## Introduction

## Deep Thought: a brief history of thought experiments

On the face of it, thought experiments are a useful way to gain new knowledge about the world, by means of 'armchair philosophy' only. And, whether they are called thought experiments or not, the approach has had an important role in not only theoretical philosophy, but in practical science over the centuries.

The ancient Greeks particularly liked to explore using the technique. Not that they had no concept of more conventional experimentation too. Empedocles (495–435 BCE), who wisely divided the world between two forces, 'love' and 'strife', also founded one of the first medical schools, from which source a fragment of writing records a very practical investigation of the circulatory system. But Heraclitus (c.500 BCE), who liked to write in riddling epithets like the famous one about it being impossible to step into the same river twice, decided that as 'all is flux', it is ultimately by the power of the mind, which can contemplate 'what is not', rather than by senses forever limited to examining merely what is, that the truly important things can be found.

Ptolemy (87–150 cE), the inspiration of future mathematicians and geometers, as well as geographers and cosmologists, describes his homely view of the universe in the first book of the *Almagest*, and gives various arguments that sit somewhere between 'thought experiments' and real experiments. In particular, Ptolemy argues that since all bodies fall to the centre of the universe, the Earth must be fixed there at the centre, otherwise falling objects would not be seen to drop towards the centre of the Earth. Now his listeners could at this point have conducted their own 'real' experiments but it was clearly enough to reflect on their own underlying assumptions of

reality to agree with Ptolemy. Only the fact that the first assumption is rather a big (and rather a dubious) one stands between them and true knowledge, a fact that is certainly a reminder of the dangers of the thought experiment technique. But then again, it is also a reminder of the dangers of the 'thoughtless' experiment technique too. It was not the lack of testing that was a problem with Ptolemy's theory; it was the assumptions underlying it.

Another of Ptolemy's experiments is designed to show that the Earth must not only be at the centre of the universe, but completely motionless – steady as a rock – too. To do this, Ptolemy asks us to consider the fact that that if the Earth moved, as some earlier philosophers had suggested, then certain bizarre consequences would have to follow. In particular, if the Earth rotated once every 24 hours, was it not intuitively obvious that an object thrown vertically upwards would not fall back to the same place, but would fall back slightly to one side?

Ptolemy's record is not encouraging, but then his experiments were not truly thought ones. For Plato, as for Heraclitus, those wishing to understand phenomena in the natural world should recognize that experience of events is a poor guide. Plato's dialogues are littered (for want of a kinder term) with thought experiments. Alongside Gyges with his magic ring exploring the nature of morality is the 'mad friend' hunting for his knife; elsewhere there is Plato himself conducting the (less well known) 'breeding experiment' in which he advances the case for eugenics for the good of society; and over there is the much repeated but little agreed upon metaphor of the prisoners in the cave, that seems to be telling us something about the nature of knowledge. Less often appreciated, but still very influential, is that the entire process of the development of society outlined in the *Republic* is actually a carefully crafted thought experiment, built on the assumption that people will not be content with all the fruits of nature - but will want to eat meat, leading to a struggle for land and resources.

But it is in Plato's account of Socrates leading the slave boy, Meno, to develop the Pythagorean principle that new knowledge appears to emerge from introspection in the best manner of the technique. Many of the Ancients valued such 'pure' knowledge, quintessentially mathematical, more highly than any that relied on actually going out and looking at real things, and the notion of 'truths' waiting to be discovered by contemplation is appropriately sometimes dubbed 'Platonist'. Thought experimenters are his fellow travellers.

Even Aristotle, who like a certain kind of scientist usually maintained the supremacy of observation, tried one or two thought experiments. In his *Metaphysics* (Book VII, iii) for instance, he offers the experiment of two individuals, Plato and Socrates, having their 'non-essential' properties stripped away leaving only their 'essence'. How many essences are there, he asks? One or two?

But Aristotle's importance in the history of the technique lies less in his use of it than in his provision of a wealth of poorly judged and false beliefs about the physical world. As Bertrand Russell has pointed out, Aristotle, in spite of his reputation, is full of absurdities. For instance, Aristotle insists that the blood of females is blacker than that of males; that the pig is the only animal liable to measles; that an elephant suffering from insomnia should have its shoulders rubbed with salt, olive-oil and warm water; and that women have fewer teeth than men. But there have also been more weighty opinions about gravity, time and space that subsequent philosophers and scientists have had to labour mightily to demonstrate the error of. And often the most compelling refutations have been not empirical, but conceptual, using the thought experiment technique. (Although not, admittedly in the matter of numbers of teeth. There is still a place for observation.)

Medieval philosophers, for example, used the idea of a lance with a sharpened handle (that is, as well as a sharpened point) to disprove Aristotle's theory that things like lances only flew through the air when released, rather than falling bemusedly straight to the ground in shocked realization that they were no longer being held, by virtue of the 'pressure' of air rushing in behind them. (The air's ability to press on the sharpened handle did not seem to them likely to affect the lance.) The medievals particularly valued such thought experiments in their debating technique of 'challenges', which used all kinds of 'common sense' experiments to settle disputes. In a formalized process known as 'obligationes', disputants were 'obliged' to either assent, dissent or doubt statements, until such time as a 'contradiction' was demonstrated in one or the other's position. A double-pointed lance, although easy enough to produce to the court, was not necessary.

But it was the Renaissance that produced the richest crop of thought experiments, including those of Galileo, Descartes, Newton and Leibniz. These were thinkers whose interests lay in 'Natural Philosophy' and who considered that the best experiments work by making conscious and obvious what any assumed laws of nature really are. Descartes used the technique particularly enthusiastically, offering in his *Meditations* (1641) the original 'brain in a vat' scenario, along with a 'possible world' peopled by automata, another run by a 'malicious Demon' (along with the general philosophical problem of whether we might all be dreaming) and finally the solitary introspection in the celebrated Second Meditation. It is there he finds that he cannot even imagine thinking away thinking and so is led to the conclusion that the only certain thing is thought itself.

Descartes took it for granted that whatever can be imagined is possible, in some sense. Yet although this might appear to give the imagination extraordinary power, he also insists that mere mortals are bound forever to the laws of logic, unable to even imagine a world in which, say, 2 + 2 did not equal 4. (Although God, Descartes piously adds, is above these laws.) Thought experimenters who dare to suppose an illogicality enter dangerous waters in which, even if they manage to survive, any findings they may eventually return with are worthless. Yet just what is 'illogical'? In another experiment, Descartes says if you remove all the matter from a chamber, the walls would touch, therefore a vacuum is impossible. Perhaps then 'imagining the impossible' is not always so foolish a thing.

Hume, like Descartes, considered that 'conceivability' equals possibility and that things which cannot even be imagined definitely could not be possible:

'Tis an established maxim in metaphysics that whatever the mind clearly conceives includes the idea of possible existence, or in other words, that nothing we imagine is absolutely impossible. We can form the idea of a golden mountain, and from thence conclude that such a mountain may actually exist. We can form no idea of a mountain without a valley, and therefore regard it as impossible.

Sometimes (but, as we have already seen, quite erroneously), the philosophical examination of thought experiments is only traced back to the Danish scientist, Hans Christian Oersted (1777–1851). Oersted saw them as not so much concerned with predictions or substituting for measurement, but as a tool for arriving at a better understanding of nature. For him the value of the technique lay in first of all supposing some kind of 'law of nature' and then asking the experimenter to apply the law in a new – perplexing – setting. This was at a time when German philosophers such as Johann Fichte (who had chosen for himself the task of finding a transcendental explanation for consciousness) or Friedrich Schelling (nowadays hailed as the father of 'post-metaphysical thinking') were indulging in 'speculative philosophy'. Oersted himself has been said (perhaps like Kant too) to have been looking for 'a middle way' between blind laboratory science and fruitless metaphysical speculation.

Certainly, in the history of science, the thought experiment has to be acknowledged as a scientific method in its own right. Galileo did not actually drop balls off the leaning tower of Pisa – it was a thought experiment. (Despite what some may say today, see experiment G.) Similarly, Leibniz's procedure for refuting Descartes' Law of Collision does not require the rolling around of variously sized billiard balls; the thought experiment alone is rightly seen as settling the matter.

Although it is not perhaps a *very* interesting experiment, it is a good example of its kind. Descartes thought that if a smaller object hit a larger one it would rebound with equal speed, and that when a larger object collided with a smaller one, then the two would move off together (in a way that conserves the total quantity of motion). Leibniz, however, asks us to imagine a series of collisions, in which one ball starts by being smaller, but the ball it is hitting is shaved down imperceptibly until the first ball actually becomes slightly the larger. At this point, according to Descartes, the behaviour of the two balls radically changes. But it seems ridiculous to suppose that such a tiny change in the ball's mass could result in it one minute bouncing off, and the next propelling the other onward, so Leibniz seems to show that Descartes is wrong.

And much of modern physics is built not upon measurement but on thought experimentation. Einstein did not carry out measurements in a rapidly descending elevator, nor did Schrödinger actually put his cat into a box with a radioactive rock; all were sufficient in themselves just as hypotheticals. They are quite *possible* as practical exercises, but the point of a thought experiment is that it really would not help to carry it out: all the information that is necessary is already there, as it were, in the hidden recesses of consciousness. And in fact Galileo, Newton, Darwin and Einstein all used them to great effect to resolve, not just explore, complex issues and scientific debates. They conjured up scenarios, obliged others to follow the logic of the tale and ultimately accept their findings. These were quintessentially experiments that took place truly in the 'laboratory of the mind'.

Einstein, a past master, used the technique to imagine what it would be like to travel at the same speed as a light ray. If you were to run down a pier, he mused, as a big wave was coming in, then the watery wave would appear to you as a stationary lump in the water. What then, for an astronaut racing a light wave – would it too appear to be stationary (experiment U)? In another thought experiment, a physicist has been drugged and wakes up in a box being pulled steadily upwards by a rope. Into this box a beam of light is projected. The 'elevator' as it became known, is designed to demonstrate the equivalence of constant acceleration and gravitational field effects, by showing that the light ray will appear to bend in both cases. From such simple musings would come the special theory of relativity. As Einstein wrote later, 'from the very beginning it appeared to me intuitively clear that, judged from the standpoint of such an observer, everything would have to happen according to the same laws as for an observer who (relative to the earth) was at rest. For how, otherwise, should the first observer know, i.e. be able to determine, that he is in a state of fast uniform motion?'

Some say all this is a little too good to be true. They worry that whilst the approach seems to offer the advantage, through being made up as you go along, of allowing extraneous detail and complications to be removed, it may equally in so doing cease to be relevant or accurate. As Wittgenstein put it, 'it is only in normal cases that the use of a word is clearly prescribed, we know, we are in no doubt, what to say in this or that case. The more absurd the case, the more doubtful it becomes.' If you imagine things differently from 'the way they are', he adds, then 'you can no longer imagine the application of certain concepts.' This sits a little uncomfortably alongside, of course, his own liberal use of the technique, including examples such as the one describing a man whose brain is removed by surgeons, and another where we are asked to imagine a world in which all human beings look exactly alike (compare with experiment I), not to forget numerous other, at least 'quasi', thought experiments supposedly highlighting aspects of language. Ones such as the comparison with the controls of the train engine, or the one with a map of a street accident, or most elaborately of all, the 'beetle' that everyone carries furtively around in a small box (see experiment W). But then Wittgenstein (or at least the 'later Wittgenstein' as aficionados put it) did believe that language is best undersood as a series of pictures, and his thought experiments are also, in their way, only the logical manifestation of that approach.

In any case, the accusation of 'abnormality' also sits rather uncomfortably alongside the rich history of mathematics, where impossible entities appear without so much as a raised eyebrow – be they dimensionless points, perfect circles, negative and irrational numbers or whatever. Mathematics is after all one of the main sources of thought experiments in a tradition stretching from the ancient philosophers up to the present day. Those proposed by Bertrand Russell, Gottlob Frege and others to resolve the so-called set paradox (a debate in which Wittgenstein himself was closely involved) are rightly recognized as central to both the philosophical and the mathematical debate. And here, in the concluding 'How to' guide, we consider mathematically (but only in the most elementary mathematics!) a kind of meta-thought experiment designed to show how the technique might quite legitimately conjure up 'new' information from old facts and established assertions. In fact, mathematics and physics operate with different rules, and should be kept apart to some extent. Physics is empirical, based on measurement, but mathematics is based on 'axioms' that are assumed at the outset. Having said that, nowadays, physicists, if not philosophers, see even mathematical knowledge as provisional and flawed. Indeed, Einstein once wrote: 'as far as the propositions of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.'

But for some philosophers, uncertainty and provisional knowledge are always going to be a bad thing. Many would-be moral scientists, such as Utilitarians equipped with simple rules based on the maximization of happiness principle, worry that thought experimenters parade 'a succession of bizarre cases' which actually warp judgement. Far from investigating moral intuitions, as they may imagine they are doing, they are replacing complexities with simplicities, while always supposing that doing so makes no difference. One critic, Alisdair McIntyre, has also objected that ethical thought experiments are 'ahistorical' in that they are detached from their original origins and debates. Others have warned that the thought experiments become like a stage play in which we are asked to become actors trapped into endlessly repeating the same scenario, proving nothing. Marxists disparage 'soft escapism', and insist that philosophers stick to practical issues.

Still others object to the reliance on intuitions, and return to the age old concern of the relationship of 'conceivability' to 'possibility', the debate that used to feature so prominently in attempts to work out whether God really did exist. But the same concerns apply to more immediate questions raised by thought experimenters. For example, on a medical theme, at what point would a person cease to be alive if body parts were progressively taken away? Such experiments appear entirely conceivable, but perhaps it is an illusion of conceivability, an unwise and fruitless adventure into hypothesis. This is certainly what Ernst Mach, who used the technique himself, meant when he complained about Newton's famous 'Bucket' experiment (experiment N), a generally mundane account of a bucket on a rope in which, nonetheless, Newton surreptitiously imagines the whole of the universe away. As Mach commented drily: 'When experimenting in thought, it is permissible to modify unimportant circumstances in order to bring out new features in a given case, but it is not to be antecedently assumed that the universe is without influence on the phenomena in question.'

The most implacable enemy of the technique, however, has been a certain kind of traditional 'analytic' philosopher, apparently concerned that he or she is being encouraged to infer conclusions from 'intuitive reactions' rather than by a sound process of rational deduction. (Although funnily enough, another philosopher, Richard Rorty, has said that thought experiments are 'circular' because our beliefs determine what happens in them. Perhaps the concern is that the technique is stepping on the toes of analysis, as there the aim is to start by assuming something, examine it a bit, and then conclude with your initial assumption, having apparently forgotten that that is what it was. No *new* information can ever be obtained by 'analysis', as any logical pedant can tell you.)

Certainly it seems these days that many contemporary philosophers' main interest is in debunking the technique. Thus, a paper in the journal *Ethics* by Tamara Horowitz spills much ink in the cause of denying the hapless Warren Quinn the right to use the technique to draw conclusions about ethical values. His experiments revolved around some 'rescue dilemmas', the infamous imaginary underground trolleys packed with different assortments of people heading forever to various sorts of disaster. In Quinn's examples, the numbers of people being rescued (or run over) stay the same, but the circumstances and indeed the language used to describe them are varied. Horowitz points out that people are inclined to forgive unfortunate effects if they are described as incidental whilst condemning those in which the consequences are described more explicitly.

In some respects, this is only common sense, but to Tamara Horowitz it shows rather that responses to thought experiments in general, and ethical dilemmas in particular, will be influenced by the wording, or what is sometimes termed the 'framing', of the question. This at least is something we must be aware of both in considering other people's, and even more so in designing our own, thought experiments, and is a point we shall return to in the 'How to' section at the end of this book.

On the other hand, there is nothing in the technique that says the experimenter is not allowed to change the wording if in fact it is shown to be skewing the results. Moreover, in a stout (if still also somewhat wishy-washy) defence of the technique, in the *Journal* of the Theory of Social Behaviour, Francis Roberts argues that at least thought experiments allow for investigations to be carried out without 'disturbing the environment' in the process, while elsewhere Jonathan Dancy, in an account of 'The Role of Imaginary Cases in Ethics', says that they can be just as good as real examples, particularly when fleshed out with details, even if they suffer from 'a certain indeterminacy' not to be found in reality.

Such as in Franz Kafka's story, *Metamorphosis* (1915), which describes what it would be like to find you had woken up and your

body had turned into that of a giant insect. If you believe in reincarnation, the possibility is quite real, of course, and in the story there is also a ready made 'mind-transfer' built in, of the kind that has spawned so much vigorous recent philosophical debate. Still, even ridiculous stories may tell us more about both our intuitions and assumptions, be they methodological or ethical, than those handcuffed to reality.

That was certainly the feeling of Charlotte Perkins Gilman, whose literary utopia *Herland* (published coincidentally in the same year as *Metamorphosis*) described how three male explorers stumble upon an all-female society in a distant part of the earth. Many generations earlier, this commune had become separated from the rest of the human race, with the men dying off. The society had evolved in a distinctively feminist way, organizing itself around raising children and living in harmony with its surroundings.

Originally published at the time of the Suffragettes in England and campaigns in Europe and America for women's equality, Charlotte Gilman's story is a vehicle for her view of male/female roles and behaviour, of motherhood, individuality, sexuality, and other topics as well. But more recent feminists have had their doubts too: that the approach utilizes a 'restrictive male form of thinking' that should instead be trained into a more holistic, inclusive and co-operative mode. Or that it elevates 'abstract principles over contextual solutions', as Carol Gilligan has put it.

This is a little unfair. One of the most celebrated of recent ethical thought experiments comes courtesy of Bernard Williams, in which he imagines a man, Jim, who arrives in the town square of a South American republic, to find 20 locals there, firmly tied up, and standing over them the Captain of the local militia. To Jim's alarm he declares that he has just quashed their rebellion and is going to shoot them. Unless, that is, perhaps Jim, as a distinguished visitor, would like to shoot the first one – in which case the rest can all go free.

This experiment is intended to challenge precisely that sort of 'elevation of abstract reason' (in this case, utilitarianism) and kind of 'contextless' thinking. In running the experiment we begin to doubt whether such matters can be swiftly resolved by calculating the 'amount of happiness' that results, and overcoming scruples in order to save the 19 unfortunates. We have also to consider whether, much as we might like to help, we would be able to justify to ourselves the sacrificing of the first one.

But in any case, other feminists recently have not been above using the approach themselves – and with some celebrity. Sissela Bok, in her book *Secrets: On the ethics of concealment and revelation*, used a thought experiment with four imaginary societies in order to examine the issues of secrecy and confidentiality in activities such as psychotherapy and spying. In experiment V we will see Judith Jarvis Thompson invent an 'unconscious violinist' to test assumptions over the ethics of abortion, this one of the most successful and fruitful of contemporary experiments. To join in the debate, Mary Anne Warren, for example, unplugged the violinist's unfortunate neighbour and asked for a volunteer instead, while Roger Wertheimer came up with a scenario asking what would people think if wombs were transparent.

And like the feminists, whatever their personal doubts, even the most analytical of philosophers have also been unable to keep off the thought experiments. The great German logician, Gottlob Frege, overcame any qualms to conjure up the possibility of a 'rational tribe' who had an 'alternative logic' before concluding, sadly, that such a tribe could no longer be counted as 'rational'. (For Frege, like his countryman Immanuel Kant, it was important to demonstrate that a rule of logic is binding on all of us, from whichever tribe.)

Amongst other more recent efforts are Peter Strawson's imaginary 'world of sounds' in which 'position' is determined in some complicated way by the gradually changing pitch of what is called the 'master-sound' (this is discussed further in the 'How to' guide) and Martin Hollis's strange village (strange, too, in its resemblance to one imagined earlier by Gottlob Frege) in which anthropologists struggle to translate their language for fear that 'the natives' may use a different kind of logic to our own. Then there's Brian Ellis who wondered whether if the universe had just one thing in it, which he calls 'e', perhaps hoping to nominate himself (Essence of Ellis), that object could still have 'quantitative' properties? How big, for example, can Essence of Ellis be considered to be, when there is nothing to compare it with? No measures or rulers, no trees, no nothing.

Or there is Anthony Quinton's (perhaps rather feeble) effort to imagine the mind of Winston Churchill in the body of a 6-year-old girl (with a view to countering the notion that certain character traits require certain physical prerequisites) and of course John Searle's Chinese Room. This last has become quite a regular both in mass media discussions and on the lecture circuit, and I myself have not shied away from offering a 'souped-up' version here as the second part of experiment R.

John Searle originally introduced his 'Chinese Room' thought experiment (then titled, 'Minds, Brains and Programs') with the explanation that 'one way to test any theory of mind is to ask oneself *what it would be like if* my mind actually worked on the principle that the theory says all minds work.' He hoped his experiment would persuade people to the opinion that theories of Artificial Intelligence that award computers human-like thinking skills are 'bunkum' or, at the very least, 'implausible'. Other analytic philosophers, however, while sharing this aim by and large, cannot accept the proof offered by a thought experiment anyway. One such, Daniel Dennett, objects that as the Chinese Room scenario is not an argument, it therefore cannot be 'sound'. Unsound, that is, in the sense that even if all the assumptions made in it were acceptable, no one would be logically bound to accept its conclusion. Of course, this cuts both ways: even if you found Searle's conclusion quite ridiculous, there is no logical reason in that to suspect either the assumptions or the method followed.

But while writing a special introduction to a new *Journal of Artificial Life*, Professor Dennett also observes that philosophers have always 'trafficked' (as he unkindly puts it) in thought experiments. These techniques, he noted, are 'notoriously inconclusive'. 'What "stands to reason" or is "obvious" in various complex scenarios is quite often more an artefact of the bias and limitations of philosopher's imagination than the dictate of *genuine logical insight*' he says. Yet even so, there is hope for the technique. Searle may turn on his swivel chair, but for Dennett at least *the computer* can make the philosophers' thought experiment worthwhile. By modelling hypotheses on computers Dennett sees the whole exercise as somehow validated. As he concludes: 'Philosophers who see this opportunity will want to leap into the field, at whatever level of abstraction suits their interests, and gird their conceptual loins with the simulational virtuosity of computers.'

Of course this is dreamy nonsense. Philosophers should leave computers, like video recorders and photocopiers, well alone. It is only worth mentioning here to show that even the most hard-nosed 'analytical' philosophers actually believe in the value of hypothesis making and testing – which is strictly speaking that most evil form of illogicality, inductive thinking. The thinking, some readers may recall, that led Bertrand Russell's unfortunate chicken to waddle down out of the coop expecting a handful of tasty grain on the day that the farmer was planning a special dinner...

Much of today's debate over the validity of thought experiments centres on this issue. The distinction is between a technique that can provide new knowledge and one that can only present old knowledge a new way. On the one side philosophers such as James Brown say thought experiments provide what they like to call *a priori* knowledge of natural phenomena, such as the mathematical entities or 'laws of nature' that mathematicians and physicists wrestle with, and which may or may not be 'out there' in the world of the Forms; while on the other, philosophers like John Norton stoutly maintain any knowledge obtained from them is not new, not 'discovered', but merely disentangled from where it is already lurking (in the disorganization of the imperfectly logical mind).

In fact it was John Norton who once grandly defined thought experiments as arguments which:

- 1 posit hypothetical or counterfactual states of affairs and
- 2 invoke particulars irrelevant to the generality of the conclusion.

But grand though it sounds, to say that a thought experiment is 'hypothetical' is to say precisely nothing, whilst to add 'or counterfactual' is to play at words. In allowing something to be either counterfactual or factual we are not venturing very much. In fact, some thought experiments are counterfactual, but many others seek on the contrary to demonstrate some facet of reality through all the elements being entirely possible, if not necessarily plausible. This leaves only the observation that thought experiments include irrelevant details as the definition which seems (in a possibly interesting way) to be exhibiting precisely the fault that it claims of the thought experimenter.

As to that, consider Galileo's Ship argument (experiment S) with its cutesy details such as the fish that swim towards the front of their bowl or the butterflies that continue their flights indifferently towards every side. It is guilty on all counts. Yet it is also the basis for much of modern physics, and created a world in which dogmatic assertion began to weaken. Up until then, as a result of such arguments, the geocentric system was forcibly thrust upon all philosophers and scientists by theologians relying not only on the guidance of divine texts but the apparent certainties offered by 'science'. Yet, for Galileo, the approach of the thought experimenter offered more certainty and greater validity than any number of measurements or predictions (leaving well alone the untouchable authority of the scriptures).

And so, if even today those who follow in Aristotle's footsteps are baffled by the whole approach, it is perhaps only the more reason to revisit the great discoveries and debates of thousands of years of experimentation in the 'laboratory of the mind'.