Thought Experiments Repositioned

Arnon Levy

Draft of Jan. 27, 2024

Abstract

Thought experiments play a role in science and in some central parts of contemporary philosophy. They used to play a larger role in philosophy of science, but have been largely abandoned as part of the field's "practice turn". This chapter discusses possible roles for thought experimentation within a practice-oriented philosophy of science. Some of these roles are uncontroversial, such as exemplification and aiding discovery. A more controversial role is the reliance on thought experiments to justify philosophical claims. It is proposed that if we adopt an underlying empiricist view of concepts, then thought experiments can be seen as affording us contact with scientific practice, despite their seemingly *a priori* character. The advantages and drawbacks of thought experiments are discussed via comparison with case studies, on the one, and simulations on the other hand. The chapter closes with some remarks on how to combine thought experiments with other methods.

1. Introduction

One of the most marked trends in recent philosophy of science is its increasing contact with scientific practice. At least relative to its 20th Century positivist ancestors[1], current philosophy of science is far more attentive to actual, concrete science – it treats it both as a subject matter and a source of examples and evidential support **(add cross-refs here?**) This 'practice turn' has many manifestations, one of which is a shift in the methods and tools of which philosophers of science avail themselves. Specifically, the practice turn has seen a pronounced retreat from—indeed, in some context a near elimination of—the use of thought experiments. What was once a common tool in philosophy of science (and remains so in other parts of philosophy) has come to be regarded as largely unhelpful, possibly even biasing and distracting. In this chapter, I aim to take another look at thought experiments. While acknowledging their limitations, I will also highlight ways in which they can have value, indeed special value. This is partly because they have a role in discovery and exemplification. And partly because, at least if some additional conditions are

[1] Though perhaps not relative to some of its deeper origins, e.g. the foundational writings of William Whewell.

met, they may be able to provide a different form of contact with scientific practice, allowing them to complement other methods.

The chapter begins with an attempt to characterize thought experiments, or at any rate by highlighting some of their distinguishing features (section 2.) I then discuss the different roles thought experiments can play in philosophical discussion, dividing them into three categories: discovery, exemplification (section 3) and justification. The latter is the most controversial use and I discuss it separately in section 4. I suggest that the use of thought experiments is not antithetical to the practice turn, at least if certain views of concept formation and hypothetical reasoning can be made good on. Section 5 further fleshes out the potential justificatory role of thought experiments by comparing them with case studies on the one hand, and with simulations on the other hand. It should be stressed that I do not reject, or even forcefully critique, the practice turn as such. I aim to stresses ways in which thought experiments can provide a specific sort of evidence, connected with scientific practice. In this spirit, section 6 suggests some ways of combining thought experiments with other methods – especially case studies and simulations.

Some parts of the present chapter, especially the comparison with cases studies, parallel an argument made in a paper jointly written with Adrian Currie (forthcoming) and receive a more extensive treatment there. To the extent that arguments made here diverge or go beyond that paper, they should be seen as representing my views, and aren't necessarily shared by Currie.

2. Some central features of thought experimentation

The debate over scientific explanation will serve as a useful source of examples for us (later I will also draw on recent work on science and values). That is because early discussions of explanation, especially from the 1950s and 1960s, occurred when thought experimentation was much more common relative to recent work. Consider the following example from Scriven (1962), in a paper devoted largely to combatting the "received view" of his time, namely the DN model. The DN model says that to explain an event one must deductively derive it from premises that include a natural law. One way in which Scriven argues against this model is by putting forward the following hypothetical scenario:

> "If you reach for a cigarette and in doing so knock over an ink bottle which then spills onto the floor, you are in an excellent position to explain to your wife how that stain appeared on the carpet, that is, why the carpet is stained (if you cannot clean it off fast enough). You knocked the ink bottle over. This is the explanation of the state of affairs in question, and there is no nonsense about it being in doubt because you cannot quote the laws that are involved, Newton's and all the others; it appears one cannot here quote any unambiguous true general statements, such as would meet the requirements of the [DN] model." (*Ibid.*, 198)

Scriven's argument draws support from the thought experiment. Specifically, it relies on the following judgement: the narrative about reaching for the cigarette and accidentally knocking over the ink bottle provides a decent explanation of the stain, even if one cannot cite any relevant laws and derive the ink stain from them. Although the text's chauvinist overtones are a bit jarring to current ears, the lesson Scriven drew from his thought experiment is quite widely accepted today: laws are unnecessary in explanation; often all that's needed to explain an event is a causal narrative leading up to it. Let us take this as a simple but typical instance of a thought experiment.

Generalizing from the Scriven case, we can regard a thought experiment as an episode in which one engages imaginatively with a certain scenario, typically hypothetical, forming a judgement about it. Thus we have two central elements: a thought experiment involves use of the imagination; and it results in a concrete judgement, often concerning whether a certain concept or principle applies in the case. Once we have these elements in place, we get something that permits (perhaps even calls for) manipulation – a third, follow-up feature. Let's look at each of these in turn.

Imagination. Part of the distinctive character of thought experiments – and part of the fun, too – consists in the fact that they are mini-fictions, little stories that call upon our imaginations. In putting forward a thought experiment, an author asks us to imagine a certain scenario or state of affairs.[2] I will not try to define imagination here, because that is a thorny task. But I will say that imagining typically involves entertaining the scenario without regard to whether it is true. In this, imagining contrasts with belief, which is typically truth-directed. Did Scriven actually explain to his wife the appearance of an ink stain on the carpet? Maybe he did and maybe he didn't. We can imagine the situation regardless.

Concrete conceptual judgements. As the Scriven example shows, when one performs a thought experiment one arrives at a judgement: this is a *bona fide* explanation, despite not including laws. But notice that this is a judgement about a particular case: it is not a claim about explanations in general but about *this* or *that* (putative) explanation. I consider this to be an important feature of thought experiments (Cf. Sorensen, 1992; Williamson, 2008). Its importance lies primarily in a feature to which we will come later in this chapter, namely that it allows us to view thought experiments as an exercise of a philosopher's implicit knowledge of scientific practice. But leaving that for later, for now we can focus on the structural aspect: a thought experiment asks us to make a judgement about the goings-on of a fictional scenario, judging whether it exhibits some state or property of interest, such an explanation. This allows us, in turn, to treat the evidence provided by thought experimentation as, broadly speaking, inductive evidence. It provides us with instances from which to draw general conclusions. In the Scriven case, the "data" given by the thought experiment (not just Scriven's but others relevantly like it) is that an explanation need not involve an appeal to laws.

Manipulation. Once we have these feature – imaginative engagement with a scenario, and generation of a concrete judgement, we have in effect a set-up which is manipulable. For we can posit other, similar scenarios, and generate judgements about them. We have a set-up in which we can tweak some aspect (the scenario in question or some specific feature thereof) and ask

[2] Sartori (2023) given an account of thought experiments which partly relies on viewing them as fictions. For more on the role of the imagination in models, thought experiments and related contexts see Salis and Frigg (2020).

how another aspect change along with it. Thus, in putting forward a thought experiment the author is not simply a telling a story for the reader's enjoyment or edification. She is, in effect, asking them to work through a scenario and to judge what depends on what. Does the fact that we have an explanation depend on the presence of laws? Or rather on the presence of a structured set of causal antecedents? Does explanitoriness depend on DN-like structure, or on causal content? Obviously, this is not a physical manipulation, as in an ordinary concrete experiment. But, arguably, it serves the same cognitive-epistemic role: it allows us to judge which changes "go together."

Having given a rough characterization of what thought experiments are, we are in a better position to describe their potential roles. I will do so in two parts. In the next section I describe relatively uncontroversial, albeit potentially important, roles: generating and exploring ideas and exemplifying a claim. The subsequent section discusses the more controversial use of thought experiments – justifying philosophical claims.

3. Thought experiments can function in illustration and discovery

The first and least problematic role that a thought experiment might play is as way of exemplifying a theory's implications or a circumstance in which it holds. Thus, if one already holds that portraying a causal narrative is a form of explanation, then Scriven's ink bottle thought experiment can be used as an example of such an explanation. Used in exemplification, the effectiveness of a thought experiment is dependent to a large extent on the communicative circumstances: what will serve as an appropriate illustration for one audience, in one context might not be for another audience in a different context. This is as with any attempt to provide an instructive example. Moreover, there may be a tradeoff between using a thought experiment for illustrative purposes and for other purposes. In particular: while working with an example that has relatively realistic scientific content can be important when it comes to discovery, and perhaps even more so to justification, it can often be a burden when it comes to exemplification: an example should be easy to convey and understand and should ideally presuppose as little background as is necessary to get the essential point(s) across.

A related role for thought experiments is to draw out an implication of a theory. Consider for instance, discussions of the counterfactual theory of causation. Almost any such discussion contains various simple thought experiments, often intended to point out implications of a particular variant of, or a condition within, the theory. For instance, Hall and Paul's extensive 2004 discussion contains about a dozen scenarios in which they envision two or more people throwing a rock at a bottle, with different verdicts as to which of the rocks is causes the bottle to shatter. For instance: "Suzy and Billy both throw rocks at the same bottle, but Suzy's gets there first, shattering it. If she had not thrown, then Billy's rock would have shattered the bottle a moment later." (Ibid., 54) Many of these thought experiments are meant to point to a particular implication of counterfactual views of causation which it would otherwise be hard to demonstrate: for instance, that there is a significant difference between different sorts of preemption – the quoted thought experiment, in particular, illustrates *late* preemption.

Another distinct role for thought experiments is to aid the discovery of new philosophical theories. This role has received a lot more attention with respect to thought experiments in science (Brown, 1986, 1991 and Norton 1996, 2004) but it also holds for philosophy. A thought experiment such as Scriven's, for instance, does not in and of itself indicate what an alternative to the DN account should look like. This might lead one to seek out scientific examples of a similar sort and to develop a more explicit view of causal explanation around them (Scriven 1959, 1969).

In a recent paper Praëm and Steglich-Petersen make an extended argument that thought experiments can and often do play a discovery role in philosophy. They suggest that there are a number of distinct ways in which they may do so. A thought experiment may pose a challenge to the necessity of a thesis, and thereby suggest what is missing or how an alternative might look – as we just noted in the case of Scriven. But it might also challenge the sufficiency of a thesis. Thus, Russell's well-known objection to the regularity view of causation: when the hooters in Manchester's factories sound this is regularly followed by the workers of London leaving their posts. But the hooters sounding in Manchester does not cause the workers in London to leave their posts. Looking at this case, and subtle variants of it, Mackie (1980, 81-87) proposes to supplement his regularity view of causation by requiring that causes and effects be connected via some process or mechanism. (It is beyond our scope here to discuss what exactly this amount to.)

This is not an exhaustive taxonomy. At the very least, there may be more open-ended thought experiments, as Praëm and Steglich-Petersen point out. These are suggestive of a theory but not by exposing a flaw in an earlier one. To pick a philosophy of science example: Glennan 1996 considers a possible explanation for the operation of a toilet. As he discusses, an appropriate explanation looks inside the device and traces its parts, their relative positions and interactions and how this brings about the system's overall operation. He follows this lead to suggest a mechanistic view of causation, which he illustrates in part via more realistic scientific examples. Glenan's is partly an illustrative thought experiment. Thus, the taxonomy is also not exclusive.

The exemplification and drawing out of implications roles for thought experiments might not be your cup of tea – you might prefer to embellish your paper with real-life science; or you might, as a reader, comprehend better when examples are less abstract and contrived. But it is hard to raise a principled objection here. To each his own example. In contrast, one might worry that thought experiments never play a genuine role in discovery. Perhaps they occur to philosophers only after the fact, once they have a theory and wish to motivate it? As Praëm and Steglich-Petersen say, it is often hard adjudicate this. In some scientific contexts, we have good evidence about the historical trajectory and can show that a thought experiment led to a theory and not vice versa (Norton, 1991). Or this might simply be a function of the relatively fluid nature of the discovery/justification distinction: whether a theory responds to a thought experiment, or gives rise to it, is often hard to tell apart. This leads to my next topic, the role of thought experiments in justification.

4. Can thought experiment justify?

A more controversial (and arguably more central) role a thought experiment might play is to justify a philosophical thesis. Specifically, thought experiments are sometimes used to generate evidence for or against philosophical claims. This is a controversial idea primarily because thought experiments seem to reflect our minds: they are a way of teasing out the concepts and conceptions with which we operate. But if so, then why think there is any objective standing to the verdicts we reach via thought experimentation? We might call this the "a prioriness" worry. Another, milder worry, is the "how" worry: if we accept that thought experiments justify, then how should we use them? I discuss a prioriness here. The how-to-use issue will occupy the next two sections.

There are several ways to respond to a prioriness concerns. A concessive response agrees that thought experiments cannot, at least not on their own, serve to justify philosophical claims. This need not involve denying thought experiments any justificatory role whatsoever. A person can refuse to grant thought experiments *independent* justificatory status while allowing that they can do so in combination with other tools, either empirical tools like real (psychological) experiments and surveys, or theoretical ones like simulations and formal models.

A second response is to say that thought experiments can play a justifying role, and that they do so in an a priori manner *i.e.,* in virtue of revealing how our concepts and conceptions work. (Chudnoff, 2013; Brown, 1991) On one conception of scientific rationality – by which I mean the set of norms governing central scientific practices such as explanation, theory construction, testing and confirmation and the like – its shape is determined largely a priori, inasmuch as rationality is a not an empirical but a normative notion. I suspect that this sort of response would be unattractive to most current philosophers of science, post practice turn, who assume that discussions of such normative notions should be at least partly sensitive to how science actually works.[3]

This gives rise to a third response which says, in essence, that thought experiments can provide justification inasmuch as they allow us to access practice, albeit indirectly. In this way of understanding (the justificatory role of) thought experiments they bypass the a prioriness worry because, in the end, they are not that *a priori*. They are a method for fleshing out concepts and conceptions that are themselves reflections of scientific practice, because they are formed in response to it, through familiarity with it. This response, in other words, is grounded in a kind of empiricist theory of concepts, one on which they reflect our experience.

There are several ways of developing such a theory of concepts, and I suspect that any number of them would allow us to offer the same sort of response to the a prioriness worry. But

[3] My own view is that the appeal to practice as a justificatory basis is grounded in the idea that science is, at least some times and in some respects, a successful epistemic enterprise. We appeal to (the relevant parts of) science because we see it (them) as a good epistemic model. But any rationale for appealing to scientific practice will have broadly similar results here, I think.

let me flesh out the point by briefly describing a theory that has recently been advanced, with thought experimentation in mind, by Michael Strevens (2019). Strevens construes concepts as categorization devices. A concept groups instances into a set, and attributes to members of that set some underlying cluster of properties which accounts for the set's members. This is a version of the 'theory theory' of concepts. As Strevens puts it with reference to an example: "the concept of (say) a horse … is a theory of horses, or to put it more plainly, a set of beliefs about horses representing some of their appearances, behaviors, their relations to other horses, and so on. The beliefs constituting such a theory represent explanatory rather than metaphysical facts – the capacities of horses, their susceptibilities and other facts about their place in the causal economy, rather than hypotheses about what ultimately makes something a horse. Their epistemic status is equally mundane: they are vulnerable to empirical disconfirmation or indeed rethinking of any sort." (2022, 303) Strevens goes on to develop a theory of reference to go along with this view of concepts, and to combine it with some other ideas to generate what he thinks of as a vindication of the use of thought experiments, and the so-called 'method of cases' more generally.[4] I won't review Strevens' theory in its full complexity, only highlight the essential idea: a thought experiment is an exercise of our ability to pick out the members of a category, where the category is shaped, at least in part, in response to ordinary observation and empirical knowledge.

I think we can apply something like this view to elucidate the workings of thought experiment in philosophy of science, while adding a specific element (which Strevens also hints at, 2019, pp. 209-210):[5] one's concept of explanation (say) can be seen as a set of beliefs, which pick out *bona fide* explanations. It is based, in part at least, on one's acquaintance with scientific practice, specifically with explanatory practice. It follows from this idea that philosophers of science use their knowledge of scientific practice in conducting thought experiments. They exercise their expertise – which consists, in part, of a familiarity with science and with explanatory practice in particular – in making judgments as part of thought experiments (and in revising these judgments, and constructing theories on their basis, etc.) In this way, the response to the a prioriness worry ends up being compatible with the practice turn. Specifically, it is compatible with treating scientific practice as an arbiter – not the sole arbiter, perhaps, but an important one – for claims in philosophy of science.

As Strevens notes, and as I want to stress, the theory of concepts he proposes is speculative. It is grounded in part in empirical psychological ideas that can be tested. So to the extent that the use of thought experiments depends on it, it should be treated as provisional. Nevertheless, it gives us a leg to stand on. The other leg – to which the next section is devoted – concerns the uses and advantages of thought experiments. Do they probe practice in an effective, distinctively effective way, such that we should seek to consult them when developing philosophical theories?

[4] Several critiques, and a response by Strevens', appear the same issue of *Analysis* as the above cited paper. See also Buckwalter (2019).

[5] The idea I am alluding to here is sometimes called "the expertise defense" in discussions of thought experiments. See Machery, 2017 (Ch. 7) for a critical stance.

I suggest that the answer is a qualified 'yes'. To flesh this out, I'll compare thought experiments to case studies, on the one hand, and to simulations, on the other hand.

5. Thought experiments in a comparative light

My comparison is deliberately selective. The aim is to expand on the previous discussion by highlighting ways in which thought experiments can give us something extra, from a justificatory point of view, but also to point to their shortcomings. The overall message is that relative to case studies, thought experiments give less direct but more manipulable access to practice. Simulations, meanwhile, allow for more specific and accurate manipulations, but aren't, as such, grounded in practice.

Thought experiments versus case studies. Case studies involve the description and analysis of an episode of actual scientific practice. Some historical research is undertaken out of an intrinsic interest in the past, or in order to understand a larger historical question. But case studies, as I construe them here, are performed in the service of a philosophical argument. The early work on explanation described above, which included thought experiments such as Scriven's, can be contrasted with more recent writing on explanation, which puts a premium on case studies. The large volume of writing on mechanistic explanation in the 21st century, for instance, has relied on cases from genetics (Darden 2002), Cell biology (Bechtel, 2005) and neuroscience (Craver, 2007). Similarly for work on non-causal explanation (Lange, 2016; Saatsi and Ruetlinger, 2017). Authors in these areas tend to argue for their view, in large part, by showing that it corresponds to cases drawn from scientific practice.

More generally, cases studies aim to demonstrate, cast doubt on, or offer support for, a more general set of claims. From a justificatory point of view, the comparison with thought experiments turns on two central aspects. On the one hand, case studies constitute a more direct, full-blooded engagement with scientific practice. On a view like the one described earlier, in which thought experiments exercise conceptual abilities that are sensitive to practice, it would not be correct to say that thought experiments are disconnected from practice. But they do not have the rich, textured content that well-executed case study has. If we want to test philosophical theories against the details of actual scientific practice, case studies have an in-built advantage.

While it is possible, in principle, to construct a very elaborate thought experiment, one that would depict a made-up scientific episode in considerable detail – not science fiction but a scientific fiction, if you will – there isn't much to recommend that. I think there are two related reasons for this. First, the point of thought experiments is the controlled exercise of our conceptual resources. This is best done on skeletal, abstract, "clean" cases. Relatedly, we have seen that part of the point of thought experiments is manipulation. It gives us the ability to examine what (conceptually) depends on what. To manipulate – even in our minds – a fully detailed depiction of a piece of science – even a fictional one – is complicated, with less certain results. So the sparsity of thought experiments is conducive to manipulability.

Here we can draw an analogy – or perhaps more than an analogy – with the distinction between observation and experiment within science itself. An observation is truer to the phenomenon and less disruptive. But it is also less probative, because it lacks manipulability. The analogy is incomplete and not tight, but it directs us to the relative advantages of thought experiments and case studies. This gives rise to the suggestion, discussed in the next section, that we should combine the methods, rather than privilege one over the other.

Thought experiments versus computer simulations. A computer simulation takes a formal version of a scenario and runs it, typically many times, to discern how it is likely to unfold. Simulations can play some of the same roles as thought explements (Mayo-Wilson and Zollman, 2019). Indeed, a well-executed simulation will often allow one to say more about dependency relations: in complex set ups, our minds are far less capable at discerning what depends on what and how than a computer. In this respect simulation is more powerful than thought experiments.

This is why simulations not infrequently lead to surprising results, in ways that would be hard to obtain through thought experimentation. Consider what's the so-called Zollman effect: the sharing of scientific results, in conditions of uncertainty, can decrease the odds that a scientific community arrives at true results (Zollman, 2007). It is hard to see how anything beyond a vague intuition about this type of phenomenon (not to mention an understanding of the exact conditions under which it occurs) can be attained via thought experimentation.

On the other hand, when one runs a simulation, one is typically engaging in an *entirely theoretical* exercise. A simulation, as such, can have internal validity; one can verify, by inspecting the simulation itself, that one has simulated what one wished to. But this falls short of external validity – a (potential) match with the phenomenon in the world. If views like Strevens', which allow us to anchor thought experiments in experience with scientific practice, are correct, then thought experiments have a kind of external validity "built in". In this respect, thought experiments have an edge over simulations.

Moreover, in some cases, at least, it is hard to see how one would get a similar result by way of simulation, rather than thought experimenting. Consider again the Scriven thought experiment. It outlines a hypothetical scenario, and asks us to make a judgement: is it an explanation or not? I am not sure we can, even in principle, run a parallel computer simulation. Who would make the relevant judgement – the computer?[6] It seems that, at least in a case like this, the whole point of running a thought experiment is to allow us to use our conceptual repertoire to make judgments that we later rely on in developing a philosophical theory. While manipulation has an important role to play in this sort of conceptual exercise, and while computer simulations permit more accurate and more elaborate manipulation, still a simulation cannot fully substitute for thought experiments because simulations are not expressions of our concepts (at least not in the same sense.)

[6] Thus, thought experiment have a special role via-a-vis normative epistemological questions. This is in contrast to mundane, concrete matters, where I'd argue for a more skeptical approach (Kinberg & Levy, 2023).

Let me make one final comparative point. For all that I have said, it is not obvious what recommends the prevalent practice in which philosophers perform thought experimentation *in the armchair*. Specifically, why prefer the philosopher ruminating on their own to an experimental philosophy study in which the same fictional episode is put to a larger sample of thinkers, and results are collated and analyzes statistically? My answer is that the "real" experimental method is, at least in most cases, preferable. It can allow us to avoid bias and idiosyncrasy and to get a more representative picture of philosophical judgements, including their distributional properties (mean, mode, variance and the like). That said, I think that, for one, thing, the above defense of thought experimentation—its "expertise" aspect, in particular—suggests that such experimental work should target the judgements of philosophers of science (and perhaps scientists, too). It is their concepts that we want to examine, given their familiarity with the scientific practice. Second, I think that even in the absence of formal "x-phi" data, thought experiments can play a justificatory role. The familiar philosophical practice in which a philosopher produces a thought experiment, makes a judgment about it, and publicizes it, thereby allowing peers to examine it and make their (hopefully independent) judgements about it is also of value, even if more rigor in eliciting such judgements and aggregating them is desirable.

6. *Combining thought experiments with other methods*

So far I've discussed discovery-related, exemplification and justification-related roles for thought experiments. That discussion should already give some indications about the kinds of uses that I think thought experiments can have. Let me now expand on one important point in this regard: bettering the justificatory use by combining thought experiments with other methods. I'll mainly look at how thought experiments combine with case studies, because that is the tool that seems most complimentary, in terms of its strengths and weaknesses, to thought experimentation. I'll also say a few words about combining thought with simulations and with "real" experimental work, primarily experimental philosophy.

Combining case studies and thought experiments. As noted previously, case studied provide us with richer, but less manipulable information about scientific practice. Thought experiments have the converse profile. So to get rich grounding in practice, and also test our claims via manipulation, we would do well to combine these methods. In principle, even a state where, at the level of the entire field, both thought experiments and case studies are used is one that involves a degree of mixture. But a more important form of combination is when thought experiments and cases studies are use *in one and the same text*. This is relatively rare in current philosophy of science, but let me mention some examples nonetheless.

A recent paper by Menon and Stegenga mixes real-life cases and hypothetical scenarios in an interesting way. These authors set out to provide a novel account of the involvement of values in science – a so-called "difference-to-inference" model. In the course of developing the account they use multiple variants of a thought experiment involving two researchers, which operate on the basis of different values, affecting their reasoning. But they also appeal to real-life cases, such as Ronald Fisher's rejection of the smoking-cancer relation, which they construe as a case of "reasoning to a foregone conclusion" – an unhealthy involvement of values, in their reckoning. The combination gives their account both a kind of robustness and a clarity in its anchoring to actual practice.

This example, and some others like it (though they aren't very common) raises questions about the relationship between the content of the case studies and thought experiments – how should it look? One basic and fairly lax constraint is that the case and the thought experiment pertain to the same philosophical question. Menon and Stegegnga, as I have noted, support some aspects of their argument with the aid of examples and other aspects via thought experiments. So their work does seem to meet the basic constraint. There are several other options, however, involving tighter relationships. One option is to begin with a real-life case and then "convert" it into a hypothetical (and simpler) case, and potentially more than one such case, to be treated as a thought experiment. This would allow the author and reader to get a grip on whether the features co-instantiated by the case are robustly connected and thus whether the claim under consideration is indeed supported by it.

A recent paper by Dellsén (2016) does this, though in a relatively minimal way. Dellsén discusses the question of whether scientific progress is aimed at knowledge or at understanding. To support the latter option, he presents a (brief) case study: Einstein's explanation of Brownian motion in one of his famous *annus mirabilis* papers (Einstein, 1905/1956). Dellsén presumes that this is an undisputed case of scientific progress. But Einstein had very partial information about the phenomenon he was accounting for, and could not have known that he had an explanation of Brownian motion. So he could not have had knowledge of the explanation – obviating the idea that progress is a matter of attaining knowledge. Dellsén then considers whether someone might dispute the idea that the kinetic theory of heat, and the phenomenon of Brownian motion, were in fact unknown at the time of Einstein's work. Still, he says, "we can easily imagine a world in which Einstein's explanation was put forward before the kinetic theory of heat became sufficiently justified to be known (e.g. shortly after James Clerk Maxwell first presented his kinetic theory in 1859). In that case, certainly, neither the explanandum nor the explanans in Einstein's work would have been known at the time. None of this would take away from Einstein's achievement, which was to show how Brownian motion is explained by the kinetic theory of heat." (ibid., 76).

Although minimal, I take this to illustrate a case study that is "converted" into a thought experiment. The real-life case is subject to an objection, and Dellsén answers this objection by performing a manipulation – eliminating knowledge in a more thorough way – to show that the relationship at issue –progress sans knowledge – still holds.

There are other possible trajectories. I will mention them although I am not aware of actual examples. One option is to come up with a hypothetical scenario and as a second step (potentially after several variants of the thought experiment have been run) search for real-life instances that match it. This way of doing things allows one to first probe the robustness of the connection alleged to hold, and then anchor it more directly in scientific practice. Another option is to try to come up with general maxims or constraints via a thought experiment, or a set of thought experiments. And then to fill in, adjust and refine those via a look at practice, especially via case studies. I think this is roughly what occurred with the move from the DN view to the causal approach to explanation – the example we began with. To recall, the DN model faced counterexamples in the form of thought experiments – from Scriven, as well as from other authors – which also pointed the way forward, because they suggested a causal approach. That approach was then shored up and refined through a wide range of case studies throughout the 1970s-1990s.

A few words on other combined uses. I also think thought experiments can be fruitfully combined with other methods, of which I will comment on two. The most obvious, already mentioned, is to combine thought experimentation with "real" experimentation – especially of the experimental philosophy variety. Perhaps 'combination' is not the right word here, as the idea is to perform the same thought experiment, or variant thereof, but in a larger sample and in a more systematic way (hence the scare quotes around 'real'). Nonetheless, I think that some of the same justifications for appealing to thought experiments, especially in their justificatory role, apply to their "real" counterparts.

Finally, I think thought experimentation can be readily and usefully combined with simulations. In particular, I think simulations can often allow us to more rigorously test and flesh out scenarios that have their initial life in a thought experiment (in this regard my view is consonant with that of Mayo-Wilson and Zollman, mentioned above). Consider an example from a recent paper by Uwe Peters (2021). Peters advances the thesis that confirmation bias, a hinderance to proper evaluation of evidence and arguments at the individual level, can have beneficial group-level epistemic effects. A central element of his argument is the following thought experiment:

"Suppose there is a group of five scientists trying to answer one of the still open questions in science, such as where life comes from ('primordial soup', a meteorite, and so on). Each of the scientists has a confirmation bias toward a different explanation of the phenomenon. As it happens, none of the five proposals enjoys more empirical success than any other. Suppose the scientists have four weeks to explore the issue and determine the most plausible account among the five views… Suppose that each of the five scientists can, and is instructed to, impartially assess all five views, and determine the most plausible through group discussion.

Suppose too that they all follow the instruction. They suspend their confirmation bias towards their own view and evaluate each of the proposals equally critically and with dispassion.

While this might seem to be the epistemically best distribution of research effort, it has a significant side effect. A confirmation bias towards a particular view, *V*, will tend to push scientists to persistently search for data supporting *V* and to invest effort in defending it... the bias may incline a scientist to consider rejecting auxiliary assumptions to *V* rather than the proposal itself…" (*Ibid.* 1069-1070)

This is an interesting thought experiment and I think that even as it stands it provides some initial support for Peters' main claim. But I suggest that it would provide firmer support if converted into a formal model and simulated. We would then be able to more readily understand the process Peters describes, gauge under what conditions it leads to the epistemic goods he envisions, and compare it to other setups. After we have some such simulations results, we may wish to go back to the drawing table and come up with a different scenario, going through the process again.

Now, admittedly there are some subtle differences between this thought experiment and ones described before. One of them is that the Peters one involves potentially complex dynamics, such that our judgement about the normative (epistemic) sides of the scenario will depend on them. It might be that these specific features make the thought experiment simulation combo attractive. There may be different cases, too. These are instances in which combining thought experiments with simulations, among other philosophical tools, can prove useful.

References

Bechtel, W. (2005) *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge University Press.

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

Brown, J.R. (1986). Thought Experiments since the Scientific Revolution, *International Studies in the Philosophy of Science*, 1: 1–15.

Buckwalter, W. (2019). Review of "Thinking Off Your feet". *Mind* 130(517): 307–320.

Chudnoff, E. (2013). *Intuition.* Oxford University Press.

Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.

Currie, A. and Levy, A. (forthcoming). Bringing Back Thought Experiments into the Philosophy of Science.

Darden, L. (2006). *Reasoning in Biological Discoveries: Mechanism, Interfield Relations, and Anomaly Resolution*. Cambridge University Press.

Dellsén, F. (2016). Scientific progress: Knowledge versus understanding. *Studies in History and Philosophy of Science*. 56: 72-83.

Einstein, A. (1905/1956). *Investigations on the Theory of the Brownian Movement* (A. D. Cowper, Trans.). Dover.

Glennan, S. (1996). Mechanisms and the Nature of Causation. *Erkenntnis*. 44: 49-71.

Kinberg, O. and Levy, A. (2023). The Epistemic Imagination Revisited. *Philosophy and Phenomenological Research*. 107(2): 319-336.

Lange, M. (2016). *Because without Cause: Non-Causal Explanations in Science and Mathematics*. Oxford University Press.

Machery, E. (2017). *Philosophy within Its Proper Bounds*. Oxford University Press.

Mackie, J. (1980) *The Cement of the Universe*. Oxford University Press.

Mayo-Wilson, C. & Zollman, K. (2021). The computational philosophy: simulation as a core Philosophical method. *Synthese. 199: 3647–3673*

Menon, T. and Stegenga, J. (2023). The Difference-to-Inference Model for Values in Science. *Res Philosophica.*

Norton, J. (1991). Thought Experiments in Einstein's Work. in T. Horowitz and G. Massey (eds.), *Thought Experiments in Science and Philosophy.* Rowman & Littlefield

Norton, J. (1996). Are Thought Experiments Just What You Thought?, *Canadian Journal of Philosophy*, 26: 333–366.

Norton, J. (2004). On Thought Experiments: Is There More to the Argument? *Philosophy of Science*, 71: 1139–1151.

Peters, U. (2021). Illegitimate Values, Confirmation Bias, and Mandevillian Cognition . *Synthese.* 199: 3647–3673.

Praëm, S. and Steglich-Peterson, A. (2015). Philosophical Thought Experiments as Heuristics for Theory Discovery. *Synthese*, 192: 2827–2842

Saatsi, J. and Ruetlinger, A. (2017). *Scientific Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations*. Oxford University Press.

Salis, F. and Frigg, R. (2020). Capturing the Scientific Imagination. In Levy, A. and Godfrey-Smith P. (Eds.), *The Scientific Imagination: Philosophical and Scientific Perspectives*. Oxford University Press.

Sartori, L. (2023). Putting the 'Experiment' back into the 'Thought Experiment'. *Synthese 201: 1-36.*

Scriven, M. (1959). Explanation and Prediction in Evolutionary theory. *Science*, 130: 477–482.

Scriven, M. (1962). Explanations, Predictions and Laws. *Minnesota Studies in the Philosophy of Science*. 3: 170-230.

Sorensen, R. (1992). *Thought Experiments.* Oxford: Oxford University Press.

Strevens, M. (2019). *Thinking Off Your Feet: How Empirical Psychology Vindicates Armchair Philosophy*. Harvard University Press.

Strevens, M. (2022). Précis of "Thinking Off Your Feet". *Analysis* 82(2) :303-306.

Williamson, T. (2008). *The Philosophy of Philosophy*. Oxford University Press.

Zollman, K. (2007). The Communication Structure of Epistemic Communities. *Philosophy of Science*, 74(5): 574–587.