Counter Thought Experiments

JAMES ROBERT BROWN

Introduction

Let's begin with an old example. In *De Rerum Naturua,* Lucretius presented a thought experiment to show that space is infinite. We imagine ourselves near the alleged edge of space; we throw a spear; we see it either sail through the 'edge' or we see it bounce back. In the former case the 'edge' isn't the edge, after all. In the latter case, there must be something beyond the 'edge' that repelled the spear. Either way, the 'edge' isn't really an edge of space, after all. So space is infinite.

This example is typical of thought experiments in general. We set things up in the imagination, we let it run, we see what happens, and we draw a conclusion. It's also quite similar to a real experiment, except that it's done in the imagination rather than in the real world. And like real experiments, thought experiments are fallible. In this case we would now make a distinction between *unbounded* and *infinite*, so that the conclusion Lucretius drew, we now clearly recognize, does not follow from what went before.

Lucretius is but one of many different types of thought experiments. Positive or constructive thought experiments support some theory, while negative or destructive ones undermine. My interest is in a special class of negative thought experiments that I shall call 'Counter Thought Experiments'. I'll largely ignore other types, except for the sake of contrast.

Examples of Negative Thought Experiments

One type of destructive thought experiment shows some existing theory to be self-contradictory. Einstein chased a light beam with a view to see what the wave front looked like. If we were to run on a pier toward the shore at the same speed as an incoming water wave, we would see a static hump in the water. Perhaps we would have a similar experience in the case of light, since light, according to Maxwell's theory, is an electromaganetic wave. The light wave, however, is dependent on change: a changing magnetic field gives

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

rise to an electric field, and a changing electric field gives rise to a magnetic field. When Einstein catches up to the front of the light wave, he would see static fields, but then a light wave cannot exist.

Galileo reduced Aristotle's theory of motion to an absurdity in a rather simple but ingenious way. The first part of his wonderful thought experiment on falling bodies is a typical *reductio ad absurdum*. Aristotle claimed that heavier bodies fall faster than light ones (H > L). Suppose we attach a heavy body and a light one together. Then the combined object must fall faster than the heavy one alone (H+L > L). But the light component of the combined body will act as a drag, slowing the whole thing down so that it is actually slower than the heavy body falling alone (H+L < L). Thus, we have an absurdity, and Aristotle's theory is destroyed. Galileo's thought experiment then takes a second step, this time a positive one. It becomes obvious how bodies must fall, given the way the absurdity was achieved. All bodies must at the same rate (H = L = H+L).

Showing an internal contradiction is not the only way to undermine an existing theory. Some thought experiments show the theory to be contrary to other established (including common sense) beliefs. Schrödinger's cat is a prime example. Schrödinger took the wierdness of the Copenhagen interpretation of quantum mechanics at the micro-level and brought it to the macro-world of everyday objects such as a cat. It was bad enough that an atom could be in a superposition of two different states (e.g., energetic and decayed), but the consequence of that view seemed to imply that even a cat could be in a superposition of living and dead. This is not a contradiction. Some physicists (Wigner) actually were willing to accept it. But it is a gross violation of common sense.

In argument terms, these thought experiments show the premisses false. That is, they show that the theory in question must be false. In the first case just mentioned, the thought experiments show that there is something wrong with the conjunction of electrodynamics and basic assumptions about moving reference frames. In the second, Aristotle's view that bodies fall at rates related to their weights is wrong. In the third, the target is the Copenhagen interpretation that allows physical systems to be in reality in states of superposition.

Of course, these are debatable outcomes. One could try to save the initial theory by putting the blame on something else. Maybe there's a difference between genuine bodies and the composite bodies of Galileo, with the true laws applying only to the genuine ones. Maybe there is a micro-macro distinction with atoms going

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

into states of superposition, but not cats. As with any real experiment, there is lots of room for rival interpretations, not to mention outright mistakes.

A second type of negative thought experiment shows a situation that undermines a crucial inference. It does not challenge the premisses the way the first type of negative thought experiment does. In terms of logic, this class of thought experiment aims to show invalidity (i.e. the premisses may be true but the conclusion does not follow from them). Consider the kind of thought experiment we would present to undermine Lucretius's thought experiment for infinite space. Imagine that we are two-dimensional bugs living on a sphere. Every time we throw a spear at an alleged edge of space it passes through or bounces back because of some barrier. In either case we would agree with Lucretius that this is not the edge of space. However, the inference that space is infinite would be clearly false, since the sphere is finite. Poincaré's disk people example works in a similar way. The measuring rods of the disk people shrink as they move toward the edge of their space, so that they might even come to mistakenly think that they live in an infinite space. If they threw spears, those spears would behave just as Lucretius says. But the disk is finite.

The third type of negative thought experiment—the one I am chiefly concerned with here—is the *Counter Thought Experiment*. The balance of this paper will be devoted to describing them and trying to determine some of their main properties. As examples, I will discuss three:

Galileo against the Aristotelians (principle of relativity)
Mach against Newton (absolute space)
Dennett against Jackson (physicalism)

One of the most interesting features about counter thought experiments is that they are not readily understood in terms of the logic of argument; that is, they are not about validity or soundness. They are directed at a given thought experiment, but they challenge neither the premisses nor the concluding inference. Instead, counter thought experiments deny the phenomena of the initial thought experiments.

Boundaries of an experiment

Experimenters do a great many things. They set up their equipment; they let it run and see what happens; they measure;

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

they calculate; they interpret; and they draw some conclusions which they publish. It is not easy to draw the boundaries of an experiment. The distinction between theory and observation, for instance, is fuzzy at best and the case has been well-made that observations everywhere are theory-laden. I readily grant this, but wish to focus on something a bit more mundane, a distinction between experiment in the broad sense and in the narrow sense.

In the narrow sense, an experiment includes the set up and the observation (which may be highly theory-laden). In the broad sense, the experiment includes background assumptions and initial theorizing, the setup, observation, additional theorizing, calculating, and drawing the final conclusion. It is this final result, with an account of how it was obtained, that we read in a journal. I doubt there could be a sharp distinction drawn between narrow and broad experiment. And what goes for real experiments goes for thought experiments, too. But there is a rough and ready distinction. The narrow part is the phenomenon, it is what we see. We could put this in a simple schematic way:

Theory & Background → Phenomenon → Result

The narrow sense of experiment (whether real experiment or thought experiment) is what we observe, the phenomenon, the middle of the schema. The broad sense includes the whole thing from theory and background assumptions to the final result.

Looking at the simple schema, it is obvious that challenges could come at different points. (NB. The arrows just mean 'is followed by', but for some purposes they might be taken loosely as deductive or inductive implications.) One could challenge the assumptions that played a role in the set up, that is, the theoretical and other background assumptions that went into it. This is what the first class of negative thought experiments do; they attack the premisses. One could also challenge the inference to the final result from what went before. This is what the second class of negative thought experiments do; they attack the alleged validity. But obviously, there is also a third way, one could challenge the phenomenon of the thought experiment; that is, one could claim that the phenomenon does not occur, or that what is observed is quite different from what was initially claimed. Let's illustrate this with some examples of each of these types of challenges. In the first set of examples to follow, the phenomena is never at issue; they are *not* examples of Counter Thought Experiments. I include them to provide a useful contrast.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

Lucretius, Searle, Thompson

Two thought experiments undermine Lucretius. They both work the same way; they accept the background beliefs of Lucretius, they accept the way the thought experiment is set up, and they accept the observations, as described. They reject the conclusion. Both provide a situation where the background and the phenomenon are as Lucretius wants, but the conclusion of infinite space is false. As I said earlier, it's similar to the way one might show a deductive argument to be invalid: provide an interpretation in which the premisses are true but the conclusion is not.

I mentioned two examples above: bugs on a sphere would not encounter an edge to their space, but their universe is finite, nevertheless. The example illustrates the distinction between infinite and unbounded, a distinction that Lucretius and others would not easily recognize until the rise of non-Euclidean geometry and modern topology.

The second example, only briefly mentioned, is more complex, but also instructive. Poincaré asks us to imagine three-dimensional beings like us inside a finite sphere. It is easier to switch this (as is commonly done) to two-dimensional beings living on a finite, flat disk. The peculiar thing about their world is that there is a force, a bit like heat, that makes all objects expand or contract as they move around the disk. The crucial thing is that *all* objects undergo this contraction as they move toward the edge, so it is utterly unobservable to the inhabitants.

The disk has a radius R and objects contract as they move toward the edge in proportion to $(R^2—r^2)/R^2$ (where r is the distance from the centre). So, if an object has length L at the centre, then its length at a distance r from the centre is $L \times (R^2—r^2)/R^2$. At the edge it shrinks to zero. These distances are as measured in the so-called embedding space, the Euclidean space in which we imagine both the disk and ourselves (with our god's eye view) to be located. If the two-dimensional beings measured their universe, they would find that it took infinitely many lengths of their measuring rods to get to the edge, so they might reasonably conclude that they lived in an infinite universe.

The original point of Poincaré's example had nothing to do with Lucretius. He wanted to show something important about how choices are made when we try to establish the geometry of our universe. Poincaré's disk people would find that the sum of the interior angles does not equal 180 degrees, as in Euclidean geometry, but rather would find that the sum is less than 180

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

degrees. So they might reasonably conclude that they live in a Lobachevskian (or hyperbolic) universe. The consequence for the status of geometry, according to Poincaré's is this: It is a conventional choice, based on practical considerations, influenced by experience but not determined by it.

I'm only using the first part of Poincaré's thought experiment, the part that involves the experience of the disk people who shrink along with their measuring rods as they move to the edge of their universe. If they threw a spear, it would sail through any alleged edge of space. But, clearly, the inference they might be tempted to make, namely, that space is infinite, is wrong.

These two negative thought experiments both accept the set up and the phenomenon of the initial Lucretius thought experiment (i.e., we never come to an edge). They deny that the conclusion (infinite space), follows from this. They (in effect), attack the validity of Lucretius's thought experiment.

John Searle and Judith Thompson have produced two of the most famous thought experiments of recent times. Searle's Chinese room thought experiment imagines a person in room with an input slot and an output slot through which pass messages in Chinese writing. The person inside has a book that tells him, on a given input, what the output should be. This set up would pass the Turing test; that is, it can think, according to the view of AI (strong artificial intelligence). But, Searle claims, obviously, the person doesn't understand Chinese at all.

There have been numerous challenges to this thought experiment, but none attack the phenomenon. No one denies that there could be such a man in a room receiving and sending messages in Chinese in accord with a book of instructions, yet not understanding the Chinese messages at all. Such an attack, were one to exist, would be what I call a counter thought experiment. Instead, the challenge is usually that Searle has drawn the wrong inference. One claim, for instance, is that it is not the man in the room that passes the Turing test, rather, it is the whole system: room, instruction book, and man. And it is the whole system that understands Chinese, not any part of it, such as the man alone.

Thompson imagined a person hooked involuntarily to a famous violinist (who happened to be unconscious of what happened). The violinist is innocent and has a right to life. The healing process will take nine months, connected all the while; and he would die without being connected for this duration. Though it might be a very generous act to donate one's life for this period, it is surely not morally required. Abortion is analogous to this and so, abortion is

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

morally permissible, even thought (for the sake of the argument), it is granted that the fetus is an innocent person with a right to life. Thompson's thought experiment helps us to make a conceptual distinction: 'right to life' does not equal 'right to what is needed to sustain life.' The violinist/fetus has the former, but not the latter, which is why abortion is morally permissible.

This thought experiment has been repeatedly criticized and rejected, but attacks have not attempted to deny the possibility of actually finding one's self hooked to a violinist who must remain connected for nine months in order to survive. In short, the phenomenon of the thought experiment is not challenged.

In each of these cases, Lucretius, Searle, and Thompson, the challenge has not been directed against the phenomenon, but rather at some other point in the thought experiment. The phenomena in each of them has been undisputed. I turn now to the interesting cases where this is not so, that is, to cases when the phenomenon of a thought experiment has been the focus of attack.

Counter Thought Experiments

As I mentioned above, there are three examples of counter thought experiments that I want to discuss at length. First, Galileo denies Aristotelian thought experiment concerning moving earth; second, Mach denies Newton's thought experiment concerning absolute space; and third, Dennett denies Jackson's thought experiment concerning physicalism.

1. Galileo against Aristotle

From the time of Aristotle through the middle ages, there was a commonly used Aristotelian thought experiment to show the earth could not move. Suppose, on the contrary, that the earth does move. Then a dropped object would fall behind us as we move along; it would not fall straight down to our feet. But, as a matter of fact, it does fall straight down. Thus, the supposition must be false; the earth does not move.

Galileo put forward a counter thought experiment. Not only is it a gem in its own right, but it played a huge role in the development of physics. It established, in effect, the principle of relativity (often now called Galilean relativity).

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

> Shut yourself up with some friend in the main cabin below decks on some large ship, and have with you there some flies, butterflies and other small flying animals. Have a large bowl of water with some fish in it; hang up a bottle that empties drop by drop into a wide vessel beneath it. With the ship standing still, observe carefully how the little animals fly with equal speed to all sides of the cabin. The fish swim indifferently in all directions; the drop falls into the vessel beneath; and, in throwing something to your friend, you need throw no more strongly in one direction than another, the distances being equal; jumping with your feet together, you pass equal spaces in every direction. When you have observed all these things carefully (though there is no doubt that when the ship is standing still everything must happen in this way), have the ship proceed with any speed you like, so long as the motion is uniform and not fluctuating this way and that. You will discover not the least change in all the effects named, nor could you tell from any of them whether the ship was moving or standing still. (*Dialogo* 186f)

Galileo's thought experiment denies the phenomenon of the Aristotelian thought experiment. If the earth were moving, a dropped object would land at our feet, not behind us at the initial thought experiment declared. Of course, this does not establish that the earth is indeed moving. The Aristotelian conclusion of a stationary earth might be true. But it does show that the empirical evidence we have is compatible with a moving and with a stationary earth. The Aristotelian thought experiment fails, since things would look the same regardless. This is a counter thought experiment. It denies the phenomenon (that objects would fall behind a moving earth), in the original thought experiment.

2. Mach against Newton

The background to Newton's famous thought experiment concerns rival understandings of the nature of space. Newton's absolutism (often called 'substantivalism'), is the view that space is a substance that exists without depending on anything else. It is the source of inertia. Relationalism is the standard rival view: space is a system of relations. If there were no bodies, there would be no space.

Leibniz, of course, is the prime representative, though he expressed his views most clearly only after Newton's thought experiment.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

> I hold space to be something merely relative, as time is; that I hold it to be an order of coexistences, as time is an order of successions. For space denotes, in terms of possibility, an order of things which exist at the same time, considered as existing together; without enquiring into their manner of existing. And when many things are seen together, one perceives that order of things among themselves. (Leibniz, *Leibniz-Clarke Correspondence*, 25f)

A 'Leibniz shift' would be moving the whole universe to the right, or mirror reflecting it, etc. But such a thing, Leibniz claimed, is impossible.

> I say then, that if space was an absolute being, there would something happen for which it would be impossible there should be a sufficient reason. Which is against my axiom. And I prove it thus. Space is something absolutely uniform; and, without the things placed in it, one point of space does not absolutely differ in any respect whatsoever from another point of space. Now from hence it follows, (supposing space to be something in itself, besides the order of bodies among themselves,) that 'tis impossible there should be a reason, why God, preserving the same situations of bodies among themselves, should have placed them in space after one certain particular manner, and not otherwise; why every thing was not placed the quite contrary way, for instance, by changing East into West. But if space is nothing else, but that order or relation; and is nothing at all without bodies, but the possibility of placing them; then those two states, the one such as it now is, the other supposed to be the quite contrary way, would not at all differ from one another. Their difference therefore is only to be found in our chimerical supposition of the reality of space in itself. But in truth the one would exactly be the same thing as the other, they being absolutely indiscernible; and consequently there is no room to enquire after a reason of the preference of the one to the other. (*ibid.* 26)

Newton expressed his absolutism in the following much quoted passage:

> Absolute space, in its own nature, without relation to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies ...(*Principia*, 6)

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

The bucket thought experiment, surely one of the most famous thought experiments ever, is described as follows:

> ... the surface of the water will at first be flat, as before the bucket began to move; but after that, the bucket by gradually communicating its motion to the water, will make it begin to revolve, and recede little by little from the centre, and ascend up the sides of the bucket, forming itself into a concave figure (as I have experienced), and the swifter the motion becomes, the higher will the water rise, till at last, performing its revolutions in the same time with the vessel, it becomes relatively at rest in it. (*ibid.* 10)

In stage I, the surface of the water is flat and the water is at rest with respect to the bucket. In stage II, the water rotating with respect to the bucket. In stage III, the water at rest with respect to the bucket, but the surface is concave. What's the difference between I and III? Newton offers what seems like the best explanation (and possibly the only one): the water (as well as the bucket), is rotating with respect to space itself.

Of course, the bucket experiment can easily be performed as a real experiment, which presents a problem. The rest of the universe is obviously present around us, something to which a relationalist might appeal. Thus, a second thought experiment is needed and is perhaps even more effective than the bucket. Newton imagines two globes in otherwise empty space.

> It is indeed a matter of great difficulty to discover ... the true motions of particular bodies from the apparent; because the parts of that immovable space ... by no means come under the observation of our senses. Yet the thing is not altogether desperate ... For instance, if two globes, kept at a distance one from the other by means of a cord that connects them, were revolved around their common centre of gravity, we might, from the tension of the cord, discover the endeavour of the globes to recede from the axis of their motion ... And thus we might find both the quantity and the determination of this circular motion, even in an immense vacuum, where there was nothing external or sensible with which the globes could be compared. But now, if in that space some remote bodies were placed that kept always position one to another, as the fixed stars do in our regions, we could not indeed determine from the relative translation of the globes among those bodies, whether the motion did belong to the globes or to the bodies. But if we observed the cord, and found

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

> that its tension was that very tension which the motions of the globes required, we might conclude the motion to be in the globes, and the bodies to be at rest ... (*ibid.*, 12)

Leibniz had no reply to this. Position and velocity are not observable, but acceleration is. The bucket and the rotating globes seemed to establish absolutism. The first serious challenge to Newton on rotation (i.e., accelerating bodies) was from Berkeley and Mach.

Mach begins his challenge to Newton with an assertion of empiricism and a new outlook on inertia. In standard Newtonian mechanics, for instance, we explain the flattening of the earth's poles and bulging of the equator in terms of the earth's rotation. And we presume that if instead of the earth rotating, the stars rotated around the earth, then the bulging of the equator would not happen. Mach takes this to be a serious mistake and that inertial forces ought to arise equally either way. This is an expression of what has come to be known as 'Mach's Principle'. With this empiricist-inspired principle in the background, we come to Mach's counter thought experiment in his *Science of Mechanics*:

> Newton's experiment with the rotating water bucket simply informs us that the relative rotation of water with respect to the sides of the vessel produces *no* noticeable centrifugal forces, but that such forces *are* produced by its relative rotation with respect to the mass of the earth and the other celestial bodies. No one is competent to say how the experiment would turn out if the sides of the vessel increased in thickness and mass till they were ultimately several leagues thick. (Mach 1960, 284)

Mach's strategy is rather clear. He proposes a new theory: The source of inertia is not space, but rather is very large amounts of mass. He rejects Newton's bucket and two spheres thought experiments in the narrow sense, that is, he denies the phenomena that Newton claimed would be observed. The water in a rotating bucket with very think walls would not climb the wall of the bucket. And in an empty universe the two balls would not act the way Newton says, but would instead move together because of the tension in the cord connecting them. Mach does not literally assert these things, I am taking a liberty. He merely remarks, somewhat rhetorically, 'who could say what would happen?' But the point is perfectly clear: The scenarios I described are as plausible as Newton's. These are counter thought experiments, they deny the phenomena of the initial thought experiments.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

3. Dennett against Jackson

Qualia are the subjective aspects of experience, feelings of hunger, pleasure, anger, and sensations of colour, smell, and so on. They are accessible to introspection. (One quale, many qualia.) The status of qualia is central to the mind-body problem. Physicalists claim that there is nothing over and above physical facts. So, qualia present some sort of challenge. Are qualia different? Can they be reduced to the physical, or perhaps eliminated? If not, then physicalism would seem to be wrong.

Frank Jackson is a long-time champion of qualia. He produced a famous thought experiment that has been much discussed for more than two decades.

> Mary is a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room via a black and white television monitor. She specialises in the neurophysiology of vision and acquires, let us suppose, all the physical information there is to obtain about what goes on when we see ripe tomatoes, or the sky, and use terms like 'red', 'blue', and so on. She discovers, for example, just which wave-length combinations from the sky stimulate the retina, and exactly how this produces via the central nervous system the contraction of the vocal chords and expulsion of air from the lungs that results in the uttering of the sentence 'The sky is blue.' (It can hardly be denied that it is in principle possible to obtain all this physical information from black and white television, otherwise the Open University would of necessity need to use colour television.)
>
> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she learn anything or not? It seems just obvious that she will learn something about the world and our visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had all the physical information. Ergo there is more to have than that, and Physicalism is false.
>
> Clearly the same style of Knowledge argument could be deployed for taste, hearing, the bodily sensations and generally speaking for the various mental states which are said to have (as it is variously put) raw feels, phenomenal features or qualia. The conclusion in each case is that the qualia are left out of the physicalist story. And the polemical strength of the Knowledge argument is that it is so hard to deny the central claim that one

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

> can have all the physical information without having all the information there is to have. (Jackson 1982, 130)

Here is the all important knowledge argument that comes from the thought experiment.

1. Mary knows all the physical facts about colour perception.
2. She has learned these facts having only black and while experiences.
3. When she experiences colour for the first time, she learns something new.

Therefore, some facts about colour are not physical facts.

Before getting to Dennett's thought experiment, let me first take note of his outright rejection of this or indeed of any thought experiment.

> Like a good thought experiment, its point is immediately evident even to the uninitiated. In fact it is a bad thought experiment, an intuition pump that actually encourages us to misunderstand its premises. (Dennett 1991, 398).

> Thought experiments depend on folk concepts; they are inherently conservative. We should expect very counter intuitive results in real science, so violating our intuitions is to be expected (Dennett 2005, 128f)

Dennett raises an important point about intuitive concepts. But his dismissal of thought experiments because they make use of them is quite unjustified. So called folk concepts—whether they are used in thought experiments or not—can and often do lead to revolutionary results. I'll take a moment to bludgeon readers with examples.

- Galileo's thought experiment on free fall led to a new mechanics.
- Poincaré's disk thought experiment lead to very rich model of non-Euclidean geometry
- Einstein's elevator thought experiment lead to the principle of equivalence which is central to General Relativity.
- Thompson's violinist thought experiment leads to new view of the morality of abortion.

It's not just thought experiment where this happens; let me mention a few other examples.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

- Arithmetic deals with very simple (folk) concepts of addition, multiplication, and division. With these we can define 'prime number' and easily prove a very profound theorem that there are infinitely many primes.
- From the very simple concepts of arithmetic we can (step by common sense step) go on to establish the remarkable result by Gödel of the incompleteness of any set of axioms for arithmetic.
- Turing computability can readily be seen as nothing more than the elaboration of common sense concepts of rule-governed calculation, but it leads to the unexpected result that there are uncomputable functions.

I certainly don't want to say that everything is at bottom based on folk concepts. 'Isospin', 'superego', 'magnetic field', and many other important notions are certainly not commonsense ideas at all, but must be introduced in some conjectural fashion. But it's a mistake to think that starting with common sense must end in common sense. Dennett's dismissal of "intuition pumps" is quite misguided. Fortunately, Dennett condescends to play the thought experiment game, anyway, and he does so with considerable success.

Dennett's attack on Jackson is in the form of the following counter thought experiment. The setup is the same as Jackson's, but the scenario is quite different.

> And so, one day, Mary's captors decided it was time for her to see colours. As a trick, they prepared a bright blue banana to present as her first colour experience ever. Mary took one look and said 'Hey! You tried to trick me! Bananas are yellow, but this one is blue!' Her captors were dumbfounded. How did she do it? 'Simple,' she replied, 'you have to remember that I know everything—absolutely everything—that could ever be known about the physical causes and effects of color vision. So of course before you brought the banana in, I had already written down, in exquisite detail, exactly what physical impressions a yellow object or a blue object (or a green object, etc.) would make on my nervous system. So I already knew exactly what thoughts I would have (because, after all, the mere disposition to think about this or that is not one of your famous qualia, is it?). I was not in the slightest surprised by my experience of blue (what surprised me was that you would try such as second-rate trick on me). I realise that it is hard for you to imagine that I could know so much about my reactive dispositions that the way blue affected

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

> me came as no surprise. Of course it's hard for you to imagine. It's hard for anyone to imagine the consequences of someone knowing absolutely everything physical about anything! (Dennett 1991, 399f)

It should be quite clear at this point what is happening. Jackson's thought experiment has the following structure: There is a set up: Mary is in black and white room, where she learns all physical facts about perception. Next comes the phenomenon: When Mary first encounters colours, she learns something new. Finally, the result of the thought experiment, i.e., the conclusion drawn: Some facts about perception are not physical, and so, physicalism is wrong. A counter thought experiment would accept the set up, but challenge the phenomenon, which is exactly what Dennett does.

I hope the general conclusion I wish to draw from these three examples is evident: Dennett = Mach = Galileo. That is, the structure of Dennett's thought experiment is the same as Mach's and Galileo's. They are all counter thought experiments. The challenge for Dennett was not: given the thought experiment we should resist the anti-physicalist conclusion (i.e., he is not against the broad thought experiment). Rather, Dennett's challenge is that the narrow thought experiment is faulty; the phenomenon is not as Jackson claims it would be. Mary would not learn anything new. Dennett, Mach, and Galileo each deny the phenomena of the initial thought experiments.

Alternative Challenges

The Newton and Jackson thought experiments might also be challenged in the broad sense (i.e., by accepting the phenomena of the thought experiment but offering a different explanation). The challenges would not be in the form of a counter thought experiment, possibly not in the form of a thought experiment at all. I'll briefly mention two examples, just for the sake of comparison.

Contra Newton, Larry Sklar introduced his notion of *absolute acceleration* (Sklar, 1976). An object or system of objects, such as the two spheres connected by a cord, might have this property. When it does, there will, for instance, be a tension in the cord joining the two spheres. They are not rotating with respect to anything, they simply have this property of absolute acceleration. It's quite bizarre, but if one thinks of quantum mechanical spin, then one gets the idea. The spin of an electron is 'intrinsic,' it

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

cannot be transformed away in any coordinate frame. Sklar's account, unlike Mach's, does not challenge the phenomenon of Newton's thought experiment; it offers a different explanation.

Contra Jackson, David Lewis proposed the 'ability hypothesis'. (Lewis 1983, 1988) It is related to the distinction between knowing how, not knowing that. One might know absolutely every fact about a bicycle, yet not know how to ride. If one learns how to ride, one is not learning a new fact or acquiring new propositional knowledge, but rather one is acquiring a skill, a new ability. When Mary leaves the laboratory and experiences red things for the first time, she is similarly learning a skill, not learning a new fact. Lewis's account, unlike Dennett's, does not deny the phenomenon of Jackson's thought experiment, but rather undermines Jackson's knowledge argument by interpreting the phenomenon of the thought experiment in a different way.

Evaluation in these cases takes the form: Who offers the best explanation or interpretation of the phenomena in the thought experiment? Sklar and Lewis do not deny the phenomena of Newton's and Jackson's thought experiments. Rather they challenge the inference drawn after accepting the phenomena.

When Does a Counter Thought Experiment Work?

The main aim of this paper is the modest one of pointing out the existence of a distinct class of counter thought experiments. But once we accept the existence of counter thought experiments and get some idea of how they work, the inevitable questions to ask are: when do they work well?, and when do they fail? What follows is but a superficial start at addressing these questions.

Clearly, a counter thought experiment will work only when it can plausibly deny the phenomenon of the original thought experiment. I don't think anyone could reasonably hope to deny the phenomena in, say, Searle's Chinese room thought experiment. Everyone is ready to allow that a person could be in a room with Chinese characters taken in that are compared by a person inside with those in a book that tell her which Chinese characters to put out. Challenges to Searle's thought experiment have all been aimed at the inference that he drew from the phenomenon. Similarly, there is no point in rejecting the phenomenon in Lucretius infinite space thought experiment, since it merely involves throwing a spear. To do so would involve a degree of scepticism that goes well beyond the case at hand. The same could be said of Thompson's

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

violinist. Of course, we could wake up with an unconscious and very ill violinist hooked up to ourselves such that he will be cured if and only if he remains connected for nine months. There could be no plausibility to denying that such things could happen. Challenges to these thought experiments must be aimed at the various conclusions the thought experimenter draws from the directly observed part of the thought experiments.

This should be uncontentious, but there are those who would disagree. Peter Geach, for instance, takes moral rules to be devine commands and he holds that God would not allow genuine moral dilemmas to exist, since we would then have to choose between different devine commands, which he takes to be absurd. Geach imagines someone saying: ' "But suppose circumstances are such that observance of one Divine law, say the law against lying, involves breach of some other absolute Divine prohibition" ' Geach then replies:

> —If God is rational, he does not command the impossible; if God governs all events by his providence, he can see to it that circumstances in which a man is inculpably faced by a choice between forbidden acts do not occur. Of course such circumstances (with the clause 'and there is no way out' written into their description) are consistently describable; but God's providence could ensure that they do not in fact arise. Contrary to what unbelievers often say, belief in the existence of God does make a difference to what one expects to happen. (Geach 1969, 128)

The upshot, according to Geach, is that perfectly consistent thought experiments might still be illegitimate and hence the phenomenon not exist, because God would not allow it to happen. I mention this outlook in passing to further illustrate the range of possible opinion on thought experiments. It is not one I think we should seriously consider.

Here are some things that seem to matter when evaluating a counter thought experiment. They are all rather obvious.

- How reliable is the initial thought experiment in the narrow sense (i.e., would the phenomenon occur)?
- How strong is the assumed background to the thought experiment?
- How similar is the phenomenon of the thought experiment to things we know and trust?

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

- How plausible is the phenomenon of the counter thought experiment?
  - How absolutely plausible?
  - How relatively plausible?

The last of these is probably the key question, but let's flesh them all out a bit by considering our three examples.

In making the case for the Aristoteleans, we might note what happens when we throw some litter out the car window—it falls far behind as we move along. (I hope introducing a car is a harmless anachronism, and I can assure readers I am not a litterbug.) The case for Galileo could be based on our experience of tossing things around in a moving car, aeroplane, etc. Motion seems to have no effect. By analogy, if the whole earth were moving, our experience should be as in a car, plane, etc.

In making the case for Newton, we might note that we have often seen water climb rotating buckets and we have felt the tension in a string holding a rock that is spinning around us. The two globes thought experiment assumes they would act the same in empty space, which seems very plausible. Mach, on the other hand, proposes a new theory: mass is the cause of inertial motion. There is no empirical evidence for this; it's motivated by his rather strict empiricism. (Remember the fate of Mach's empiricist-inspired anti-atomism.) However, it would seem that acceleration should be on par with position and velocity—*relative*, otherwise not detectable.

Given that Mach's account is possible, he undermines to some extent the degree of belief we had in the phenomenon of Newton's thought experiment, i.e., that the cord's tension would be maintained. But Mach's counter thought experiment is certainly not as plausible as Newton's. Consequently, it is a weak attack on Newton's thought experiment, and hence, a weak attack on absolute space.

The Case for Jackson might begin by noting that in general, mental things don't seem like physical things, and more specifically, when people, for instance, acquire eye-sight late in life, they appear to learn something new. Mary the colour scientist seems like an extreme case of this; hence, Jackson's narrative appears initially plausible. But we don't know anyone who knows everything about anything, much less all the physical facts. So, the analogy with things we already know is very weak. There is a superficial similarity to Plato's cave, but with vastly greater assumptions.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

The case for Dennett might begin with noting that for various philosophical reasons, physicalism seems right (i.e., problems with dualism, etc.). And, to repeat, we have no idea what it would be like for someone to know all the physical facts. As a story, Dennett's narrative about Mary seems coherent and intelligible. It would appear then that Dennett's counter thought experiment is just as plausible as Jackson's, even though neither is very plausible in its own right. Dennett is quite aware of this:

> My variant was intended to bring out the fact that, absent any persuasive argument that this could not be the way Mary would respond, my telling of the tale had the same status as Jackson's: two little fantasies pulling in opposite directions, neither with any demonstrated authority. (Dennett 2005, 105)

Thus, Jackson is neutralized, if not refuted, and the Mary thought experiment is a failure.

Comparatively, I would say that Galileo is completely triumphant; the Aristotelian thought experiment is destroyed. Newton is slightly weakened but not seriously damaged; Mach is an alternative, but it is nowhere near as plausible. Jackson is nullified, since Dennett's alternative story is equally plausible. It's a tie, as far as the thought experiments go, which probably leaves Dennett the winner in this particular battle.

These evaluations, of course, are very rough and open to objection. They are only preliminary and should not be taken too seriously. They merely illustrate the kinds of consideration involved. My main aim is to determine how counter thought experiments work in general, not to evaluate particular instances.

I do, however, wish to explore the comparative nature of the thought experiment-counter thought experiment pair by briefly examining a simple proposal. Seeing its shortcomings will, I hope, stimulate some interest in others in the further investigation of counter thought experiments.

A Ratio Test

For quite some time, it has been common to think of the evaluation of scientific theories as taking place comparatively. Kuhn's paradigms and Lakatos's research programmes are evaluated (at least in part), by comparing them with rivals. Much the same can be said of counter thought experiments. There are, however, important differences. Comparative theory evaluation is usually

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

over the long haul and it is the whole theory/paradigm/ programme that is being compared. By contrast, thought experiments and counter thought experiments go head to head and the evaluation is direct and immediate.

In trying to capture this comparative aspect, we might try the following *Ratio Test*, as I shall call it. Assign a probability to the phenomenon of a thought experiment, given the thought experiment set up. I should readily admit and even stress that this not intended to be realistic; I doubt these things can be quantified. But it might shed a little light on the structure of counter thought experiments.

Let *initial phen* = the phenomenon in the original thought experiment (e.g., action of Newton's two spheres in empty space, or Mary learning something new when she leaves the laboratory), and let *counter phen* = the phenomenon in the counter thought experiment (e.g., action of the two spheres according to Mach, or the actions of Mary in Dennett's thought experiment). Assign probabilities to these, e.g., Prob(Mary learns something new) = r. Probability here is meant to be something like degree of belief.

It would seem that a counter thought experiment is successful, if: Prob(counter phen)/Prob(initial phen) >1, and is not successful, if: Prob(counter phen)/Prob(initial phen) << 1.

Why does the second claim not use ≤, which would be a simple denial of the first? The reason has to do with a complication I have not mentioned, but will soon be obvious.

Presenting a counter thought experiment is perhaps like the defence presenting an alternative account of the facts of a legal case. The prosecution must make its case 'beyond a reasonable doubt'. The defence need not match that high standard, but need only make a case for a slightly plausible alternative. This asymmetry in standards will upset the ratio test, or at least would greatly complicate it. Even if we think the prosecution's story is more likely, the possibility presented by the defence is enough to undermine our initial confidence. In probablilistic terms, the defence can do its job successfully even when its case has a probability well below ½, just as long as the probability isn't too low. Mach's counter thought experiment might plausibly fall in this range. It may not be plausible in its own right, but it could be plausible enough to undermine our initial assessment of Newton's thought experiment.

In general, the range of plausibility of counter thought experiments is great. Some counter thought experiments might be highly compelling in their own right, as was Galileo's. Others

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

might be weak in their own right, but still strong enough to cast doubt on the main account, as Mach's perhaps did to Newton's.

There are also cases where the ratio test might break down badly. This will happen in 'how possible' thought experiments. These are thought experiments that don't try to establish a result concerning how things are, but only try to show how something is possible. Darwin provided examples in discussion of the evolution of particular characteristics that seemed problematic. How could the eye evolve or the giraffe acquire a long neck? Darwin's thought experiment would show a possible evolutionary route. It was not intended to be true, only to show that the particular characteristic is not a counter example to the theory of evolution. As long as its probability is not equal to zero, the thought experiment is a success. One of Darwin's foremost early critics, Fleming Jenkins, produced counter thought experiments (involving 'blended inheritance'), that aimed to show the evolutionary account Darwin provided is not possible. In other words, he constructed counter thought experiments with probability virtually equal to one. (See Lennox 1991 for an account of Darwin and Jenkins.)

The ratio test is quite inappropriate in cases such as these, since almost inevitably Prob(counter phen)/Prob(initial phen) >> 1. This will happen even when the counter thought experiment is only moderately plausible, since the initial thought experiment is only meant to show possibility, not likelihood. The ratio test is at best a first stab; it is certainly not adequate as it stands. Counter thought experiments that aim to undermine 'how possible' thought experiments will have to be evaluated some other way.

Concluding Remarks

There is an interesting class of negative thought experiments, which I have called *Counter Thought Experiments*. Galileo against the Aristotelians on the motion of the earth, Mach against Newton on absolute space, and Dennett against Jackson on physicalism are instances. Evaluation of these counter thought experiments seems to be essentially comparative. A simple proposal, a ratio test, works reasonably well in some cases, but it will certainly need supplementing, since it flounders on 'how possible' cases.

Is it possible to give a general account—perhaps quite different from the one I have sketched—of what makes a counter thought experiment effective? This is wholly unexplored territory, but

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

definitely worth further investigation, as are all areas of the remarkable topic of thought experiments.

Bibliography

Brown, J.R. (1986) 'Thought Experiments Since the Scientific Revolution', *International Studies in the Philosophy of Science*, Vol I, no 1. 1986.

Brown, J.R. (1991) *Laboratory of the Mind: Thought Experiments in the Natural Sciences*, London: Routledge.

Brown, J.R. (1999) *Philosophy of Mathematics: An Introduction to the World of Proofs and Pictures*, London: Routledge.

Brown, J.R. (2003a) 'Peeking Into Plato's Heaven', *Philosophy of Science*, vol. 71, 1126–1138.

Dennett, D. (1991) *Consciousness Explained*, New York: Little Brown.

Dennett, D. (2005) *Sweet Dreams*, Cambridge, MA: MIT Press.

Galileo (*Dialogo*), *Dialogue Concerning the Two Chief World Systems* (Trans from the *Dialogo* by S. Drake), second revised edition, Berkeley: University of California Press, 1967.

Galileo (*Discoursi*), *Two New Sciences*, (Trans from the *Discoursi* by S. Drake) Madison: University of Wisconsin Press, 1974.

Geach, P. (1969) *God and the Soul*, London: Routledge and Kegan Paul.

Horowitz, T. and G. Massey (eds.) (1991) *Thought Experiments in Science and Philosophy*, Savage MD: Rowman and Littlefield.

Jackson, F. (1982) 'Epiphenomenal Qualia', *Philosophical Quarterly*, Vol. 32, No. 127, 127–136.

Kuhn, T. (1964) 'A Function for Thought Experiments', reprinted in Kuhn, *The Essential Tension*, Chicago: University of Chicago Press, 1977.

Leibniz, G. (1956) *The Leibniz-Clarke Correspondence*, H. G. Alexander (ed.) Manchester: Manchester University Press.

Lennox, James G. (1991). 'Darwinian Thought Experiments: A Function for Just-So Stories', in Horowitz and Massey (1991), 223–245.

Lewis, D. 1983. Postscript to 'Mad Pain and Martian Pain'. In his *Philosophical Papers*, Vol. 1. New York: Oxford University Press, 1983, 130–32.

Lewis, D. 1988. 'What Experience Teaches'. In *Proceedings of the Russellian Society*. Sydney: University of Sydney, 1988. Reprinted in *Mind and Cognition*, W. Lycan (ed.), Oxford: Blackwell, 1990, 499–518.

Lucretius, *De Rerum Natura*, Cambridge,Ma, Loeb Library.

Mach, E. (1960) *The Science of Mechanics*, (Trans by J. McCormack), sixth edition, LaSalle Illinois: Open Court.

Mach, E. (1976) 'On Thought Experiments', in *Knowledge and Error*, Dordrecht: Reidel.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press

Newton, I. (Principia) *Mathematical Principles of Natural Philosophy* F. Cajori (trans.), Berkeley: University of California Press
Norton, J. (1991) 'Thought Experiments in Einstein's Work', in Horowitz and Massey (1991).
Norton, J. (1996) 'Are Thought Experiments Just What You Always Thought?' *Canadian Journal of Philosophy*.
Poincaré, H. (1969) *Science and Hypothesis*, New York: Dover.
Searle, J. (1980) 'Minds, Brains, and Programs', *Behavioral and Brain Sciences* 3, 417–424.
Sklar, L. (1976) *Space, Time, and Spacetime*, Berkeley, University of California Press.
Thompson, J.J. (1971) 'A Defense of Abortion', *Philosophy and Public Affairs*, 1/1 (Fall): 47–66.

https://doi.org/10.1017/S1358246100009784 Published online by Cambridge University Press