Armchair physics and the method of cases

Abstract

The method of cases, i.e., the informal elicitation of judgements in thought experiments for the purpose of philosophical theorizing, has been much criticized in recent years. In this paper we point out that the use of the method of cases is not peculiar to philosophy: it can also be found in physics, where it has made a valuable contribution to the probing and formulation of theories. The method of cases *per se* should therefore not be treated as intrinsically methodologically flawed. This paper has also a narrower goal: to show that when the analogy underlying the so-called 'expertise defense' is drawn between judgments in thought experiments in philosophy and in physics, several objections can be avoided.

1 Introduction

Over the last two decades, experimental philosophers have gathered an impressive array of data on judgments in philosophical thought experiments, also known simply as *cases*. One of the several strands that has been discerned within the experimentalist movement is the so-called "negative program" (Knobe and Nichols 2017). Proponents of the negative program have argued that the traditional method of informally eliciting *case judgments*[1] from philosophers is deeply flawed (Alexander and Weinberg 2007, Machery 2017): (i) it doesn't take into account the diversity of case judgments of subjects of different demographics and subjects without philosophical training, and (ii) it is ignorant of the sensitivity of case judgments to presentation effects (order and framing of cases).

The conclusions that some proponents of the negative program have drawn from their results come in varying degrees of skepticism. At the moderate end, experimental philosophers caution against the uncritical use of case judgments and recommend the scientific investigation of when and where case judgments can be used as evidence (Alexander and Weinberg 2007, 2014, Mortensen and Nagel 2016). On the more radical end, philosophers have called for "a radical restriction of the deployment of [case judgements]

[1] Oftentimes judgements in cases have been described as "intuitions". But friends and foes of the method of cases have criticized the characterization of these judgments in terms of their psychological properties (Williamson 2007, Cappelen 2012, Machery 2017). Machery (2017) aptly refers to judgments in cases simply as "case judgments".

as evidence" in philosophy (Alexander and Weinberg 2007, 61), and have recommended that we "suspend judgement[s] in response to most philosophical cases", in particular case judgments that are used to support "modally immodest philosophical views" (Machery 2017, 7; see also Machery 2011).

This paper has a narrower and a broader goal. The broader goal of the paper is to show that philosophers are not alone in using judgments in thought experiments for the purposes of theorizing. Practitioners in one of our most cherished sciences, namely physics, have done so too quite successfully. The method of cases, we argue, therefore shouldn't be viewed as somehow intrinsically flawed, methodological oddity of philosophy. The narrower goal of the paper is to argue that a popular response to these experimental findings, namely the so-called *expertise defense* may be better off than sometimes claimed.

The expertise defense has emerged as one of the most popular ways of reacting to the experimental findings regarding case judgements by the 'folk'. Its proponents have argued that the apparent error proneness of case judgements by the folk does not give us a good reason to question the reliability of case judgments by philosophers, just as the error proneness of the folk in matters scientific wouldn't give us a good reason to question the judgments by scientists (Hales 2006, Ludwig 2007, Horvath 2010, Devitt 2011, Williamson 2011). Unfortunately, both the adherents and the critics of the expertise defense have remained rather vague on what kinds of judgments in science the expertise defense builds on. This has resulted in criticism that is threatening to undermine the expertise defense (Ryberg 2012, Nado 2014a, Rini 2014, Nado 2015, Machery 2017).[2] In this paper, we want to propose that it is best build on *judgments in thought experiments in science*.

This paper is structured as follows. In Section 2 we show that thought experiments in physics are very similarly structured as thought experiments in philosophy and that judgments in both contexts serve an evidential function. We'll also argue that the epistemology of these judgments is very similar. In Section 3 we'll argue for our broader goal, namely that thought experiments are not intrinsically problematic. In Section 4 we'll argue that the expertise defense can be strengthened by drawing the underlying analogy between judgments in thought experiments. In Section 5 we conclude by pointing to further open issues which the proponents of the expertise defense will still have to be addressed.

[2] Other issues have been raised by Weinberg et al. (2010), Machery (2011, 2017).

2 Armchair physics

Debates about the method of cases have focused on the judgments made in philosophical thought experiments. Although there is a rich literature on thought experiments in science, there is little discussion about judgments figuring in thought experiments. Instead, philosophers of science have been mostly interested in the nature and function of thought experiments in science. The most prominent views range from constructivist accounts according to which thought experiments are capable of pointing us to weaknesses of current theories (Kuhn 1977, Gendler 1998, Camilleri 2014), to accounts which construe thought experimentation as a form of mental modelling (Nersessian 1992), to Platonist views which conceive of some thought experiments as windows to laws of nature (Brown 1991), and to the view that thought experiments are identical to simply arguments (Norton 2004). Although Norton's view has been criticized by many commentators as too restrictive (Bishop 1999, Bokulich and Frappier 2017, Brendel 2017), a weaker thesis, namely that thought experiments can be reconstructed as arguments, strikes us as not too problematic.[3] Häggqvist (2009), who has analyzed one important class of thought experiments in philosophy and in science in terms of counterfactual conditionals (see also Williamson 2007), proposes the following reconstruction:

1. *A counterfactual scenario*: It is possible that A is the case.
2. *The theoretical conditional:* If theory T is true, and if it were the case that A, a consequent, C, should be true.
3. *A counterfactual conditional*: If it were the case that A, it would be the case that not-C.
4. *Conclusion (by modus tollens)*: Therefore, T is false.

That is, thought experiments, which seek to undermine a theory T, can be reconstructed as consisting of a consideration of a counterfactual scenario, where A is the case and where it is judged that C is not the case, contradicting a theory that entails C. The evidence for the falsehood of T is provided by the counterfactual judgement that C. This scheme works for thought experiments from science and philosophy alike. Consider for example Schrödinger's cat:

S1. It is possible that a cat can be situated in a way described by the thought experiment.

[3] In fact Norton tries to establish the identity thesis largely by challenging his critics to find a thought experiment that cannot be reconstructed as an argument.

S2. If the Copenhagen interpretation is true, then, if a cat were situated this way, there is no fact of the matter as to whether the cat is dead or alive.
S3. If the cat were situated in this way, it would have to be either alive or dead.
S4. Therefore, the Copenhagen interpretation is false.

Now consider the notorious Gettier cases:

G1. It is possible for somebody (S) to be in a Gettier situation with regard to p.
G2. If knowledge is justified true belief, then if S were in a Gettier situation with regard to p, S would know that p.
G3. If S were in a Gettier situation with regard to p, S would justifiably believe the true proposition p without knowing that p.
G4. Therefore, knowledge is not justified true belief.

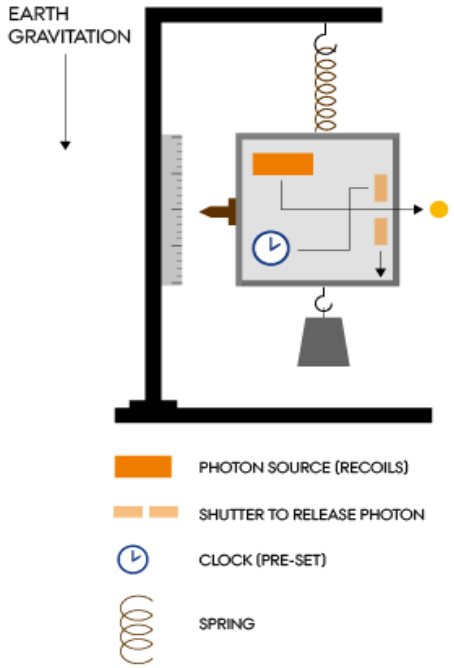

Figure 1: The clock-in-the-box thought experiment. Proposed by Einstein to challenge the Heisenberg uncertainty relation, it was supposed to show that it was possible to measure the energy of a photon at a particular time.

Again, in *both* Schrödinger's cat and the Gettier cases, the relevant counterfactual judgment figures *as evidence* in the attempted refutation of the theory in question (the Copenhagen interpretation of quantum mechanics and the theory that knowledge is justified true belief, respectively). Although not all thought experiments are meant to refute a theory (there are also some meant to support theories)[4], many important ones do fall into that category.

[4] A thought experiment in physics that supports a theory is e.g. Einstein's elevator. A thought experiment from philosophy may e.g. be Putnam's twin earth supporting the causal theory of reference for natural kinds. See also Praëm and Steglich-Petersen (2015) for a discussion of constructive and destructive thought experiments in both philosophy and in science.

For some reason, many well-known thought experiments in science are from physics.[5] So let's consider yet another thought experiment from physics, namely the so-called clock-in-the-box thought experiment, which Einstein used to question the Heisenberg uncertainty relation (see Figure *1*). The following reconstruction again follows Häggqvist (2009):

E1. It is possible (in principle) that a single photon exit a box equipped with an arbitrarily exact timer and an arbitrarily sensitive spring-balance.

E2. If the uncertainty principle holds, then the time and energy of its passage would not be simultaneously measurable to any degree of accuracy violating the inequality $\Delta E \times \Delta t > h$ (where h is Planck's constant / 2π).

E3. If a single photon exited a box equipped with an arbitrarily exact timer and the box were then weighed, the time and energy of its passage would be simultaneously measurable to any degree of accuracy.

E4. Hence, the uncertainty principle doesn't hold.

Once again, it is the counterfactual conditional judgment which is meant to provide the evidence on the basis of which a theory (here: an interpretation of quantum mechanics) is questioned. Whether or not the evidence is good, is of course another matter. In this case, it wasn't (see Bishop 1999 for more details).[6] Regardless, the intended function *was* evidential.

More precisely and more correctly, it isn't so much the judgments but rather the *content* of the judgment that provides evidence for theorizing (Williamson 2007, Machery 2017). To see this, consider for example the argument in which one judges that <"Gödel" refers to Gödel and not Schmidt> and where this judgment is the premise for the conclusion that <"Gödel" refers to Gödel and not Schmidt>. In such an argument, the premise (i.e., the judgment) is not available before the conclusion (which the judgment is about). Hence, as long as the conclusion is not asserted or accepted, the premise cannot be true. Accordingly, the judgment cannot serve as evidence for the relevant conclusion; only its content can. For simplicity's sake, however, we shall continue to speak mostly of the evidential function of judgements (rather than of the content of the judgments).

[5] See Stuart et al. (2017) for (less standard) thought experiments from the other sciences.

[6] It turns out that Einstein ironically ignored relativistic effect that would arise from the movement of the box in a gravitational field, affecting the measurement of the energy of the photon.

2.1 Are judgements in thought experiments *actually* evidential?

Judgments in thought experiments in physics can have a very similar evidential function as they do in philosophical thought experiments, as we've just seen. But are they also effectively treated as such?

Judgements elicited in *philosophical* thought experiments have been instrumental in the history of the discipline of philosophy in rejecting previously widely held philosophical views. For example, the relevant judgements in the Gettier cases have been instrumental in undermining the JTB theory of knowledge, Kripke's Gödel cases have helped undermine descriptivism, Mary's room have convinced many that physicalism about the mind is false, etc. Although it has of course been contested by experimental philosophers that judgements elicited by thought experiments *ought to* be used as evidence, it is denied by few that they have *in fact* played an evidential role in philosophical practice. That is, it is denied by few that case judgements have historically constituted important reasons for rejecting certain philosophical accounts and supporting others (see e.g. Machery (2011, 2017)).[7]

So again, in what sense—if any—have judgements in thought experiments in physics played an evidential role in the practice of physics? Clearly, physicists have better sources of evidence at their disposal than judgments in thought experiments, on which they can rely when assessing theories. But this doesn't mean that judgements in thought experiments *never* serve as evidence in physics. On the contrary, especially in times of theoretical reform judgments in thought experiments can play an important role in highlighting the problems of previous theories and help explore new ones (Kuhn 1977). For example, in the mid-16th century Galileo famously showed by means of a thought experiment of two connected falling objects of different weights that there was an inherent contradiction in Aristotelian physics (Gendler 1998). In the late 1920s, Schrödinger used his cat thought experiment to argue against the plausibility of the Copenhagen interpretation of the new quantum mechanics, and in particular against an ontological interpretation of quantum uncertainty. As he concluded after the presentation of his thought experiment matter-of-factly: "*That*

[7] Cappelen (2012) has argued that intuitions have not played any evidential role in philosophical theorizing. This is controversial. But as far as we are concerned, Cappelen may be right and case judgments still played the evidential role that most philosophers accept they did.

prevents us from so naively accepting as valid a 'blurred model' for representing reality" (Schrödinger 1980, 157).

In a famous debate between Einstein and Bohr concerning the foundations of the new quantum mechanics in the 1920s, Einstein used the aforementioned "clock in the box" thought experiment to challenge Bohr on the plausibility of Heisenberg's uncertainty relation. Bohr's reaction to this thought experiment was *not* to reject it out of hand because it wasn't a real experiment. On the contrary, Bohr was deeply disturbed by it. As one of his students later reported:

> It was quite a shock for Bohr ... he did not see the solution at once. During the whole evening he was extremely unhappy, going from one to the other and trying to persuade them that it couldn't be true, that it would be the end of physics if Einstein were right; but he couldn't produce any refutation. I shall never forget the vision of the two antagonists leaving the club: Einstein a tall majestic figure, walking quietly, with a somewhat ironical smile, and Bohr trotting near him, very excited. ... The next morning came Bohr's triumph. (Rosenfeld, cited in Pais 1982, 446-7)[8]

That is, Bohr accepted that Einstein's thought experiment elicited the judgement also in him and that the uncertainty relation would be violated in such a scenario. His reported reaction illustrates how seriously he took this thought experiment as *evidence* against the uncertainty principle (luckily for him, he found a way to avoid this threat (Bishop 1999)).

Of course, this is only anecdotal evidence and a more systematic investigation would be required to assess the actual historical impact thought experiments have had in the relevant contemporary debates.[9] Yet it seems undeniable that thought experiments were used at crucial junctures of theory formulation and in laying theoretical foundations at times of scientific progress (again see Kuhn 1977). Galileo's thought experiments, for example, were part and parcel of his arguments for a new science of motion (McAllister 2004). Thought experiments such as Schrödinger's cat and Wigner's friend have played a pivotal role in the critical thinking about interpretations of quantum mechanics even long after they were originally advanced.

[8] See also Bishop (1999) for a philosophical discussion.

[9] Such a systematic historical investigation, as far as we are aware, is not yet available.

2.2 The epistemology of judgements in thought experiments

In many thought experiments, the truth of the counterfactual judgements arguably cannot be determined on an experimental basis – again, the lack of the appropriate experimental resources is often a reason for resorting to thought experiments in the first place! Moreover, many thought experiments in physics cannot even possibly be carried out in our world. For example, both the already mentioned Einstein's clock-in-the-box and Schrödinger's cat thought experiment describe imaginary scenarios which are practically impossible: it's impossible to build an apparatus that would measure the weight of a photon (photons don't have any mass) and it's impossible to determine the state of the cat in the box without employing some sort of (measurement) interaction. Likewise, in Einstein's famous elevator thought experiment, it seems impossible (by our means) to pull an elevator with a person inside with the required speed through space that has no nearby gravitational fields. But if there is no experimental evidence to determine the right answer of some thought experiments in science, then how do physicists evaluate the truth content of judgements about thought experiments in physics? The short answer is: in a similar way as philosophers evaluate the truth content of case judgments, it seems.

As mentioned above, standard accounts reconstruct thought experiments as counterfactual arguments and the ability to 'conduct' thought experiments as the ability to evaluate counterfactual conditionals (Norton 1991, Williamson 2007, Häggqvist 2009). Roughly, when one engages in counterfactual reasoning, one 'enriches' the antecedent via mental simulation and with the help of one's background knowledge. If this enrichment results in the truth of the consequent of the counterfactual, the whole counterfactual conditional is acceptable (Williamson 2007). This seems to be a plausible model not only thought experiments in philosophy, but also for thought experiments in science. In thought experiments in science one also uses one's imagination and one's background knowledge in order to produce the relevant scenarios before "one's eyes". Likewise, in thought experiments in science one also needs to judge whether the enrichment of the antecedent of the relevant counterfactual leads to the truth of the consequent. How does one evaluate the truth of a counterfactual?

In accordance with the standard, Lewis-Stalnaker semantics of counterfactuals, a counterfactual is true exactly when the consequent is true in a possible world that is closest

to the actual world in which the antecedent is true.[10] There is one crucial difference between counterfactuals associated with thought experiments in philosophy and science respectively. Thought experiments in philosophy often not only diverge from the actual facts but also violate the laws of nature. Metaphysically possible worlds are therefore often more relevant than physically possible worlds in the assessment of the relevant counterfactual. In contrast, in thought experiments in science the laws of nature set strong constraints on what is considered possible—even if many thought experiments in physics are not actually physically possible, given the physical facts of this world (our limited ability to produce energy, for example). Regardless, the metric and the procedure for evaluating the closeness of any of such worlds in a semantics of counterfactuals should be the same.

One may be tempted to think that, perhaps, in philosophical thought experiments one can evaluate the truth of the relevant counterfactual without considering any kind of empirical knowledge, whereas in physics the only relevant facts are the empirical ones. But neither of these things is quite true. Even though we're often interested in metaphysically (or logically) possible worlds in philosophical thought experiments, the evaluation of the relevant counterfactual is often a posteriori; we have to consult our background knowledge about the *actual* world in order to be able to judge whether the worlds considered are sufficiently similar (Williamson 2007). On the other hand, in thought experiments in physics, one often has to abstract away from the actual matters of facts and focus merely on whether or not a situation is physically possible by virtue of the laws of nature. So neither thought experiments in philosophy nor thought experiments in physics are either "fully" a priori or fully a posteriori.

Let us take stock. We have seen that judgments in many thought experiments in philosophy and physics serve the same function (as evidence in arguments against some theory) and that judgments in both kinds of thought experiments have a similar epistemology. In the following two sections we will build on these results to argue for the broader and narrower goal of this paper: there is nothing intrinsically suspect about case judgments, and the expertise defense can be strengthened by drawing the underlying analogy between judgments in thought experiments in philosophy and physics.

[10] The account provides a "similarity metric", i.e., a set of criteria on the basis of which one can evaluate how close a possible world is to the actual world.

3 The broader goal: no intrinsic flaws in the method of cases

As mentioned in the introduction of this paper, the negative program in experimental philosophy has drawn rather pessimistic conclusions about the epistemic prospects of case judgements. The critics are quite adamant that their skepticism about case judgements does not extend to other forms of judgments (Machery 2011, 2017). But what is it that makes specifically case judgments, rather than some other forms of judgments, somehow intrinsically problematic as a source of information for philosophical theorizing? Some proponents of the negative program have taken it upon themselves to identify reasons for why this may be so.

Machery (2017) identifies three "disturbing characteristics" of philosophical cases, which he suggests are responsible for the "fundamental unreliability" of case judgments (111-120): cases are (i) unusual, (ii) pull apart properties that usually go together, and (iii) contain "superficial content". Cases are unusual: body- and brain-swapping cases ask subjects to asses situations they may have never read or thought about before. Cases pull apart properties: in the notorious trolley cases, we are asked to consider cases in which causing harm can actually do more good than bad. In Gettier cases, the "usual strategies" of identifying knowledge may fail because some crucial properties of the belief forming process that normally hold (e.g. safety) do not hold in these cases. Cases contain superficial content, that is, narrative content that is irrelevant to the philosophical point in question and that may distract subjects from what's at stake. For example, in Gettier cases, it doesn't matter whether the scenario is about somebody in the office driving a brown Ford or about a broken clock showing the right time. In sum, Machery argues that the epistemic uncertainty associated with these three "disturbing characteristics" is likely to result in individual differences in subjects' responses and to make subjects' responses sensitive to confounders.

Machery singles out specifically *philosophical* thought experiments as environments that cause the unreliability of judgments. Yet *physical* thought experiments exhibit *all* the very characteristics that Machery describes as "disturbing". For example, Schrödinger's cat is a highly 'unusual' scenario in Machery's sense (we infrequently encounter it and rarely read texts about it). That is, it is highly unusual for cats to be trapped in a box with mechanisms that have a certain probability of killing them. Properties that usually go

together are also being pulled apart. That is, quantum mechanical properties are usually associated with micro-objects like electrons and photons, not so much with cats. Finally, Schrödinger's cat also contains plenty of narrative detail irrelevant to the point under consideration: it is irrelevant whether the animal in the box is actually a cat, a dog, or another animal entirely, it is irrelevant what exactly the mechanism is that would kill the animal (e.g. poison or an explosion), etc.

We seem to be faced with the following option: either Machery is right that thought experiments in philosophy exhibit disturbing characteristics that cause the unreliability of specifically case judgments or he isn't. If he's right, then we seem forced to conclude that thought experiments in physics, by virtue of them exhibiting the very same characteristics that we find in philosophical cases, too should be environments which cause judgments to be unreliable. But this seems too strong. Although there are clearly some thought experiments in physics that may be considered misleading or even "failed" (Norton 2004, 2017), there are also very successful ones – some of which we discussed – which elicit judgments that have been important reasons for rejecting or for clarifying the targeted physical theories. There is now even some empirical evidence that judgments made in thought experiments in physics can be considered reliable – at least when made by experts (Schindler and Saint-Germier forthcoming). A *general* skepticism about judgments in thought experiments seems therefore unwarranted; there is little reason to think that there is anything intrinsically problematic about thought experiments, be it thought experiments in physics or thought experiments in philosophy.

All this is of course not to say that there could never be any biases or confounders in the judgments made in thought experiments in philosophy or physics; there may well be. But for all that we know, we shouldn't be *more* worried about judgments made in thought experiments than about other kinds of judgments—pace the armchair sceptics.

Finally, we should note that, if the characteristics of thought experiments identified by Machery are not as problematic as he makes them out to be, then this undermines not only his singling out of case judgments as unreliable, but it also sheds doubt on what's at the heart of the experimentalists' challenge to armchair-philosophy: the inductive leap from the data on the unreliability of case judgments by the folk to the (inferred) unreliability of case judgments by philosophers. That is, if thought experiments do not constitute *especially* problematic environments for reliable judgments, then there are no substantive reasons for

why the critics' skeptical conclusions about folk judgements should generalize to *all* case judgments; there may simply be some groups that make more reliable case judgments than other groups. But instead of arguing this point in detail here, we'll now turn to our related, narrower goal.

4 The narrower goal: strengthening the expertise defense

A central way in which the expertise defense is motivated is by analogy to other fields of expertise such as mathematics, physics, linguistics and palaeontology (Hales 2006, Ludwig 2007, Horvath 2010, Devitt 2011, Williamson 2011). It would strike us as outright absurd to think that, for example, the expertise to accurately perform mathematical operations or the ability to construct proper physical experiments should in any way be undermined by experiments showing that the folk diverge significantly in their responses when asked about a mathematical problem, the setting up of a physical experiment, and so on. Why, the proponents of the expertise defense ask, should we then be disturbed if experiments showed that case judgements of laypeople were unreliable?

While most proponents of the expertise defense draw the analogy to natural science or math, they have appealed to a (just mentioned) motley bunch of science expert judgments. For the analogy to be convincing, however, the similarities between expert judgments must be of the *right kind*. For example, drawing an analogy between apples and oranges might be quite mistaken when drawn between the contents of these fruits, but (despite the proverb) it might also be quite acceptable, for example, when drawn between their shapes. Is the analogy underlying the expertise defense sound? This has been denied by Nado (2014a, 2015), Ryberg (2012), Rini (2014), and Machery (2017). Here we will show that the expertise defense can avoid some of these criticisms if the analogy is drawn not just between case judgments in philosophy and (generic) judgments in science, but rather, between case judgments and judgments in *thought experiments in science.*

4.1 How to avoid objections to the analogy

Nado (2014a, 2015) offers perhaps the most comprehensive criticism of the analogy underlying the expertise defense.[11] In particular, Nado has argued that expert judgments in

[11] Ryberg (2012) and Rini (2014) focus on the expertise defense in the moral realm. See also Andow (2015) for a critique of the Ryberg.

science – such as a physicist's judgment that a moving object will follow a certain trajectory – (i) are reflective (rather than intuitive), (ii) are based on explicitly accessible and learned principles (rather than on opaque ones), (iii), do not function as evidence for theories. Nado takes these three points to show that not only that the analogy fails, but also that it's implausible that philosophers would be better than the folk at controlling confounders of their case judgments (as argued by the expertise defense). She also adds that, in contrast to philosophers, scientists have developed methods for controlling their own biases. Until philosophers have not developed "similar compensatory procedures", she claims, the expertise defense doesn'thave much plausibility (Nado 2014, 10).

In what follows we will show that Nado's three points of criticism do not apply when the analogy is drawn between case judgments and judgments in thought experiments in physics (rather than some generic and vague category of 'scientific judgments'). Let us consider Nado's three points in more detail in turn.

First, Nado argues that, in contrast to case judgements (or intuitions) in philosophy, which are "immediate and unreflective" (1032) and made "in the absence of introspectively obvious conscious reasoning" (1029), judgements in science are *reflective* judgements (Nado 2015, 1032). However, Nado herself points out that this difference is not accepted by people like Devitt who hold that philosophical (and linguistic) judgements are theory-laden and that can be improved by better theories (Devitt 2010, 2011).[12] Likewise, philosophers who do not think that case judgments should be defined in terms of their psychological properties (e.g., Williamson (2007), Cappelen (2012), and Machery (2017)) would probably not accept that a contrast between "intuitive" and "reflective" captures an adequate difference between case judgments and judgments in science.[13]

Even if one were to accept Nado's distinction between intuitive and reflective judgments, it looks as though there are judgments in thought experiments in physics, which satisfy Nado's first class of judgments—the class, that is, which she reserved for philosophical case judgments. It is for example intuitive that Schrodinger's cat is alive or

[12] See Colaço et al. (manuscript) for a detailed critical discussion and an empirical test of this claim.

[13] Williamson (2007) warns against "psychologizing the data" in the method of cases (5) and proposes to get rid of the term 'intuition' altogether (220). Cappelen (2012) argues that intuitions play no substantive role in philosophical practice. Machery (2017) rejects psychological characterizations and prefers to speak of case judgments (as we do). See also fn. 1.

dead even before we open the box. It is also intuitive that the Stevin's chain will not move (Figure 2) and that the objects on Galileo's ship will follow the same trajectories as when dropped on a stationary ship (see also below).

Second, Nado argues that scientific judgments are "mediated by explicit beliefs and theories acquired in the classroom", whereas case judgments are not (Nado 2015, 1032). Scientists accordingly have "explicit access" to principles justifying their judgments and those principles are "consciously available and straightforwardly articulable" (7-8). In contrast, in philosophy, these principles are "frustratingly opaque" (8). For example, when they judge the shape of the trajectory of a cannon ball, physicists can justify their judgements in a way that philosophers cannot, namely by appealing to the well-established principles of Newtonian mechanics. No such principles are available for case judgments, Nado claims, because if they were, "we likely could have avoided, for instance, several decades of post-Gettier literature" (8).

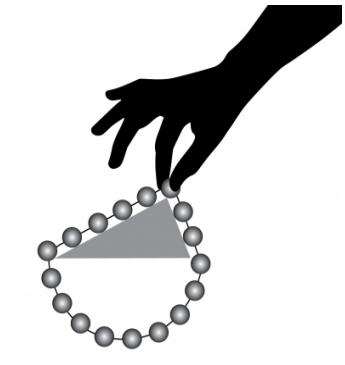

Figure 2: Stevin's chain. Does a chain of evenly divided balls draped over two inclined planes start moving to the right or left under its own weight? No, Stevin argued, because if it would, and each ball the chain had moved to the position of the previous ball, the chain would be indistinguishable from its initial position and the movement shall continue indefinitely, in contradiction with the impossibility of perceptual motion machines.

Nado's assessment applies to (generic) judgments in science, such as predicting the path of a projectile. But consider now judgments in thought experiments in science. Are the principles that underlie judgements in thought experiments in science always transparent and accessible to those making the judgments? It seems not. For example, in order to demonstrate that a moving earth was reconcilable with the manifest physics, Galileo in Day Two of his *Dialogue Concerning the Two Chief World Systems* invited his readers to "shut yourself up with some friend in the main cabin below decks on some large [moving] ship" and to observe the motion and trajectories of various objects, such as a fish in a water tank, water drops from a bottle falling on the floor, a ball thrown to a 'friend', etc. Since the physics on a uniformly moving ship (under ideal conditions) is indistinguishable from one that is standing still, there is every reason to think that a moving earth shouldn't imply a

different physics than the one that manifests itself to us. The principle justifying the judgment (that the physics on a moving ship is no different from a standing ship, or more generally that the laws of motion are the same in all inertial frames) became to be known as *Galilean invariance*.

When Galileo proposed his thought experiment, this principle was only informally and intuitively available to Galileo; he had no means for expressing it formally as part of a theory of motion. Likewise, the intended judgment in so-called Stevin's chain (1586), presupposes the principle that perpetual motion machines are impossible (see Figure 2). The theory of thermodynamics, which rules out perpetual motion machines, however, would not be formulated until the early 19th century. In yet other thought experiments in science, it's actually quite obscure what the principles might be that underlie the relevant judgments. For example, what is the principle that underlies the judgment that Schrödinger's cat is either dead or alive, even before the box is opened? Some kind of principle of familiarity with everyday experience?

In sum, the principles underlying judgments in thought experiments in physics are not as transparent as Nado makes out in her comparison of (generic) judgments in science and case judgments. Conversely, one may also question whether the principles underlying case judgments are really always as "frustratingly opaque" as Nado claims. In Gettier cases, which Nado singles out specifically, we have all developed a sense that the reason we make the judgement that "Smith doesn't know" has got something to do with the widely accepted fact that *one cannot know by luck*. Overall then, expert judgments in philosophy and in science can be quite similar with regard to their (in-)transparency.

Apart from the question of transparency of principles underlying judgments, there is another issue implicit in Nado's second critique. Nado implies that in science, the learned and explicitly accessible principles help scientists to avoid bias and mistakes (whereas no such principles are available in philosophy). But that is only the case, if these principles are (approximately) true and they are known to be true. Only then can one disqualify a judgment (say, about the shape of a projectile) as being erroneous, when it violates such principles, or when it is based on the wrong principles. In the context of thought experiments, however, this assumption is highly problematic. As we've mentioned already in Section 2.1, thought experiments have been proposed in historical periods in which the

theoretical foundations of the science in question were very much in doubt and new ideas yet still to be proven (Kuhn 1977). For example, the early 20th century witnessed a flurry of thought experiments when classical physics made way for quantum mechanics and the theory of relativity. Many of the most well-known thought experiments in physics in fact stem from this period: Schrödinger's cat, the clock-in-the-box, the Einstein-Podolsky-Rosen paradox, Einstein's elevator, Einstein's twins, etc. Other famous thought experiments stem from what we know as the scientific revolution: Galileo's tower and ship thought experiments, and Newton's bucket and cannon (for details see e.g. Brown and Fehige 2019 and Stuart et al. 2017). In these revolutionary periods, thought experiments were used to convince sceptics that the new science could make sense of previous obstacles such as "why do objects, when shot vertically up in the air, not land to the west from where they departed, if the earth really would rotate eastward?" (a great concern of Galileo), or how circular planetary motion could be accommodated with terrestrial physics (as neatly illustrated by Newton's cannon). It is in these revolutionary periods of science scientists could not appeal to the truth of the principles underlying their judgments, since this still had to be established.

Third, Nado identifies a "crucial methodological difference" between scientific judgements and philosophical judgments in cases: whereas armchair philosophers often treat case judgements as (defeasible) evidence for philosophical theories, judgements in science are not used to support scientific theories. Instead, theories are used to support judgments. As Nado says, "the support goes the other way round" (1033).

We've already seen in detail in Section 2 that judgments in thought experiments in science also serve an evidential function. That is, the judgment that the cat in Schrödinger's thought experiment is either dead or alive is meant to undermine the Copenhagen interpretation of quantum mechanics. The judgment that, in some scenarios, the energy of a photon can be measured at a particular time was supposed to undermine the Heisenberg uncertainty relation. So in thought experiments in physics, the evidential support *does* often run in the right direction (from judgment to theory, not the other way around).

Of course, physicists are much less dependent on judgments in thought experiments than philosophers are: physicists can often conduct actual experiments to probe their arguments. In fact, there are important thought experiments (such as the Einstein-Podolsky-

Rosen paradox), (parts of) which were successfully translated into an actual experimental set up (Aspect 1999). There are however other thought experiments, where actual experiments are not (and probably never will be) available (see Section 2.2 for examples).

Regardless of whether thought experiments in physics will eventually be realized or whether other (better) evidence sources become available, it is important to note once more that in their original historical contexts, these additional evidence sources were *not* available. It is in these contexts in which judgments in thought experiments fulfil their (intended) evidential function.

To the three points just discussed, Nado adds that judgments in science are different from judgments in philosophy in that scientists have successfully developed by methods for controlling biases affecting their judgments, such as double-blinding and randomization. Philosophers are yet to come up with "similar compensatory procedures" in philosophical practice (Nado 2015, 1035).[14] Although this is correct as far as many scientific judgments are concerned, things look different when it comes to thought experiments in science. Quite in contrast to real experiments, there is no obvious sense in which thought experiments could be controlled, for example, by double-blinding or randomization.

In sum, the objections that Nado and others have raised against the analogy underlying the expertise defense can be avoided if the analogy is drawn between case judgments and judgments in thought experiments in science (in particular in physics). Combined with a recently published result that physicists are indeed better at judging thought experiments in physics than the folk, the expertise defense seems considerably strengthened.

4.2 On the relevance and prospects of the expertise defense

The vast majority of the experimentalist challenge to the method of cases comes from studies with students or Amazon Turk subjects (see Machery 2017 for a comprehensive overview). Here the expertise defense applies rather straightforwardly. However there have also been studies with professional philosophers (Schulz et al. 2011, Machery 2012, Schwitzgebel and

[14] Similarly, Machery (2017) criticizes the analogy underlying the expertise defense by pointing out that whereas scientists "have extensively calibrated their instruments … there is nothing analogous with respect to the cases used by philosophers" (161).

Cushman 2012, Tobia et al. 2013, Vaesen et al. 2013, Horvath and Wiegmann 2016). Here the expertise defense seems to be in significant trouble.

It is worth noting that most of the extant studies with philosophers concern cases from the moral realm. For example, Schwitzgebel and Cushman (2012) and Tobia et al. (2013) present data showing that even the moral intuitions of philosophers are susceptible to order effects and framing effects, respectively. For all that we know, however, case judgements may be generated by different mental capacities and may therefore have different degrees of reliability (Nado 2014b). Rini (2014) in fact singles out specifically moral philosophy as a domain in which the expertise defense may be applicable only with certain caveats. So even though the expertise defense may have lost plausibility at this point with regards to moral judgments, it still seems to be a live option with regards to other kinds of philosophical judgments.

To date, there are still only a handful of studies with philosophers on non-moral judgments (Machery 2012, Vaesen et al. 2013, Horvath and Wiegmann 2016). In Machery (2012) study on reference, the majority of experts from linguistics and philosophy report the standard Kripkean judgement. The proportion varies across groups (66.7% to 88.6%), with philosophers and semanticists being more likely to report the standard Kripkean judgement than linguists with other specialisations. This is compatible with the expertise defense (see also Devitt 2012). Horvath and Wiegmann (2016), in their study on (non-)knowledge cases, found that philosophers in the majority of cases agreed with the 'textbook consensus' and the folk didn't (in agreement with the expertise defense). However, they found that even though experts' responses in fake-barn scenarios were on average closer to the disagree end of the scale than the folk with regards to knowledge ascriptions (as expected), more experts agreed than disagreed with the textbook consensus (contrary to expectations). Vaesen et al.'s (2013) study is problematic in that it doesn't seem to be about case judgements in the first place, but rather about judging mere stipulations. So, for example, instead of asking subjects about some version of Gettier cases, subjects were asked to judge whether "Steffi knows" qualifies as knowledge in a sentence like "Steffi knows that water is H_2O". There was no scenario in which Steffi knew that water is H_2O merely by luck.

Thus, although the expertise defense does seem to be in trouble with regards to moral case judgments, the case still has to be made that other kinds of case judgments are similarly problematic when made by trained philosophers.

5 Conclusion

This paper pursued a broader and a narrower goal. The broader goal was to show that the method of cases should not be treated as an intrinsically problematic methodological abnormality. This we argued by pointing out that thought experiments have been used also in the history of physics to successfully challenge, clarify, and to support theories. Philosophers therefore shouldn't be too apologetic when it comes to the use of the method of cases. Our narrower goal was to show that the analogy underlying the expertise defense can be strengthened by drawing it between case judgements and judgments in thought experiments in physics. We don't pretend to have delivered a blanket victory to the defenders of traditional philosophical methodology; there are still many open issues, such as whether it actually makes sense to talk about *expertise* in making judgments in philosophical cases (Machery 2017). We do believe, though, that we have clarified the basis on which the debate of the expertise defense should take place.

References

Alexander, Joshua and Jonathan M Weinberg. 2007. Analytic epistemology and experimental philosophy. *Philosophy Compass*, **2** (1): 56-80.

———. 2014. The "unreliability" of epistemic intuitions. In *Current Controversies in Experimental Philosophy*, Edouard Machery and Elizabeth O'Neill (eds.), New York: Routledge, 128-145.

Andow, James. 2015. Expecting moral philosophers to be reliable. *Dialectica*, **69** (2): 205-220.

Aspect, Alain. 1999. Bell's inequality test: more ideal than ever. *Nature*, **398** (6724): 189-190. https://doi.org/10.1038/18296.

Bishop, Michael A. 1999. Why thought experiments are not arguments. *Philosophy of Science*: 534-541.

Bokulich, Alisa and Mélanie Frappier. 2017. On the Identity of Thought Experiments: Thought Experiments Rethought. In *The Routledge companion to thought experiments*, Michael T Stuart, Yiftach Fehige and James Robert Brown (eds.), London: Routledge, 545-557.

Brendel, Elke. 2017. The Argument View. Are thought experiments mere picturesque arguments? . In *The Routledge companion to thought experiments*, Michael T Stuart, Yiftach Fehige and James Robert Brown (eds.), London: Routledge, 281-291.

Brown, James Robert. 1991. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*. New York: Routledge.

Camilleri, Kristian. 2014. Toward a constructivist epistemology of thought experiments in science. *Synthese*, **191** (8): 1697-1716.
Cappelen, Herman. 2012. *Philosophy without intuitions*. Oxford: Oxford University Press.
Colaço, David, Markus Kneer, Joshua Alexander, and Edouard Machery. manuscript. On second thought: A refutation of the reflection defense.
Devitt, Michael. 2010. Linguistic intuitions revisited. *The British Journal for the Philosophy of Science*, **61** (4): 833-865.
———. 2011. Experimental semantics. *Philosophy and Phenomenological Research*, **82** (2): 418-435.
———. 2012. Whither experimental semantics? *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia*, **27** (1): 5-36.
Gendler, Tamar Szabó. 1998. Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science*, **49** (3): 397-424.
Häggqvist, Sören. 2009. A model for thought experiments. *Canadian Journal of Philosophy*, **39** (1): pp. 55-76.
Hales, Steven D. 2006. *Relativism and the Foundations of Philosophy*. Cambridge, MA: MIT Press.
Horvath, Joachim. 2010. How (not) to react to experimental philosophy. *Philosophical Psychology*, **23** (4): 447-480.
Horvath, Joachim and Alex Wiegmann. 2016. Intuitive expertise and intuitions about knowledge. *Philosophical Studies*, **173** (10): 2701-2726.
Knobe, Joshua and Shaun Nichols. 2017. Experimental Philosophy. *The Stanford Encyclopedia of Philosophy (Winter 2017 Edition)*, edited by Edward N. Zalta, https://plato.stanford.edu/archives/win2017/entries/experimental-philosophy/.
Kuhn, Thomas S. 1977. A function for thought experiments. In *The Essential Tension*, Thomas S. Kuhn (ed.), Chicago: University of Chicago Press, 240–265.
Ludwig, Kirk. 2007. The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy*, **31** (1): 128-159.
Machery, Edouard. 2011. Thought experiments and philosophical knowledge. *Metaphilosophy*, **42** (3): 191-214.
———. 2012. Expertise and intuitions about reference. *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia*, **27** (1): 37-54.
———. 2017. *Philosophy within its proper bounds*: Oxford University Press.
McAllister, James W. 2004. Thought experiments and the belief in phenomena. *Philosophy of Science*, **71** (5): 1164-1175.
Mortensen, Kaija and Jennifer Nagel. 2016. Armchair-Friendly Experimental Philosophy. *A Companion to Experimental Philosophy*: 53-70.
Nado, Jennifer. 2014a. Philosophical Expertise. *Philosophy Compass*, **9** (9): 631-641.
———. 2014b. Why intuition? *Philosophy and Phenomenological Research*, **89** (1): 15-41.
———. 2015. Philosophical expertise and scientific expertise. *Philosophical Psychology*, **28** (7): 1026-1044.

Nersessian, Nancy. 1992. In the theoretician's laboratory: thought experimenting as mental modeling. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, **2** (291-301): 39.

Norton, John D. 1991. Thought experiments in Einstein's work. In *Thought Experiments in Science and Philosophy*, T. Horowitz and Gerald J. Massey (eds.), Savage, MD: Rowman and Littlefield, 129-148.

———. 2004. Why Thought Experiments Do Not Transcend Empiricism. In *Contemporary Debates in the Philosophy of Science*, Christopher Hitchcock (ed.), Oxford: Blackwell, 44-66.

———. 2017. The worst thought experiment. In *The Routledge companion to thought experiments*, Michael T Stuart, Yiftach Fehige and James Robert Brown (eds.), London: Routledge, 454-468.

Pais, Abraham. 1982. *Subtle is the Lord: The Science and the Life of Albert Einstein: The Science and the Life of Albert Einstein*. New York: Oxford University Press.

Praëm, Sara Kier and Asbjørn Steglich-Petersen. 2015. Philosophical thought experiments as heuristics for theory discovery. *Synthese*, **192** (9): 2827-2842.

Rini, Regina A. 2014. Analogies, moral intuitions, and the expertise defence. *Review of Philosophy and Psychology*, **5** (2): 169-181.

Ryberg, Jesper. 2012. Moral intuitions and the expertise defence. *Analysis*, **73** (1): 3-9.

Schindler, Samuel and Pierre Saint-Germier. forthcoming. Are thought experiments "disturbing"? The case of armchair physics. *Philosophical Studies*. https://doi.org/10.1007/s11098-019-01333-w.

Schrödinger, Erwin. 1980. The Present Situation in Quantum Mechanics. A Translation of Schrodinger's "Cat Paradox" Paper. (Trans. John D. Trimmer) *Proceedings of the American Philosophical Society*, **124**: 323-338.

Schulz, Eric, Edward T Cokely, and Adam Feltz. 2011. Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition*, **20** (4): 1722-1731.

Schwitzgebel, Eric and Fiery Cushman. 2012. Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, **27** (2): 135-153.

Stuart, Michael T, Yiftach Fehige, and James Robert Brown, eds. 2017. *The Routledge companion to thought experiments*. London: Routledge.

Tobia, Kevin, Wesley Buckwalter, and Stephen Stich. 2013. Moral intuitions: Are philosophers experts? *Philosophical Psychology*, **26** (5): 629-638.

Vaesen, Krist, Martin Peterson, and Bart Van Bezooijen. 2013. The reliability of armchair intuitions. *Metaphilosophy*, **44** (5): 559-578.

Weinberg, Jonathan M, Chad Gonnerman, Cameron Buckner, and Joshua Alexander. 2010. Are philosophers expert intuiters? *Philosophical Psychology*, **23** (3): 331-355.

Williamson, Timothy. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

———. 2011. Philosophical expertise and the burden of proof. *Metaphilosophy*, **42** (3): 215-229.