dictionary equidecomposable into two copies of itself. Imagine a dictionary that is simply a list of all the possible words of finite length that can be spelled with two letters, say A and B. Every finite string of As and Bs is a word in this language, and they can all be arranged in alphabetical order. All words are finite, but the number of possible words is infinite. This is a lot of words, so after the initial one-volume printing, the dictionary came out in two volumes: one containing all words starting with A, the other all words starting with B. Since all the words in each volume start with the same letter, it was considered redundant to keep repeating it. Thus all the initial As in volume one, and all the initial Bs in volume two, are left out of the second printing.

The paradoxical result is that the two volumes are now identical, all of the entries read the same, and they are also identical to the one-volume original edition. Something like this strange doubling is involved in the Hausdorff result, except in that case it involves sequences of rotations of the spheres.

Fitting a Camel Through the Eye of a Needle

If a pea-sized ball can be partitioned into discrete parts that can be rearranged to form a sun-sized ball, then it hardly seems a miracle for a camel-sized region of space to be dissected and the parts recomposed into a camel small enough to fit through the eye of a needle. A miracle involves the violation of physical law—but the Banach–Tarski Theorem is a form of mathematical law. It is not an exception to a necessity, but an expression of it.

1081_BBo-Paradoxes_132-147.indd 144-145 04/02/2013 15:00

Exercise 7

The Toxin Paradox

THE PROBLEM

How to form an impossible intention. Let us suppose a billionaire will an altruistic passion for paradox puts to you the following proposal, the terms of which lawyers and relevant experts of the most scrupulous sort have verified as accurate. You are to receive $1 Million simply for forming an appropriate intention at midnight tonight, an intention to perform a particular act at noon tomorrow. You will not be required to perform the later act, only to intend at midnight—based on this offer alone—to commit that act tomorrow at noon. Infallible brain scanning technology will reveal at midnight whether or not you in fact intend to act at noontime the next day.

The act which you must intend to later perform is the ingestion of a certain toxin, which will make you very sick for 24 hours, but you will fully recover without side-effects. To get the $1 Million, you must intend at midnight to ingest the toxin the following noon. Show that it is impossible to form such an intention.

THE METHOD

It seems that it can be done, for nothing is so easy as to frame an intention, especially when the responsibility of carrying it through is so conveniently relaxed, as in this case. No one would willingly ingest this toxin, were it not for the $1 Million. But acquiring $1 Million is clearly preferable to avoiding the illness due to poisoning. So, intending to act is both easy and preferable, and when you add in the facts that not actually following through would bring in this case no penalty, and spare you a day of misery, the prudent person would chose to intend to take the toxin, but then later to refrain from it. This is still honest, since completion of the action is not in fact required to get the $1 Million.

But here arises a problem. To intend at midnight to take the poison next noon, requires that you not intend later to forego taking it. One cannot honestly or consistently intend to do what you also intend not to do. If you know in advance you are not going to take the poison, then you cannot intend to do so. Of course, you could *pretend* to intend, and merely *say* you are going to take the toxin later; but this won't work, as the infallible intention-detection technology will catch you in the lie.

So you can't honestly intend to do what you know you later won't. From this perspective, the unlikely billionaire's money is secure in his hands.

Various ruses may be tried to get another chance at winning that money. For instance, one could ignore the probable fact that one will not take the toxin if it is no longer necessary, and instead focus exclusively on, and try to strengthen, the resolve to swallow the poison. Think about the money that will come, not the illness that one will endure. This is a form of self-deception, deceiving oneself that something isn't true, when you know very well that it likely is. But the mere appearance to oneself of having formed an intention is not any better than the mere appearance to others, and is no more likely to succeed. It is not enough to persuade oneself that you have an intention which in fact you do not; you must in fact have the intention, and not be deceived, by yourself or anyone else, into thinking you have it. The lie-detection technology is just that good.

Incidentally, there is a related paradox of self-deception. How can one deceive oneself at all, since you are privy to any plans that you might hatch to fool yourself. To deceive others, it is necessary that they not know they are being deceived. To fool yourself, it ought likewise to be impossible *not to know* that one is being deceived, since the ploy you are using is, after all, *yours*, and so not unknown to you.

Also related to the toxin paradox is what is known as Moore's paradox: something can be true, though you don't believe it; and yet it would seem to be impossible for anyone ever consistently to assert of any definite proposition, P, that "P is true but I don't know P".

Perhaps you could persuade yourself that you intend at midnight to take the toxin at next noon by making further arrangements that compel you to do so. For instance, you could sign a legal agreement requiring you to incur a $10 Million cost if you do not drink the toxin at noon tomorrow, having committed to. This gives you plenty of reliable incentive to ingest on time the poison, though again the offer itself does not strictly require you do so (only the intention to at midnight is needed). The billionaire's offer, however, also required that you intend to consume the toxin only on the basis of the billionaire's offer. So such additional incentives are ruled out.

THE SOLUTION

The inventor of this paradox is Gregory Kavka. He was led to conclude that "one cannot intend whatever one wants", that "intentions are only partly volitional", and that "intentions are constrained by reasons for action". In much the same way, one cannot believe whatever we want, belief cannot be wanton, for belief is constrained