Forthcoming in A. Booth & D. Rowbottom (Eds), *Intuitions* (Oxford University Press)

Intuitions in Science:
Thought Experiments as Argument Pumps

Darrell P. Rowbottom
Department of Philosophy
Lingnan University

DarrellRowbottom@ln.edu.hk

1. Thought Experiments and Intuitions

If intuitions play a distinctive role in the empirical sciences – if they are significant in scientific method not only in so far as scientists occasionally do logic or mathematics – then this is presumably in thought experiments.[1] Indeed a rather seductive view, *prima facie*, is that intuitions (and intuition statements) are to thought experiments as perceptions (and observation statements) are to experiments; and as such, that they play an evidential role. (I use 'intuition' to mean 'intuitive judgement'. Following Gopnik and Schwitzgebel (1998: 77), a judgement is intuitive if it is 'not made on the basis of some kind of explicit reasoning process that a person can consciously observe'.[2]) At the very least, this appears to be correct for many thought experiments, such as the one due to Simon Stevin that I will present in the next section.

Thought experiments are interesting because they appear, on the surface, to be means by which to delimit ways the world might be from the armchair. In some special cases, what's more, such delimitation results in only one possibility remaining; and one is then able to derive its actuality. This might not be terribly surprising if the kinds of possibility involved were strictly or narrowly logical – if they merely

[1] Naturally there may be other uses. Following Gopnik and Schwitzgebel (1998), for example, one might say that intuitions can be used as hypotheses too. But my own view is that the real work of science goes on in the so-called *context of justification* rather than the *context of discovery*. In short, what we do with our hypotheses when we have them is the crucial part; see Rowbottom (2011a). Nevertheless, it is worth adding that the limits of our imagination (and hence our intuitions) may bear on the realism issue. For example, if our hypotheses are appropriately constrained by experience, and the unobservable world is significantly different from the observable one, then this tells against scientific realism. See Rowbottom (2012).
[2] For a treatment of the many different uses of 'intuition', see Jenkins (This Volume).

concerned the laws of logic and the definitions of terms, e.g. recognitions that an object cannot be light and heavy simultaneously – but they sometimes appear to be metaphysical or nomic.

This paper is written with a spirit of humility. I do not think that we are in a position to determine what thought experiments in science rest on (and the nature of the intuitions therein); instead, I believe the best we can do is to explore the possibility space, and continually refine the various resultant positions in response to immanent and transcendent criticisms.[3] This is a kind of gentle progress where we get a better understanding of the options, and their relative merits and demerits; and I think it is progress worth making. Personally, I am interested in the prospects of the view that thought experiments (and the intuitions therein) rest, ultimately, on experience. In particular, I am concerned with articulating and defending the claims that: (A) the raw conceptual materials of natural scientific thought experiments are derived from experience; and (B) our 'intuitions' about how these concepts interrelate also derive from experience (where this includes learning how to use words, e.g. by ostensive definition). Call this *scientific intuition empiricism*, to be contrasted with *scientific intuition rationalism* (which involves the denial of at least one of these theses). I will try to persuade you of its viability and attractiveness, by appeal to theoretical virtues such as simplicity, but I do not think I have any evidence that it is true. (Equally, I don't think its opponents have any evidence that it is false). But naturally, those who disagree with my stance on inference may think that I succeed in doing rather more than I take myself to be capable of!

After giving an example of a thought experiment, I will look at one of the most developed – and one of the most bold and interesting – versions of scientific intuition rationalism, which is due to James Robert Brown. I will then endeavour to tease out a novel problem for this. My strategy is to be as sympathetic as possible, and to buy into many of Brown's premises that I personally reject, rather than concentrate on the more foundational issues that divide us. In short, I am aiming for an immanent

[3] This is mainly because of my stance on ampliative inferences, i.e. that they are not truth-conducive in science itself, let alone in (relatively) speculative metaphysics and epistemology. In this kind of context, in particular, inference to the best explanation is often appealed to; but I take this to be pragmatic, at best. (One problem, of course, is that we cannot be sure what needs explanation and what doesn't; we have to have a terminus somewhere.) For more on this, see Rowbottom (2011a).

approach to criticism; and one of the virtues of this, I hope, is that it will help the reader to seriously entertain Brown's position on intuitions and thought experiments.

I will subsequently devote my attention to an alternative account of thought experiments championed by John Norton, according to which they are arguments. I will argue that this is not quite correct either, with reference to different understandings of the thought experiment of Stevin, but that thought experiments are best seen as argument *pumps*. In this, I will suggest, they are similar to actual physical (or 'real') experiments.

2. Stevin's Thought Experiment

Before I launch my discussion of different accounts of thought experiments, it will help to have an example in mind; the one I choose is not only a personal favourite, but also has special significance in the account of Brown that I discuss in the next section. It concerns a contemporary of Galileo's, namely Simon Stevin, who derived the Principle of the Inclined Plane by considering a string of equally spaced balls, each of equal mass, draped over a triangular prism. (If liked, one may follow Ernst Mach in instead imagining a uniform chain.)

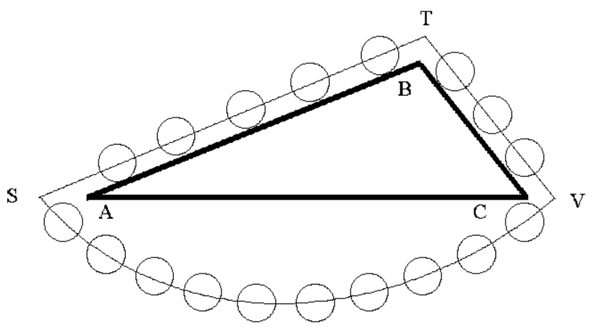

Figure 1

Stevin's Thought Experiment

Stevin's reasoning was simple but brilliant. First, the chain will remain stationary. Second, the lower portion of the chain (between S and V below the prism) is symmetrical, and therefore pulls the remainder of the chain (between S and T and T and V) equally clockwise and anti-clockwise. Now we need only imagine cutting the chain at S and V, and can see that the remainder of the chain will remain stationary.

'Hence: on inclined planes of equal heights equal weights act in the inverse proportion of the lengths of the planes' (Mach 1893: 34). We have our principle of mechanics.

Where does an intuition enter? At the very least, this is in the initial claim that the chain will remain stationary; whether there are any intuitions involved in the other steps, or whether appeal to (either folk or somewhat more developed formal) physics is all that is needed, we can leave as moot. Personally, I am sympathetic to what Mach says here: that we expect the chain not to move on the basis of our experience of similar scenarios, e.g. of hanging a coat on a peg. Others take a different view.

Before we continue, I should point out that some might object to the very way that I have presented the thought experiment. I said only 'the chain will remain stationary'; and this follows the summary of Mach (1893: 34), according to which, 'the assumption from which Stevin starts [is] that the endless chain does not move'. But Brown (2010: 3) treats matters rather differently, by starting instead with only the upper portion of the chain present on the triangular prism (i.e. between S and V *on top of* the prism). He continues:

> How will the chain move? ... There are three possibilities: It will remain at rest; it will move to the left, perhaps because there is more mass on that side; it will move to the right, perhaps because the slope is steeper on that side. Stevin's answer is the first: it will remain in static equilibrium. The second diagram below [equivalent to my figure 1] clearly indicates why. By adding the links at the bottom we make a closed loop which would rotate if the force on the left were not balanced by the force on the right. Thus, we would have made a perpetual motion machine, which is presumably impossible.

An interesting question, which I will come back to, is whether Brown and I have presented two different (but highly similar) thought experiments. (In fact, Mach's initial presentation was different again; so, as we will later see, is a reconstruction due to Norton (1996).[4]) But I wish to emphasise that Brown (2010: 4) thinks: 'The assumption of no perpetual motion machines is central to the argument, not only from a logical point of view, but perhaps psychologically as well.' My presentation did not mention perpetual motion at all; nor do I think it should have. When Brown writes of 'three possibilities', I see infinitely many possibilities. The chain could move for n seconds and stop (and maybe start again m seconds later, and so on, but not perpetually). Or have a small propensity to move and a large propensity not to move. Or explode. Or transmute into a bird and fly away. That is, I think, for all we know independently of experience.

Presumably Brown would agree that these are possibilities, of a logical variety, if pushed. Ultimately, our disagreement would be on *why* it is reasonable to disregard them, or to rule them out. We will return to this.

3. Brown's Account of Thought Experiments

Brown (1991, 2001) argues that we can grasp some laws of nature *a priori*, by intuition alone. To some, this view will appear absurd. But it is less so, at least, when it is set in proper context. First, Brown has an independent view that intuitions play a central role in mathematical inquiry, in so far as they allow us to comprehend mathematical entities *qua* abstracta; and hence, they already play a role in science, albeit indirectly, in so far as mathematical results are required in science.[5] Second, Brown thinks that laws of nature should be understood to reflect connections between universals, following Fred Dretske (1977), Michael Tooley (1977), and David Armstrong (1983). In the words of the latter:

[4] Here's what Mach (1893: 33) initially says, considering the scenario in figure one: 'The chain will either be in equilibrium or it will not. If we assume the latter to be the case, the chain, since the conditions of the event are not altered by its motion, must, when once actually in motion, continue to move forever, that is, it must present a perpetual motion, which Stevinus deems absurd. Consequently only the first case is conceivable. The chain remains in equilibrium.' I discuss this presentation in detail later.

[5] Note also that nothing I say here questions the existence of mathematical *explanations* of physical phenomena, as suggested by Mark Colyvan (2001) and Alan Baker (2005).

> Suppose it to be a law that *F*s are *G*s. *F*-ness and *G*-ness are taken to be universals. A certain relation, a relation of non-logical or contingent necessitation, holds between *F*-ness and *G*-ness. (Armstrong 1983: 85)

Brown also holds this view independently of his view on intuitions, in light of the problems with other accounts of laws, especially the regularity account championed by the likes of Alfred Ayer (1956). In particular, the notion that laws are just universal statements to which we take particular attitudes runs the risk of conflating epistemic and metaphysical issues, according to some philosophers. The line goes that there may be laws even if we cannot identify them. So even if we can only glean whether a given regularity lacks exceptions – in fact, on my own view, it is hard to see how we could even do that[6] – it does not follow that some regularities are not 'special', in so far as they are necessary in some non-epistemic sense.

I am unconvinced that we should be Platonists concerning mathematics – unconvinced, in particular, that the prospects for an empiricist line on mathematics are as bad as Brown (2010: 75–78) suggests – and also doubt that the regularity view of laws, or a somewhat more sophisticated derivative form of anti-realism concerning laws, is dead. But naturally I cannot treat these issues seriously in this chapter. So what I propose to do is just to accept these premises of Brown. If anything, this stacks the deck against empiricism concerning thought experiments.

Now Brown (2010: ch.2) also develops a taxonomy of thought experiments, as depicted in figure 2. The overarching division is between destructive and constructive types. The former refute and/or disconfirm existent theories/hypotheses. The latter result in new findings. They may articulate existing theories (in mediative cases), lead to the positing of theories (in conjectural cases), or result in (and confirm) new theories (in direct cases).[7] There is also a special class of thought experiments, Platonic ones, which are simultaneously destructive and constructive (and direct).

[6] See, for example, Rowbottom (2010).

[7] Conjectural cases are different from direct cases in so far as they do not result in any one theory in particular. Rather, the 'point of such a thought experiment is to establish some (thought-experimental) phenomenon; we then hypothesize a theory to explain that phenomenon' (Brown 1991: 40). Stevin's thought experiment, which Brown classifies as direct, does not seem to be like that. It appears instead that a lone theory 'falls out'.

Brown takes these, and these alone, to grant insight into connections between universals.

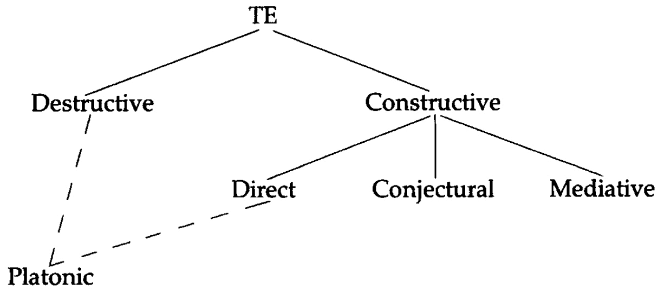

Figure 2

Brown's Taxonomy of Thought Experiments

Now according to Brown (2010: 40), Stevin's thought experiment, discussed previously, is direct but not Platonic because it does not contain a destructive component; specifically, it is not true that it 'destroys or at least presents serious problems for a theory' (Brown 2010: 33). However, it seems to me that the distinction between direct and Platonic is arbitrary *from an epistemic perspective*, in so far as it historically contingent. To be more specific, any direct thought experiment will implicitly rule out (or tell against) theories that are inconsistent with those that it tells in favour of. Imagine, for example, that Stevin had started with the general theory that 'For any non-looped chain, there will be at least one triangular prism – with angle ACB at an appropriate value, between zero and ninety – that it will fall off when draped over'. (Perhaps he did start with this theory, as a matter of fact, but abandoned it in his presentation of the thought experiment.) Then the thought experiment would have been Platonic, it seems to me, and have involved grasping contingent relations between universals on Brown's view. I therefore take it to be a pertinent example. (Note, nonetheless, that nothing I will argue in the remainder of this section requires that Stevin's thought experiment be classified one way or the other. My argument will still go through.)

Having accepted many premises that I personally reject, I will now argue that empiricism concerning thought experiments is nonetheless a reasonable position, due to a key lacuna in the Platonic account of Brown. This involves a mystery that lies at its heart, concerning how we perceive/intuit relations between universals. But before I come to explain this, there is an important caveat.

It may be tempting to think that intuitions are mysterious in a way that perceptions are not at all, but I follow Brown (2010: 81–82 & 108) in thinking that this is a mistake. Even if we take all of our physics at face value – in a way that I, as a philosopher with skeptical tendencies, am hesitant to do – it seems that our story ceases at a crucial juncture. Electromagnetic waves of a range of frequencies enter the eye, forming an image on the retina. Waves of particular frequencies cause our photoreceptor cones (and/or rods, and/or ganglions) to fire, meaning that the photoreceptive pigments in these change conformation. Ultimately, through a complex biochemical pathway, this results in a membrane potential and an electrical signal being generated. In some cases, e.g. in the case of rods, this signal is strongly amplified (so that even a lone photon can be detected).

But now what? I do not claim to be an expert in neuroscience, but I know that although the story may go on a little further, i.e. that we may be able to say which parts of the brain 'light up' when we're seeing things (rather than, say, when we have our eyes closed), it ends before it gets anywhere near explaining how we generate perceptual beliefs. To see this, just consider the distribution of photoreceptor cones, which are primarily responsible for our colour vision in daylight, in the eye. They are most densely packed in the central region (about 2% in total) of the retina, the macula, which we use when we perform tasks that require higher resolution vision (such as reading) than is needed, say, to be aware of movement. Yet although this explains some aspects of our visual experiences, e.g. that things appear sharper in the centre of our visual fields, it fails to explain others.[8] For example, why do so many people fail to notice when they are losing peripheral vision (when they *would* notice if their macula degenerated)? Saying that half of the visual cortex is devoted to processing macular signals does not answer the question. Even if we add the dubious assumption of mind-brain identity, and disregard the problem of qualia, what precisely *is* the brain doing? There are gaps between photoreceptor cones. But there are no gaps in our visual field. We don't *really* understand what is going on. And this is after granting much far more than I ever would if pushed.

[8] Even here, I am perhaps being too generous. In the words of Rossi and Roorda (2010): 'Visual resolution decreases rapidly outside of the foveal center. The anatomical and physiological basis for this reduction is unclear.'

In fact, it is plausible that the mystery of perception underlies several traditional debates in philosophy of science, in so far as (often relatively undefended) assumptions about perception serve to influence stances on the aim of science and the means by which to achieve it. For Mach (1893), for example, the aim of science is to save the phenomena just because there are no physical objects, above and beyond bundles of sense impressions, to speak of. And for Bas van Fraassen (1980), the distinction between the observable and the unobservable has special significance, in part if not in whole, because he is a direct realist.[9] So on this view, we do not *infer* the existence of observable physical things, in the way we do the existence of unobservable physical things on the account of a scientific realist.

However, I *do* think that there is a key mystery in Brown's account that lacks a correlate in our best accounts of perception, and which may be avoided by appeal to (relatively uncontroversial claims concerning) perception. Let's start, to reiterate, by accepting Brown's premises; let's imagine that there are universals, genuine intuitions involving abstracta in mathematics, and that laws of nature involve connections between universals.

But now let us consider *how* we come to be acquainted with – or to use the language of Bertrand Russell (1911), gain knowledge by acquaintance of – the universals relevant to the laws of natural science. The short answer, again following Russell (1912), appears to be that this is via *their instantiation by observable things*. So in short, we become acquainted with universals such as circularity via acquaintance with circular concrete things.[10] At the very least, there are many universals used in scientific laws that we would not have become acquainted with in the absence of specific sensory modalities. It is uncontroversial, for example, that no human can

[9] James Ladyman (2000) concurs that van Fraassen appears to be a direct realist. At the very least, it is clear from the following passage that he is not an indirect realist:

> Such events as experiences, and such entities as sense-data, when they are not already understood in the framework of observable phenomena ordinarily recognized, are theoretical entities. They are, what is worse, the theoretical entities of an armchair psychology that cannot even rightfully claim to be scientific. I wish merely to be agnostic about the existence of the unobservable aspects of the world described by science—but sense-data, I am sure, do not exist. (Van Fraassen 1980: 72)

[10] A critic might urge that we only ever see approximately circular things, after Plato. However, this seems presumptuous in so far as we may possess sense impressions, distinct from the physical things we observe, or at least *ways of perceiving circularly* (on an adverbial theory of perception). In any event, for present purposes I am content to accept that seeing things that *approximately instantiate* some universal is sufficient for grasping that universal.

grasp redness unless they have seen a red thing; so if we were all born blind, and remained blind, no-one would speak of redness at all. It is also hard to deny that experience is the sole means by which we become acquainted with the universals *that are the special concern of natural scientific theories*; those relating to charge and mass, for example. And Brown does not offer any argument to the contrary.

So far, there is no problem for Brown's view. But let us now ask how we become acquainted with the *contingent* relations between universals that Brown discusses. Let us concede, for the sake of argument, that some *necessary* relations are apparent; let's accept it is clear that all circular objects are shaped objects, for instance, due to a connection between the relevant universals (circularity/circular-ness and shaped-ness). (As the connection between these two kinds of property is manifest, it seems we *must* accept this when we've granted the existence of the universals.[11]) This presumably entails that we have the ability to intellectually grasp some *metaphysical* modalities. But how about the *contingent* relations that Brown takes to be the basis of scientific laws, i.e. that determine modalities in the *nomic* dimension? How exactly does one spot the connection between being extended and being massive, for example?

Here, I contend, Brown has nothing to say. But the empiricist who buys into the same account of laws may, by contrast, say that we come to posit (and perhaps even confirm the existence of) such connections through experience of *co-instantiation* of the relevant universals.[12] There is a long tradition in the philosophy of science of thinking in this way, which goes back to at least Francis Bacon (1620).

Moreover, the empiricist can explain why we only *sometimes* spot the contingent connections between universals; this is because we only sometimes spot that the universals in question are co-instantiated, and/or only sometimes pay attention to the relevant patterns of co-instantiation. I cannot see a rival rationalist story. In summary,

[11] I should add that I think that there is a mere conceptual necessity, not a metaphysical necessity, here. For the record, I do not believe in universals.

[12] For an anti-inductivist, which Brown (2010: 32) is emphatically not, the desirability of appealing to intuition may be greater. This is because experience could only ever lead us to falsify claims about how universals were contingently linked. It could never confirm them. See Rowbottom (2010).

then, the empiricist is able to provide an answer to two questions that the rationalist cannot *on Brown's very own preferred account of laws*.

4. Norton on Thought Experiments as Arguments

If Brown's account of thought experiments is rejected, then what are we to put in its place? The most popular alternative, which has been championed by Norton (1991, 1996, 2002, 2004) in a series of papers, in which he covers an impressive array of historical examples, is that thought experiments are simply *arguments* which:

> *(i) posit hypothetical or counterfactual states of affairs,*
> *and (ii) invoke particulars irrelevant to the generality of the conclusion.* (Norton 1991: 129)

What *kinds* of arguments are they? 'A very broad range of argument forms should be allowed… [including] inductive argument forms.' (ibid.) And it follows that some thought experiments are considerably better than others: 'A good thought experiment is a good argument; a bad thought experiment is a bad argument.' (Norton 1991: 131) It would therefore seem that deductive cases, where the 'particular-free conclusion follows deductively from the premisses' (ibid.) are exemplars of good thought experiments; and gladly, Norton (1996: 350) thinks Stevin's is just such a case.

So on Norton's view, we can decide between empiricism and rationalism concerning scientific thought experiments by examining the grounds for belief in (or sources of) their premises.[13] (Naturally it might turn out that neither epistemological view is correct across the board; the premises of different thought experiments may be grounded in a variety, even a wide variety, of fashions.) In the case of Stevin's thought experiment, at least as I have presented it, experience appears to be enough. It is a matter of experience that chains don't move in actual circumstances similar to those hypothetical circumstances outlined by Stevin.

[13] Indeed, Norton (1991: 142) states: 'I do not rule out the possibility that thought experiments can allow us to gain access to universals as Brown argues. However I urge that they may do so only insofar as these universals can be accessed via argumentation.'

But are all thought experiments really arguments? Michael Bishop (1999) argues convincingly that they are not, by asking us to consider a well-known exchange between Niels Bohr and Albert Einstein. Without getting bogged down in unnecessary detail, the story is as follows. Einstein presented a thought experiment that he took to refute a view of Bohr. And although Bohr could not respond satisfactorily at the time, he was able to do so the very next day *by discussing the same hypothetical/counterfactual circumstances*. What's more, Einstein agreed with Bohr's response; he accepted, that is to say, that he had previously been mistaken.

Now Bishop's concern is that this would mean that Einstein *changed his mind* about what was a good argument on Norton's view. In the words of Bishop (1999: 539–540):

> [On Norton's view,] two tokens of a thought experiment are tokens of the same thought experiment-type if and only if they are tokens of the same t-argument [thought-argument] type…
>
> [But] the assumption that the arguments Bohr and Einstein proposed were type-identical makes a muddle of the episode. Consider what happened. At first, Einstein proposed a t-argument that Bohr did not accept… Later, Bohr proposed a t-argument that forced Einstein to disown his original t-argument. Here is what one must say if one believes that Einstein and Bohr were presenting tokens of the same t-argument-type: Bohr did not accept Einstein's t-argument, so he presented Einstein with a token of that same argument; and then when faced with a token of his own t-argument, Einstein proceeded to disavow that very argument.

But might Norton not respond by suggesting that Bohr presented a different (type of) thought experiment in response to Einstein? Bishop (1999: 540) suggests that this is wrong because 'Einstein could have rightly accused Bohr of changing the subject', and he did not. So the natural thing to say, Bishop thinks, is that Einstein and Bohr were discussing the same (type of) thought experiment, but developed *different* arguments on the basis of it.

However, Norton (2002) has a convincing response. He points out that the thought experiments (and thus arguments) might be construed as different yet highly similar. In particular, he suggests that there is just one key difference: Einstein's thought experiment occurs in classical spacetime, whereas Bohr's occurs in relativistic spacetime. As such, Einstein could not reasonably have accused Bohr of changing the subject. Rather, it became clear that his initial thought experiment (implicitly) presumed that his very own view of spacetime was incorrect! The relevance of Bohr's thought experiment, which introduced relativistic considerations but was otherwise identical, was evident.

5. Thought Experiments as 'Argument Pumps'

I agree with the verdict of Bishop (1999) that thought experiments are not arguments, but wish to argue for it with a different example. Specifically, I wish to return to Stevin's thought experiment, and consider how different philosophers present it. In particular, I will be interested in the different way that they present the argument for the principle of the inclined plane.

Recall my initial presentation of Stevin's thought experiment; I took it to be a fundamental premiss that the joined string, depicted in figure one, does not move.[14] And I probably took this from Mach (1893: 34), who writes of 'the assumption from which Stevinus starts, that the endless chain does not move' at one point. However, this is not a premiss of the argument presented by Mach (1893: 33) on the page beforehand:

> The chain will either be in equilibrium or it will not. If we assume the latter to be the case, the chain, since the conditions of the event are not altered by its motion, must, when once actually in motion, continue to move forever, that is, it must present a perpetual motion, which Stevinus deems absurd. Consequently only the first case is conceivable. The chain remains in

[14] This is how I remembered Mach's presentation of Stevin's argument when I began work on this chapter. I also have evidence for this claim, because I have written on Stevin's thought experiment previously in a handbook entry (although this has not yet appeared). See Rowbottom (Forthcoming).

> equilibrium. The symmetrical portion… may, therefore, without disturbing the equilibrium be removed.

Assuming that the translation is faithful, this treatment is imprecise. Clearly the second case *is* conceivable; I can picture, in my mind's eye, such a chain moving unceasingly.[15] Such movement is *judged* impossible because of assumptions such as 'the conditions of the event are not altered by its motion' and 'perpetual motion… [is] absurd'. But there is nothing in the scenario described by Stevin – as depicted in figure one – to suggest that such assumptions are true. These are imported. One might think that 'The chain remains in equilibrium' for many different reasons; indeed, one might be convinced that the chain would not move, as I was, without even considering the notion of perpetual motion. (I believe my conviction rests on experience of other similar scenarios; that is, on old experience.)

Now Brown (2010: 3), as we saw in the initial presentation of Stevin's thought experiment, instead considers an open chain lying over a triangular prism, and then considers what would happen if links were added so as to create the situation depicted in figure one: '[by] adding the links at the bottom we make a closed loop which would rotate if the force on the left were not balanced by the force on the right'. This proceeds in the opposite direction to Mach, who *begins* with the scenario in figure one and asks us to work out what will happen when the string is *cut* at particular points (or more accurately, when the symmetrical portion is somehow 'removed'). I take this to show that it is only central to the thought experiment to consider the difference between two possible setups; the means by which the transition is made is irrelevant. (And in fact, there is no need for a transition at all. It suffices to consider two different chains, draped over two different prisms of the same type.) However, it is fascinating that such transitions *do play a part* in the arguments offered by Mach and Brown. And this suggests to me that thought experiments are not arguments.

I suppose one might instead conclude that Mach and Brown present slightly different thought experiments. This seems rather odd, however, when both take themselves to

[15] One might raise the difference between *prima facie* and ideal conceivability, introduced by Chalmers (2002), in an attempt to defend Mach's view. However, it seems to me that the second case is ideally conceivable. See also Sidelle (2002, §1).

be reporting on one and the same experiment. In short, it seems to me more natural to say that they have constructed different arguments on the basis of the same thought experiment than it does to think that they have constructed different thought experiments *as well as* different arguments. I contend that both understood the key point that Stevin sought to make, and the importance that understanding what happens in one kind of setup (with a string/chain forming a closed loop) has for understanding what occurs in another (involving an open string/chain).[16] Bishop (1999: 540) offers an additional line of argument in support of this conclusion, by driving a wedge between the notion of repeating an experiment and that of repeating an argument:

> Thought experiments, like real experiments, can be repeated. And in order to repeat an experiment, it is not necessary (or even possible) to duplicate the original in all its details. In fact, if one thinks that an experiment has been botched, it would be folly to try to duplicate it, mistake and all. (Bishop 1999: 540)

Even putting the aforementioned difference of presentation to one side, moreover, there are many other differences in the way Stevin's thought experiment, and/or the argument to be derived from it, are presented. Mach asks us to consider two possibilities; equilibrium and non-equilibrium. Brown asks us to consider three; the chain stays still, the chain moves clockwise, and the chain moves anticlockwise. Mach appears *not* to think the mention of perpetual motion is crucial to Stevin's argument.[17] But Brown (2010: 4) thinks that 'the assumption of no perpetual motion machines is central'. And Norton (1996: 350) seems to agree, when he presents the argument in the following 'capsule form':

> 1. Assumption: The experimental arrangement is as shown in… Figure 1.

---

[16] Alternatively, an advocate of the view that thought experiments are arguments might suggest that I have reconstructed Stevin's argument in a superior, or simply different, way. Indeed, it seems that Mach (1893), from whom many philosophers will first have learned of Stevin's thought experiment, does a rather fine job of improving on Stevin's lengthy and complicated initial presentation. At the very least, the presentations are strikingly different. See Kühne (2001: 320–322).

[17] At least, recall, this seems true if Mach (1893: 34) genuinely thought that 'the assumption from which Stevin starts [is] that the endless chain does not move'. As I earlier pointed out, however, Mach *does* mention perpetual motion in his initial discussion of the thought experiment, just one page before. So perhaps the correct view is that his treatment is not internally consistent, and/or that he had two distinct arguments in mind.

2. Assumption: Perpetual motion is impossible.
3. From 1 and 2: The chain is in static equilibrium.

Let's stick with Norton, for a moment, and also look at his treatment of the scenario depicted in figure one in slightly greater detail:

> If the distribution of weight of the chain were to lead it to rotate about the prism, then it would do so endlessly, since the distribution of the homogeneous chain about the prism would be unaltered by such rotation. (Norton 1996: 349)

Again, there are some puzzles here. Why should we assume that the distribution of the homogenous chain about the prism would be *causally* unaltered by such rotation? And must the density of the material forming the chain remain uniform under rotation? (Imagine that the motion warmed the balls on the string to a temperature where they started to melt.) My point in raising these possibilities is just to illustrate how much is being imported in order to make these arguments seem plausible. Certainly it is hard to see *explicitly* how the 'argument is deductive', as Norton (1996: 351) claims, unless we are expected to assume a good deal of physics.[18] But if we

---

[18] It may help to compare the more detailed reconstruction of Laymon (1990: 170):

> Tx = x is a situation of the sort described by Stevin, namely, one where there's a prism with hypotenuse parallel to the ground, with a rope wrapped around it that has equally spaced balls attached and that is friction free, &c.
>
> Ex = x is in equilibrium.
>
> Px = x is a situation where there is perpetual motion (of Aristotelian type)…
>
> [O]ur three premises are:
>
> (1) $\exists x(Tx \ \&\ x = a)$
> (2) $Ea \vee {\sim}Ea$
> (3) ${\sim}\exists x(Px)$
>
> Stevin's argument that the rope is in equilibrium can be represented as:
>
> (4) $\exists x(Tx \ \&\ x = a) \ \&\ {\sim}\exists x(Px) \ \&\ (Ea \vee {\sim}Ea) \rightarrow Ea$ (claimed logical fact)
> (5) Ea (by 1, 2, 3, 4 and some logical cousin of *modus ponens*)

In short, my worries concern premiss four, which I take Norton to be arguing for in the previous quotation.

assume this physics, the worry is that the principle of the inclined plane is also assumed (although not explicitly), and is not believed *directly* on previous experience.

So why do I think that all these different arguments derive from the same thought experiment? The answer is that they all concern highly similar '*hypothetical or counterfactual states of affairs*'. To be more specific, they all involve the positing of the same hypothetical objects (e.g. chains) undergoing similar kinds of manipulations (e.g. being cut or extended and joined up). Put simply, my view is that a thought experiment is just an experiment that involves '*hypothetical or counterfactual states of affairs*', and which serves as an argument pump. Not any old argument will appropriately relate to those states of affairs, in any given context; but many arguments will.

Does the 'argument pump' view have any special virtues? First, following something akin to Roy Sorensen's (1992: 214) principle of parity, it is satisfying that we may say that actual experiments are 'argument pumps' too.[19] We often disagree about what an actual experiment shows in so far as we construct different arguments on the basis of it, *as opposed to questioning its results*. For example, it is commonplace to 'explain away' apparent conflicts between theory and experimental results by appeal to factors that have not been considered in generating predictions. Thus any good secondary school pupil can explain why experiments in mechanics don't give quite 'the right answer' by appeal to friction, and so forth.

Second, it becomes explicit that the kind of theory-ladenness involved in observations in standard experiments also occurs in thought experiments. *In each kind of experiment, theoretical context affects which arguments are generated and/or thought to be sustainable on the basis of the states of affairs*. The only difference is whether the states of affairs are actual rather than hypothetical and/or counterfactual.

Third, the 'argument pump view' is attractive because it is explicitly compatible with the operational perspective on scientific thought experiments championed by Marco

[19] Sorensen (1992: 214) proposes 'a parity thesis: thought experiments are arguments if and only if experiments are arguments'. And my proposal is compatible with 'thought experiments are argument pumps if and only if experiments are argument pumps'.

Buzzoni (2008), where 'all [scientific] thought experiments may conceivably become real experiments and all real experiments may be conceived as realized [scientific] thought experiments'. I take this to be an independently appealing viewpoint, at least in so far as 'experiment' takes the same referent. However, I personally think that thought experiments can conceivably 'become real experiments' not in the sense that they might become actual, but only in so far as they could be performed in different possible worlds. (I do not believe in actual frictionless planes, rather than approximations to these, for instance. And part of the distinct value of thought experiments is that they can involve situations that are physically impossible, in addition to those difficult or impossible to realize in practice.) In any event, it is interesting to see what Buzzoni (2008: 69) says about Stevin's experiment:

> Stevin's thought experiment is not demonstratively powerful because it can be reconstructed as an argument; rather, it can be reconstructed as an argument because, as soon as we see the apparatus built by Stevin and follow through the few steps of his experiment, we are persuaded of its validity. If we reconstructed Stevin's thought experiment as an incorrect argument, we would question our reconstruction rather than the experiment.

So it is possible for the holder of the 'argument pump' view to agree with the gist of the idea that 'we analyse and appraise thought experiments by reconstructing them explicitly as arguments and testing them against just those standards which we apply to arguments of other forms' (Norton 1991: 142). What she will deny is that there is any *reconstruction of the thought experiment*, rather than *reconstruction of the line of reasoning of some author based on the thought experiment*. And in fact, the holder of the 'argument pump' view can venture an explanation of *why* it is necessary for so much reconstruction to occur. The problem is not typically in grasping the hypothetical and/or counterfactual states of affairs. Rather, it consists in grasping what someone else thinking about those states of affairs takes them to show. As already implied, in the previous mention of theory-ladenness, this kind of thing happens with standard experiments too. For some realists, the Casimir effect – that two parallel uncharged plates, in a vacuum, attract one another when in close proximity – shows the existence of virtual photons. For me, and indeed some scientists, it does not; and that's because we take virtual photons to be mere aids to calculation, mere instruments, rather than legitimate physical posits.[20]

The 'argument pump' view is also as amenable to scientific intuition empiricism as Norton's alternative. Specifically, its advocate may hold that thought experiments are: 'not some kind of mysterious new window onto the physical world' (Norton 1991: 142). Rather, the premises that play a part in the (good) arguments *that it is possible to pump from* thought experiments may rest on experience, rather than an additional faculty.

Might Norton rejoin that thought experiments are not (precisely) arguments, but instead classes of highly similar arguments? My response is twofold. First, this would constitute a genuine shift in his position, towards mine. (Indeed, I do not think this is a shift that Norton would want to make. But it is worth considering nonetheless.) Second, this view seems undesirable for pragmatic reasons, at the bare minimum. (And this is true even if one thinks that the matter reduces, when such a shift is made, to a verbal dispute about 'experiment'.) It requires, unnecessarily, that we use 'experiment' differently in one context than in another; for physical (or 'real') experiments can be repeated, we are apt to say, just by bringing about relevantly similar actual states of affairs. If we can say the same about thought experiments, without any loss in explanatory power, so much the better. Among other things, doing so encourages us to look to physical experiments to understand the conditions for hypothetical and/or counterfactual states of affairs to be part of an experiment. For example, just as to accidentally knock a glass over is not to perform a physical experiment, so writing fiction is not necessarily to perform a thought experiment. Experiments involve states of affairs intentionally brought about in order to solve problems.[21]

---

[20] For more on this, see Rowbottom (2012).

[21] Moreover, rather interestingly, it appears that physical experiments can be fit, roughly at least, into the classification scheme for thought experiments proposed by Brown, which I discussed in section 3. Certainly, physical experiments can be either destructive or constructive on most views of scientific method; they can refute and/or confirm/corroborate theories. Furthermore, they can illustrate phenomena that lead us to conjecture theories (i.e. be 'conjectural'). They can also 'articulate' theories (i.e. be 'mediative') on the view of Kuhn (1996, ch. 3); see Rowbottom (2011b, 2011c) for some examples. Whether they can be 'direct' may be somewhat more controversial. But think of it this way. They can prompt 'direct' deductive arguments for one theory in particular, if appropriate premises are assumed.

Acknowledgements

I am especially grateful to Jim Brown and John Norton for detailed feedback on an initial version of this chapter. I have also benefitted from comments from Peter Baumann, Tony Booth, Duncan Pritchard, and Oxford University Press's anonymous referees. Finally, I should also like to thank audiences at Dubrovnik (39th Annual Philosophy of Science Conference), Lingnan University, and Durban (39th Annual Conference of the Philosophical Society of Southern Africa, 2013). I cannot hope to name everyone who has said something useful, but I recall having benefitted, in particular, from comments made by Jamin Asay, Marco Buzzoni, Rafael de Clerq, David Papineau, Dean Peters, and Neven Sesardic. My work on this essay was supported by the Sydney Centre for the Foundations of Science (where I held a Visiting Fellowship), the British Academy (by way of a Postdoctoral Fellowship), and Lingnan University (by way of travel grants).

Bibliography

Armstrong, David. 1983. *What Is A Law of Nature?* Cambridge: Cambridge University Press.

Ayer, Alfred J. 1956. 'What Is A Law of Nature?', *Revue Internationale de Philosophie* 10, 144–165.

Bacon, Francis. 1620. *Novum Organum*. Translated by P. Urbach and J. Gibson. La Salle: Open Court, 1994.

Baker, Alan. 2005. 'Are There Genuine Mathematical Explanations of Physical Phenomena?', *Mind* 114, 223–238.

Bishop, Michael A. 1999. 'Why Thought Experiments Are Not Arguments', *Philosophy of Science* 66, 534–541.

Brown, James R. 1991. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge, 1st Edition.

Brown, James R. 2010. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge, 2nd Edition.

Buzzoni, Marco. 2008. *Thought Experiment in the Natural Sciences: An Operational and Reflexive-Transcendental Conception*. Würzburg: Königshausen and Neumann.

Chalmers, David J. 2002. 'Does Conceivability Entail Possibility?', in T. S. Gendler & J. Hawthorne (eds), *Conceivability and Possibility*, 145–200. Oxford: Clarendon Press.

Colyvan, Mark. 2001. *The Indispensability of Mathematics*. Oxford: Oxford University Press.

Dretske, Fred. 1977. 'Laws of Nature', *Philosophy of Science* 44, 248–268.

Gopnik, Alison & Schwitzgebel, Eric. 1988. 'Whose Concepts Are They, Anyway? The Role of Philosophical Intuition in Empirical Psychology', in M. R. DePaul & W. Ramsey (eds), *Rethinking Intuition*, 75–91. Lanham: Rowman & Littlefield.

Jenkins, Carrie. This Volume. '"Intuition", Intuition, Concepts and the A Priori'.

Kuhn, Thomas S. 1996. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kühne, Ulrich. 2001. 'Die Methode des Gedankenexperiments.' Unpublished PhD Thesis, University of Bremen. (Abridged version published as: *Die Methode des Gedankenexperiments*, Frankfurt am Main: Suhrkamp Verlag, 2005.)

Ladyman, James. 2000. 'What's Really Wrong with Constructive Empiricism? Van Fraassen and the Metaphysics of Modality', *British Journal for the Philosophy of Science* 51, 837–856.

Laymon, Ronald. 1991. ‘Thought Experiments by Stevin, Mach and Goy: Thought Experiments as Ideal Limits and as Semantic Domains’, in T. Horowitz and G. Massey (eds), *Thought Experiments in Science and Philosophy*, 167–191. Lanham: Rowman & Littlefield.

Mach, Ernst. 1893. *The Science of Mechanics: A Critical and Historical Account of Its Development*. La Salle: Open Court, 1960, 6th Ed.

Norton, John D. 1991. ‘Thought Experiments in Einstein’s Work’, in T. Horowitz and G. Massey (eds), *Thought Experiments in Science and Philosophy*, 129–148. Lanham: Rowman & Littlefield.

Norton, John D. 1996. ‘Are Thought Experiments Just What You Thought?’, *Canadian Journal of Philosophy* 26, 333–366.

Norton, John D. 2002. ‘Why Thought Experiments Do Not Transcend Empiricism’, in C. Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*, 44–66. Oxford: Blackwell.

Norton, John D. 2004. ‘On Thought Experiments: Is There More to the Argument?’, *Philosophy of Science* 71, 1139–1151.

Rossi, Ethan A. & Roorda, Austin. 2010. ‘The Relationship between Visual Resolution and Cone Spacing in the Human Fovea’, *Nature Neuroscience* 13, 156–157.

Rowbottom, Darrell P. 2010. 'Evolutionary Epistemology and the Aim of Science', *Australasian Journal of Philosophy* 88, 209–225.

Rowbottom, Darrell P. 2011a. *Popper's Critical Rationalism: A Philosophical Investigation*. London: Routledge.

Rowbottom, Darrell P. 2011b. 'Stances and Paradigms: A Reflection', *Synthese* 178, 111–119.

Rowbottom, Darrell P. 2011c. 'Kuhn Vs. Popper on Criticism and Dogmatism in Science: A Resolution at the Group Level', *Studies in History and Philosophy of Science* 42, 117–124.

Rowbottom, Darrell P. 2012. 'The Instrumentalist's New Clothes', *Philosophy of Science*, 78, 1200–1211.

Rowbottom, Darrell P. Forthcoming. 'Educational Research as Science?', in A. Reid, M. Peters, C. Russell and P. Hart (eds), *Companion to Research in Education*. Dordrecht: Springer.

Russell, Bertrand. 1911. 'Knowledge by Acquaintance and Knowledge by Description', *Proceedings of the Aristotelian Society* 11, 108–128.

Russell, Bertrand. 1912. *The Problems of Philosophy*. London: Oxford University Press.

Sidelle, Alan. 2002. 'On the Metaphysical Contingency of Laws of Nature', in T. S. Gendler & J. Hawthorne (eds), *Conceivability and Possibility*, 309–336. Oxford: Clarendon Press.

Sorensen, Roy A. 1992. *Thought Experiments*. Oxford: Oxford University Press.

Tooley, Michael. 1977. 'The Nature of Laws', *Canadian Journal of Philosophy* 7, 667–698.

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Clarendon Press.