Studies in History and Philosophy of Science 83 (2020) 17–27

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: http://www.elsevier.com/locate/shpsa

The material theory of induction and the epistemology of thought experiments

Michael T. Stuart

University of Geneva, Department of Philosophy and Centre for Philosophy of Science, Rue de Candolle 2, Geneva 1211, Switzerland

ARTICLE INFO

Keywords:
Material theory of induction
Thought experiments
John D. Norton
Role of logic in reasoning
Pluralism
Empiricism

ABSTRACT

John D. Norton is responsible for a number of influential views in contemporary philosophy of science. This paper will discuss two of them. The *material theory of induction* claims that inductive arguments are ultimately justified by their material features, not their formal features. Thus, while a deductive argument can be valid irrespective of the content of the propositions that make up the argument, an inductive argument about, say, apples, will be justified (or not) depending on facts about apples. The *argument view of thought experiments* claims that thought experiments are arguments, and that they function epistemically however arguments do. These two views have generated a great deal of discussion, although there hasn't been much written about their combination. I argue that despite some interesting harmonies, there is a serious tension between them. I consider several options for easing this tension, before suggesting a set of changes to the argument view that I take to be consistent with Norton's fundamental philosophical commitments, and which retain what seems intuitively correct about the argument view. These changes require that we move away from a unitary epistemology of thought experiments and towards a more pluralist position.

1. Introduction

John D. Norton has had a tremendous impact on contemporary philosophy of science, not just through his written work, but also through his efforts to increase open-access policies in philosophy of science, and by using his role as director of the Center for Philosophy of Science at the University of Pittsburgh to make the culture of philosophy of science more inclusive and collegial. In terms of his written work, he has published over one hundred articles and chapters on topics ranging from quantum mechanics to the paradoxes of sailing (Norton, 2012). At least one-third of his work concerns relativity, and more than one-fifth spotlights Einstein, many of whose successes, failures, and cryptic remarks Norton has illuminated (some personal favourites: Norton, 2005a; 2010a, 2013). His work on induction (one-tenth) and thought experiments (one-tenth) also make up substantial pieces of his philosophical pie.

The Nortonian corpus is united by several fundamental commitments. First, his philosophical claims are always connected to and informed by actual scientific practice. Second, there is a consistent focus on the epistemology of scientific reasoning. Third, Norton is a committed empiricist. Each of these are identifiable in his work on induction and thought experiments. For example, according to his material theory of induction, it is facts about the world discovered through experience that justify the inductive inferences that amplify our knowledge, and Norton argues for this using case studies drawn from the history of science. In his work on thought experiments, Norton argues for an empiricist position by claiming that thought experiments are just arguments, and their epistemology is just the epistemology of arguments. Again, he supports his position using historical case studies.

Given that Norton always sails by the same philosophical compass, it is natural to assume that the conclusions he reaches on different topics will be mutually supportive. Still, Norton himself hasn't anywhere explained how his ideas about induction and thought experiments fit together. In the next section I will present Norton's views on induction. Then I will discuss his work on thought experiments to show that, surprisingly, there is a tension between the two. I will examine several possibilities for easing this tension, and argue that one preserves more of Norton's claims and fundamental commitments, as well as what seems right to me about the argument view.

2. The material theory of induction

For Norton, we can characterize the power of induction to amplify knowledge formally or materially. Formal theories attempt to do for

E-mail address: michael.stuart@unige.ch.

https://doi.org/10.1016/j.shpsa.2020.03.005
Received 5 April 2019; Received in revised form 19 January 2020; Accepted 25 March 2020
Available online 11 April 2020

induction what logicians have already done for deduction: identify universal schema for inference, like modus ponens and tollens, which are valid under all substitutions of content. To do this, formal theories of induction separate the material aspects of an inductive inference (the content) from its formal aspects (the structure) in order to isolate what is common to the structures of successful inductive inferences. According to formal theories of induction, "valid inductive inferences are distinguished by their conformity to universal templates. They may be simple, such as the template that licenses an inference from some past A's being B to the conclusion that all A's are B. Or they may be more complicated, such as the requirement that degrees of inductive support conform to the probability calculus" (Norton, 2014, 673).

However, Norton argues that formal theories of induction have not, and cannot succeed. Take the following example. Suppose we find that a pure sample of bismuth has a melting point of 271 °C. From this, we infer inductively that all pure samples of bismuth will have that melting point. Formal theories justify this induction with reference to formal features of the universal schema that the inference instantiates, which in this case, is enumerative induction. But, and here is the problem with formal theories, the fact that a sample of *wax* melts at 91° does *not* tell us that all other samples of wax will melt at 91° (Norton 2003, 649). Why not? Both inferences, the good one about bismuth and the bad one about wax, instantiate exactly the same formal schema: some past A's are B, therefore all A's are B. In the case of deduction, the meaning of the logical connectives guarantees the truth of the conclusion given the truth of the premises. But this can never be the case for induction. In a series of papers and forthcoming book, Norton argues that no formal theory of induction is capable of identifying universal schema for successful inductive inferences (2003, 2005b, 2010b, 2014, forthcoming).

Individual enumerative inductions can of course be warranted, but it is not formal properties that warrant them: it is background facts about the subject of the inference. Norton calls these background facts "material principles." In general, the material principles that "authorize" inductive inferences are truths of the domain we are interested in. In the case of bismuth, the material principle is the fact that pure samples of chemical elements all have the same melting point (and this fact is underwritten by other facts in chemistry and physics) (2003, 650). Such facts "warrant a local mini-logic, peculiar to the context," in which we are licensed to proceed from some particular evidence to a more general conclusion (forthcoming ch. 1, 23). In the case of bismuth, the local mini-logic might be: If x is a pure sample of chemical element y, and x has a melting point of z, generally, all other pure samples of y will have melting point z. This is only a mini-logic because we cannot eliminate reference to terms like "chemical element" and "melting point." And it is the "generally" that makes the inference inductive and not deductive. The appearance of that word indicates an admission of inductive risk. Bismuth, or the next element, might be the first to undermine our material principle.

To put Norton's argument against formal theories of induction another way, he notes that these theories have tried to follow the example of deductive logic, but in fact they should have proceeded in a manner that is just the opposite. Deductive logic can be pursued relatively independently of experience, and its results can then be applied to identify and evaluate actual inferences. For induction, justification proceeds in the other direction: We discover what is generalizable first, via empirical investigation, and through these discoveries (material principles) we decide which inductive inferences are warranted.

3. The argument view of thought experiments

In 1986, James R. Brown pointed out (following Kuhn, 1977) that thought experiments in physics seem to provide new knowledge that comes neither from valid logical inference, nor from new experience. According to Brown, the best explanation for the success of some thought experiments is that they occasionally (and fallibly) allow us to "see" or "intuit" the relations between universals that constitute the laws of nature (Brown, 1986, 1991, 1992, 2004, 2007). Norton denies Brown's conclusion: thought experiments never provide a priori empirical knowledge about nature. Instead, they are just arguments. Since arguments merely rearrange existing knowledge, no rationalist insight will be required to explain scientific advancement via thought experiments (Norton, 1991; 2004a; 2004b). This position has been dubbed "the argument view" (e.g., Bishop, 1998, 22; Brendel, 2018, 291), and it serves several purposes for Norton: it defends empiricism, gives Norton a way to evaluate a tool wielded so powerfully in science by Einstein and others (see e.g., Norton 1991, 2013, 2018), and most importantly, tells us how "thought experiments are supposed to give us knowledge of the natural world." We should like to know, "From where does this knowledge come?" (Norton, 2004b, 44). The answer is, from previous knowledge, augmented or amplified through deduction or induction.

The account has developed over the years, becoming "a kind of moving target" (Brendel, 2018, 283). Following Elke Brendel (2018, 283), we can divide the argument view into seven theses:

> *Identity Thesis*: Thought experiments are (type)-identical with arguments.[1]
>
> *Reconstruction Thesis*: Thought experiments "can always be reconstructed as arguments" (Norton, 2004a, 1142).[2]
>
> *Reliability Thesis*: A thought experiment is a "reliable mode of inquiry" only if the argument into which it can be reconstructed justifies its conclusion (2004b, 52).
>
> *Elimination Thesis*: "Any conclusion reached by a (successful) scientific thought experiment will also be demonstrable by a non-thought-experimental argument" (Gendler, 2000, 34).[3]
>
> *Epistemic Thesis*: Thought experiments and the arguments associated with them have the same epistemic reach and epistemic significance.
>
> *Empirical Psychological Thesis*: "The actual conduct of a thought experiment consists of the execution of an argument" (Norton, 2004b, 50).[4]
>
> *Empiricist Thesis*: The result of a thought experiment can only come from experience: "The result of a thought experiment must be the reformulation of … experience by a process that preserves truth or its probability" (Norton, 2004a, 1142).

These theses are supported by various considerations. Here is my own reconstruction of what I think is the most detailed and mature version of the argument that connects them all (drawing on Norton, 2004a; 2004b).

A) Assume empiricism. Good empiricists should maintain a healthy skepticism against "epistemic magic" (Norton, 2004b, 45). For empiricists, we only obtain and increase our knowledge by experience and logical manipulation of that experience, and we should develop an epistemology of thought experiments that reflects this. (Empiricist thesis)

B) As a matter of fact, we notice that all the thought experiments we have investigated can be reconstructed as arguments.

   Inductive step: generally, *all* thought experiments will be reconstructible as arguments. (Reconstruction thesis)

C) As a matter of fact, we notice that all the thought experiments we have investigated have the same epistemological strengths and weaknesses as their logical reconstructions.

[1] See Bishop (1999) and Häggqvist (2009) for criticism.
[2] See Brown (1992, 275) and Stuart (2016) for criticism.
[3] See Gendler (1998) for criticism.
[4] See Nersessian (1992) and Brendel (2018) for criticism.

Inductive step: generally, *all* thought experiments will be equally justified as their logical reconstructions.

D) If a thought experiment is not an argument, we need some other way of telling whether the thought experiment is justified or not. And as a matter of fact, there is no such way that does not reduce in the end to logic (Reliability thesis)

E) From A), B), C) and D): thought experiments are eliminable from discussions about the epistemology of science, at least in principle. (Elimination thesis)

F) From A), which implies a desire for ontological parsimony, plus B), C), D), and E): we can treat thought experiments as arguments. (Identity thesis)

G) From C), D), E) and F): we are licensed to pursue the epistemology of thought experiments exclusively as the epistemology of arguments. (Epistemic thesis)

H) The only (or best) explanation of B), C) and D) is that thought experiments are not just treatable as arguments; they are exactly the same as arguments, even *psychologically*. The execution of a thought experiment is thus always just the execution of an argument. (Empirical Psychological thesis)

This is the argument for the argument view.[5] Its success depends heavily on what Norton means by "argument." Logic textbooks tell us that arguments are sets of statements that admit of the division between premise and conclusion. Norton demands an even more encompassing definition: "I stress here that this thesis invokes a notion of argumentation that is far more general than the one usually invoked in logic texts" (Norton, 2004b, 52). Indeed, Norton allows that diagrams may be premises (2004b, 58), premises may be tacit (1996, 339) and inferences may be tacit (1991, 142-3 note 2).

More positively, Norton remarks that anything "licensed" or "governed" by a deductive, inductive, abductive, or informal logic is an argument (2004b, 64 and 2004a, 1140 respectively). Norton's notion of argument therefore also depends on his notion of logic. For Norton, logic is any systematization that distinguishes between form and content, and focuses on form (2004b, 53). It need not work upon propositions, e.g., Bayesianism is a logic despite the fact that it can operate upon probability distributions rather than propositions. A logic also need not preserve truth or operate only using true premises, and it need not be restricted to systems that avoid contradictory conclusions (2004b, 53). An argument, therefore, is anything that has formal as well as material components, and whose formal components allow logic to classify it as (in)valid or (non)cogent. Finally, even if something can only be classified and evaluated by a future logic, it may still count as an argument now (2004b, 54–55).

The argument view aims to explain how scientific thought experiments increase our knowledge. For Norton, there is a defining "mark" that tells us whether a given thought experiment is valid or cogent. The mark is completely internal: a structural feature that can be identified by "merely reading its text" (2004a, 1143). The mark identifies justified inferences by flagging formal features, including the logical relations that connect conclusions to premises. While Norton claims that there is "something in the logic that evidently confers the power of a thought experiment to justify its conclusion" (Norton, 2004b, 54), I think we should read this claim not in the sense that logic *itself* justifies arguments, but that it identifies and explains the justification already present in valid or cogent arguments, which they have due to their logical properties. For instance, an argument that instantiates modus ponens is a good one not simply because logicians say so, but because the logical form of modus ponens is such that it guarantees the truth of the conclusion given the truth of the premises.

There are questions that it would be natural to raise at this point. For one, Norton tends to stress validity and cogency over soundness. For example, he claims that the mark of a good thought experiment "is *just* that the thought experiment either uses an argument form licensed by a logic or can be reconstructed as one" (2004b, 54; my emphasis). This may be worrying, because an argument will only give us new knowledge if it is sound: validity and cogency are not enough. Also, there is a threat of triviality, since, for Norton, arguments are whatever can be evaluated by a present or future logic, and ideal future logics will eventually be able to reconstruct all inferences. Thus, every inference becomes an argument by definition (see Stuart, 2016). But in any case, the present purpose is not to argue against Norton's argument view, but rather to see how it combines with the material theory of induction.

4. Combining the two accounts

To combine the material theory of induction with the argument view of thought experiments, we must first identify some inductive thought experiments, i.e., thought experiments that, on Norton's view, are to be identified as inductive arguments. For Norton, these are all the thought experiments that cannot be reconstructed as deductive arguments, thus, any that proceed via inference to the best explanation or diagrammatic reasoning, or whose conclusion is logically stronger than its premises. For example, in Einstein's elevator thought experiment, one observer is placed in an opaque chest that is in a gravitational field (e.g., on the surface of the earth), while another is placed in a different opaque chest, away from any gravitational fields, but which is being accelerated upwards in a way that creates the same downward pull experienced by the observer in the first chest. Einstein challenges us to find a way to tell the difference, observationally, between the two cases. We find that we cannot. Next, we *induce* from the equivalence of the phenomena experienced by the two imagined observers in the two opaque chests to the equivalence of *all* phenomena in such conditions. Once this is done, we can identify inertial frames with gravitational fields. The material principles Norton identifies (though not by that name, as he was writing before he published on the material theory) are that "the case is typical" and "the presence of the chest and observer are inessential to the equivalence" (Norton, 1991, 137). The formal properties of this inference do not determine its quality; the material principles do. If they are doubtful, the inference fails.

Another example is Newton's bucket and Mach's response. Both of these appear to be inferences to the best explanation. For Newton, the best explanation of the concavity of the water's surface is rotation with respect to absolute space. One interpretation of Mach's response is that the best explanation of the concavity of the water's surface is rotation with respect to the other masses in the universe. On a formal theory of induction, we are told to identify which is the better explanation. But how do we do that? Formal measures have not provided any workable account of what makes one explanation better than another. In practice, we don't appeal to formal accounts of explanatory quality, but consider the reasons given against the other explanations, and the reasons given in favour of the preferred explanation, as well as the plausibility of the background assumptions. And this is precisely what the material theory of induction recommends we do to decide between Newton and Mach.

The general idea, then, is that a combination of the argument view and the material theory of induction will take thought experiments that amplify our knowledge, identify them with inductive arguments, and provide an analysis of the justification of their conclusions not in terms of the formal aspects of the arguments, but the material aspects. I'll start by noting some exciting upshots of this combination of views.

First, the above-mentioned reconstructions of Einstein's elevator and Newton/Mach's bucket (while needing to be fleshed out) do seem prima facie plausible, and we can imagine how the same could be done for other inductive thought experiments. In other words, by combining

[5] Notice that it uses a combination of argument styles. A) is argued for by appeal to tradition, popularity, or intuition. B) through G) present a series of steps that begins with the practice of science as data, and infers to the identity of thought experiments and arguments via a combination of induction and deduction. H) is an inference to the best explanation.

these views, we gain a new and possibly powerful way of evaluating inductive thought experiments, via the identification and analysis of material principles.

Second, Norton seems to have presaged the material theory in his work on thought experiments. He argues in 1991 that there will always be some particulars that appear in the premises of inductive thought experiments that will not appear in the conclusion (Norton, 1991). This is because the set up of any inductive thought experiment is always a specific scenario, not all of whose features are relevant to the more general conclusion. The elimination of particulars to arrive at something general is accomplished by an "inductive step" (1991, 131). "This step might involve the assertion that the case involving the particulars is 'typical' or that the particulars are 'inessential,' so that the result derived holds in other cases as well" (131). This seems to be evidence that it was already clear to Norton in 1991 that material principles are needed to license the inductive steps found in thought experiments.

Third, a consequence of this combination of views is that we are never licensed to say how justified the conclusion of an inductive thought experiment is until we have examined the relevant material principles that justify it. And it is unlikely that we will ever be able to say how justified a thought experiment is in any "objective" or acontextual sense, because the uncertainty in the material principles relies on the uncertainty of other facts in the web of scientific knowledge, which is in a constant state of change. The most we can say about a scientific thought experiment is that *as far as we know* it is justified. And this reflects scientific practice. If you bring up a famous thought experiment to a scientist and ask them if it was a good one, they usually say "Well, at the time, given what was then known, yes [or no]. But now…" and then they praise or condemn the thought experiment depending on how much relevant change there has been to the state of knowledge since the thought experiment was introduced. (This development contradicts Norton's earlier claim that the conclusion of a thought experiment is justified acontextually, since the justifying "mark" is logical and therefore independent of context (Norton, 2004a, 1143). But perhaps Norton would now agree that inductive thought experiments cannot be evaluated without reference to their context).

Despite some harmonic opening notes, however, discord threatens. Recall that on the argument view, thought experiments and arguments are justified in virtue of their formal properties. Yet on the material theory of induction, arguments are justified in terms of their material (and *not* formal) properties. Let us consider the tension in more detail.

Where exactly does it stem from? First, there is Norton's oft-repeated claim that the argument view should hold for all thought experiments, including inductive ones. Second, there are readings of some of the seven theses mentioned above which contribute to the tension if "logic" and "argument" are understood formally, as originally intended. For example, the Identity thesis claims that all thought experiments are arguments. This need not be problematic, but, as we saw, "arguments" were intended to be formal entities governed by a logic, which is an evaluative system that "governs" inferences via a focus on formal properties. This cannot be held consistently with the material theory of induction, since a material logic of induction does not govern via a focus on formal properties. Second, the Elimination thesis claims that the epistemology of thought experiments can be reduced to the epistemology of arguments. This is problematic because the epistemology of arguments that Norton originally presented was the *formal logic* of arguments. If the material theory of induction is correct, no epistemological account of inductive inference that focuses only on formal features could ever succeed, thus, we cannot reduce the epistemology of thought experiments to a formal epistemology of arguments. Finally, the Reliability thesis states that if a thought experiment cannot be justified by a (formal) logic, there is no way to tell if it is justified or not. Again, this contrasts with the material theory, according to which there *are* ways to tell whether an inference is justified other than by reference to its formal properties: namely, by investigating the inference's material principles.

What about Norton's argument for his position: is this compatible with the material theory of induction? The main sources of support for the seven theses are premises A), B), C) and D) in the reconstruction given above. A) is just the Empiricist thesis, which coheres very nicely with the material theory of induction since that theory tries to explain the justification of inductive inferences not in terms of abstract logical forms, but facts, which (it is assumed), ultimately reduce to experience. But A) is an assumption, and therefore cannot be relied upon (or denied) without begging the question. D) is the Reliability thesis, which (as mentioned immediately above), clashes with the material theory of induction if it is read as originally formulated. The premises B) and C) are left, and I think they must be the main pieces of evidence for the argument view. They are both inductive inferences. In B), Norton notes that all thought experiments he knows of can be reconstructed as arguments. From this he induces to the logically stronger claim that, generally, *all* thought experiments will be reconstructible as arguments. In step C), he notes that all thought experiments he knows of have conclusions that are equally justified as those of their logical reconstructions. From this he induces to the logically stronger claim that *all* thought experiments have conclusions that will be equally justified as those of their logical reconstructions. These are both inductive inferences, and according to the material theory of induction, such inferences succeed or fail based on their material principles. So what are the relevant material principles in these cases?

Remember bismuth: inducing from the melting point of one pure sample of bismuth to the melting point of another pure sample is (relatively) safe because bismuth is *a special kind of thing* such that all its samples generally have the same melting point. In the cases we are interested in, we need some material principle about the *kind of thing* that thought experiments are, which will support the inductive inferences in B) and C).

For B), we need a reason to believe that the nature of thought experiments is such that all thought experiments will always be amenable to logical reconstruction. The material principle underlying this inference is that thought experiments all admit of the distinction between form and content, and their forms can be used to identify them. Since a logic (for Norton) is a system that evaluates inferences based on their formal properties, then, if all thought experiments are as described, logicians should be able to reconstruct all thought experiments. Now, there is a sense in which we may individuate arguments according to their formal properties, even on the material theory. For example, even Norton identifies some inductive arguments as inferences to the best explanation and others as arguments by analogy. However, in a more fundamental sense, there is still tension here, since Norton *denies* that we can (or should) separate the formal from the material elements of an inference. He writes, "if one adopts a material theory of induction, one no longer separates factual content from the rules of inductive inference" (2014, 672). In other words, this material principle (which might work for deductive thought experiments) contradicts Norton's new characterization of inductive logic, according to which the formal properties of inferences are irrelevant to their status as inductive arguments. But I will not dwell on this, since, as mentioned, there is a sense in which we can still use formal properties to identify arguments, even assuming the truth of the material theory of induction.

The case of C) is more interesting and more difficult. In C), we induce from the fact that the conclusions of many thought experiments enjoy exactly as much justification as the conclusions of their logical reconstructions do, to the claim that the conclusions of *all* thought experiments will be exactly as justified as the conclusions of their logical reconstructions. To justify this move, we need a material principle that grounds the conclusion of the inductive argument in some fact about the nature of thought experiments. The material principle assumed by the original argument view seems to be something like this: thought experiments (and arguments) are formal objects, and the only way to justify a formal object is via some formal logic. Thus, the conclusion of any thought experiment will always be exactly as justified as its

reconstruction, because they have the same source of justification: namely, their formal properties. But once we introduce the material theory of induction, we no longer use the formal properties of inferences to determine their justificatory status. Norton writes, "merely stating an inference schema does not automatically make it a good one. In familiar deductive cases, we discern that they are good because of the meaning of the connectives. We cannot do the same for inductive schemas" (forthcoming, prolog, 7). Thus, the formal properties of inductive thought experiments and arguments are irrelevant to their justificatory status, and we lose our reason for thinking that thought experiments will always be equally justified as their reconstructed arguments are, if we use the presumed material principle.

In sum, the inductive generalization in step C) of Norton's argument requires a material principle, and the material principle that Norton would have chosen is no longer available once we introduce the material theory of induction. And this blocks the passage to E), F), G) and H) in his argument for the argument view.

To eliminate the tension between the two views, something has to go. Perhaps the easiest thing to do is to break apart the argument view into a deductive and inductive account. Norton is clear that the (formal) argument view is meant to work for all thought experiments, whether deductive or inductive. But he could alter this, so that we have a formal argument view for deductive thought expermiments and a material argument view for inductive thought experiments.[6] For deductive thought experiments, the argument from A) to H) then remains as it was originally. But in the case of inductive thought experiments, we need a new material principle that suggests that the justificatory status of all thought experiments will be exactly the same as their reconstructed arguments. The old material principle was that all thought experiments admit of the distinction between form and content, and are justified in virtue of their formal properties. Perhaps the new material principle could be that all inductive thought experiments (and arguments) admit of the distinction between what warrants (the material principles) and what is warranted (the conclusion of the argument or thought experiment). Since both thought experiments and arguments admit of this distinction and are justified in the same way, we can infer that they will always be equally justified.[7]

However, there are worries here as well. First, on its own, the new material principle doesn't suggest that thought experiments will always be *as justified* as their reconstructions. All it says is that thought experiments admit of the distinction between what warrants and what is warranted, and their conclusion will be warranted by material principles. It seems perfectly possible that a thought experiment and its reconstruction could be warranted by different material principles. It also seems possible that two different inferences, warranted by the same material principles, might be unequally justified.

Second, it is possible that many things which aren't arguments will admit of the distinction between what warrants and what is warranted. For example, laboratory experiments admit of that distinction, yet they are not arguments. Or consider justification via perception: when I see raindrops through my window, I justifiably come to believe that it is raining. This event admits of the distinction between what warrants and what is warranted (the perception warrants the belief). Yet, no one would claim that I am arguing. Or consider emotional episodes. States of affairs warrant certain emotions (like righteous indignation), but surely not all emotional episodes are logical arguments.[8]

Third, whatever the material principle is, it must be consistent with the material theory of induction. On the material theory, Norton takes inference to be a *non-psychological* relation of logical support. He restricts "notions of inference and logic to relations of deductive and inductive support between propositions, independently of our beliefs and thought processes" (forthcoming ch. 1, 1–2). "An inference from proposition A to proposition B is a logical relation between the two propositions as sanctioned by some logic…This usage is to be contrasted with a psychologized notion of the term 'inference' that will not be employed here" (forthcoming, ch. 1, 2). Surely arguing requires inferring. But what is the mental act of inferring, on this account? It seems to be only *recognizing* that some logical relation holds. But thought experiments involve agency: when we perform a thought experiment, we do more than merely observe the presence of logical connections between facts. This gets even murkier when Norton claims that logical relations of support are themselves facts (forthcoming ch. 1, 22). If thought experiments are just facts transformed by other facts, where do humans fit in? If the agent is no more than a spectator, this doesn't match the phenomenology of thought experiments.[9]

Finally, even if Norton succeeded in defending this version of the argument view, it would be accepted by all parties to the debate about thought experiments. Its main claim, that the conclusion of a thought experiment is justified by virtue of its content and not its logical form, only contradicts Norton's original account and no one else's. The disagreement in the literature on thought experiments is not about whether the conclusions of thought experiments are justified by facts, but about how we mentally mobilize the facts in our possession to produce new knowledge from those facts.

Given these difficulties, we must make more drastic changes. One possibility is to modify the material theory to bring it in line with the argument view. Another is to hold the material theory constant and modify the argument view. I'll quickly consider the first before pursuing the second. In the end, I will suggest a new version of the argument view that avoids the tension.

5. Modifying the material theory of induction

One way to save the argument view would be to allow that induction is still (partially) formal. In this case, thought experiments can be (partially) justified in virtue of their logical properties, as the argument view claims. Perhaps Norton would be willing to make such a move; after all, inductions have local forms, or "mini-logics," which might contribute to a "dual view" of inductive justification, according to which both material principles *and* local logical schema contribute to the justification of an inductive argument.

But Norton makes several claims that seem to disqualify this as an option. For example,

> The material theory of induction arises when we assume that the truth of these background factual presumptions is all that is needed for the inductive inference to be warranted. One might imagine that this might not be so. The facts, we might suppose, play only a partial role in warranting the inductive inference. Might there still be a residual universal formal schema or inductive rule that contributes to

[6] There are hints that Norton's material theory of induction might account for deductive inference as well (see Brigandt 2010). I think this interpretation of the material theory is exciting. And aspects of the argument view even suggest it. For example, as an anonymous referee pointed out, since knowledge-producing deductive thought experiments must be sound, we have to justify their premises, and those justifications might be inductive. But on the solution I am now considering to the above-mentioned tension (which I think is Norton's preferred solution), we must rule out treating deduction in a material way. If we were to pursue that option, we would have to abandon the original argument view completely.

[7] Norton suggested this idea to me by personal communication.

[8] One could try to exclude perceptions and emotions by claiming that they are non-propositional, and arguments must have propositions as parts. This would be controversial, but even if successful, it would not help Norton because Norton does not want to limit arguments to manipulations of propositions (2004b, 53).

[9] This was also a problem for the original argument view. I thank an anonymous referee for pointing this out.

> the warrant? Such a schema or rule, however, would in turn be subject to the same analysis just given. (Forthcoming ch. 2, 10)

Norton appears to be claiming that facts provide all the warrant for inductive inferences, which implies that formal schema provide none of the warrant. However, it is possible that Norton only means that inductive inferences do not receive any justification from instantiating *universal* schema, leaving open the possibility that they receive some justification from instantiating *local* (non-universal) schema. But what exactly does a local schema contribute to the overall justification of an inductive inference? A complete theory requires an answer. We do not want an account that only explains when we were warranted in *using* local inductive schema without explaining how the schema themselves are justified. This would be like explaining when we are warranted in using violence to solve a problem, and then claiming this as an explanation of how violence solves problems.

However Norton seems to rule out this reading as well. For instance,

> These examples illustrate the general point: the factual assumption that ours is a hospitable world is the fact that, if true, warrants the inductive inference. It may not always be apparent that this fact warrants the inference. It may appear that the warrant is still provided by some sort of schema. The inference to a future snowy winter, we may think, is still warranted by the schema:
>
> All past A's have been B
> Therefore, the next A will be B.
>
> … This schema, if used at all, has a purely intermediate role. It does not have universal applicability. We can use it in the snowy winter case only because the requisite background facts authorize it, when we make the specific substitutions: 'winter' for A and 'snowy' for B. That is, there is a cascade of warrants that may pass through a schema. *The cascade terminates in facts that are the final warrant of the inference*." (Forthcoming ch. 2, 9; my emphasis)

The *final warrant of the inference* is different from the final warrant for *using* a schema in an inference. This seems to imply the purely material reading of the account.

In sum, if local schema partially justify inductive arguments then Norton's account does not explain this extra source of justification, and it is incomplete as an account of induction. If local schema provide no additional justification to inductive inferences (and I think this is what Norton intends), then we cannot rescue the argument view by introducing partially formal justification. We must try something else.

6. Modifying the argument view

I think we can modify the argument view such that it still serves as *part* of a complete epistemological account of thought experiments. First, as mentioned above, Norton can keep the original argument view for deductive thought experiments. When it comes to inductive thought experiments, he can save a version of the argument view, which retains the Empiricist thesis in full, claims weaker (but more plausible) versions of the Identity, Reconstruction, Reliability and Epistemic theses, and rejects the Elimination and Empirical Psychological theses. Thus it would retain (plausible versions of) five out of seven of Norton's theses, while also satisfying Norton's more fundamental commitments. This and the next section will offer such a view for consideration.

Norton claims that thought experiments are arguments, and the cognitive performance of a thought experiment is the cognitive execution of an argument. But it isn't clear what that means if "argument" is defined by Norton in terms of non-psychological relations of warrant between premises and conclusion. Rather, it's more plausible to analyze thought experiments in the other direction: thought experiments are complex cognitive actions performed by agents, parts of which can be reconstructed as explicit arguments. Indeed, this is precisely what Norton elsewhere suggests we do when it comes to the epistemology of scientific *analogies*. With respect to analogy, Norton finds

> a curious divergence between the philosophical literature and the scientific literature. The philosophical literature categorizes analogy as a form of inference to be analyzed using some version of the formal methods of logical theory. The scientific literature approaches analogies as factual matters to be explored empirically; or at least it does so for the important analogies that figure centrally in the sciences. (Forthcoming, ch. 4, 2)

Norton claims that we should stop trying to create general formal accounts of analogies, and instead evaluate each one as it is used in science, taking into account the facts of the matter relevant for each analogy. The same considerations should hold for thought experiments. In this case, we begin by pointing out

> a curious divergence between Norton's argument view and the scientific literature. The argument view categorizes thought experiments as a form of inference to be analyzed using some version of the formal methods of logical theory. The scientific literature approaches thought experiments as factual matters to be explored empirically.

In other words, we should treat thought experiments in the way that Norton encourages us to treat analogies. We should take them as they are, individually, without trying to create a unified formal logic that explains how they work in general. Treating thought experiments as we treat analogies (and other kinds of inference like inference to the best explanation – see Norton forthcoming chs. 8–9) brings thought experiments under the domain of the material theory of induction, and it also coheres nicely with Norton's commitment to take scientific practice seriously.

The resulting account, I will argue below, should be this: the *conclusions* of inductive thought experiments, insofar as they produce new knowledge, are justified according to the material theory. That is, according to their material (and not formal) properties. However, inductive thought experiments are not wholly identical to inductive arguments, and the epistemology of inductive thought experiments is not exhausted by the epistemology of inductive arguments.

As mentioned above, everyone in the literature on thought experiments would accept this. For example, Brown would allow that what warrants us in believing the conclusion of a given thought experiment will be certain background facts that obtain. The difference is that the facts, for Brown, can include facts about relations between universals "perceived" through a priori intuition, while for Norton they must be reducible to sensory experience. Still, everyone in the debate would agree with Norton that it is facts, not logical forms, that provide the ultimate justification for the conclusion of a thought experiment.

Yet, a material argument view can still claim something unique. Namely, it will usually be the case that the inductive step of all inductive thought experiments *should* be reconstructed as an argument. That reconstruction is required for identifying and justifying the thought experiment's material principles.[10] This weaker version of the argument view would not insist on the Eliminability thesis, because, as I'll argue below, there is more to the epistemology of thought experiments than the inductive extension of their conclusions. It would also give up the Empirical Psychological thesis as well. That thesis was motivated at least partially by the Eliminability thesis: if we can eliminate thought

[10] Some authors have already encouraged Norton to soften the argument view so that it only *recommends* reconstruction. For example, Richard Arthur writes that "the reformulation of thought experiments as arguments is a vital part of the scientific process" (1999, 228), without endorsing the elimination of thought experiments via their identification with arguments. See also Lennox (1991). However, what is being suggested now is not merely a softening of the argument view, but a way of saving and defending a material theory of inductive thought experiments.

experiments from our epistemology of science, it must be because there is nothing more to thought experiments than arguments. Therefore, the psychological performance of a thought experiment must be nothing more than the execution of a logical argument. But since the Eliminability thesis is no longer supported by the material theory, we lose the support it provided for the Empirical Psychological thesis. Also, if Norton is committed to taking local elements of scientific inference seriously, including the context and content, then he should be interested in the epistemologically relevant features that thought experiments have outside of their formal reconstructions.

In sum, here is how Norton could frame a new argument view. Given a thought experiment, we first determine whether it is deductive or inductive (assuming that this is always possible in practice). Deductive thought experiments are treated according to the original argument view: they can be reconstructed into arguments whose formal properties determine their validity. Inductive thought experiments, however, must be treated differently. When faced with an inductive thought experiment we take each on its own terms and no longer claim that its epistemology should be pursued by formal reconstruction. And, for Norton, the most interesting aspect of the inductive thought experiment, whatever other aspects it might have, is the part of the thought experiment in which previous knowledge is extended, via induction, to new knowledge. This will be justified via a material principle. But it might not always be easy to see which material principle is required, and so reconstructing this aspect of the thought experiment might be helpful.[11]

Now, when we speak of reconstructing the thought experiment as an argument, we are only speaking about reconstructing a *part* of the thought experiment as an argument: namely, the part in which the inductive step occurs. There may be many other epistemologically relevant parts of a thought experiment, but the new argument view does not concern itself with those other aspects (more on this in a moment). It allows that they exist, and does not deny epistemological accounts of them, because the new argument view does not claim exhaustiveness with respect to the epistemology of thought experiments.

Finally, worries about the notion of "argument" involved in the new argument view disappear entirely. On the original view, Norton needed a notion of argument that was broad enough to include diagrammatic thought experiments and those with tacit premises and inferences. This threatened to make that view trivial, because if arguments are anything categorizable by a final or future logic, then every epistemologically interesting inference becomes an argument by definition. However, on the new proposal, the notion of argument loses all importance: all we need is the ability to (re)present the inductive step of a thought experiment such that the relevant material principle(s) can be identified. Reconstructing this part of a thought experiment may require teasing apart discrete inferences and identifying relevant bits of information hidden in premises, but it is not at all controversial that this can be done, indeed, we do it all the time. Because we are no longer claiming that thought experiments are (and are only) arguments, we can even require that the bits of information be presented in propositional form. Our reconstructions can be propositional without requiring that thought experiments themselves be propositional.

This version of the argument view avoids the tension mentioned above because it provides different accounts for inductive and deductive thought experiments, does not claim that the epistemology of arguments exhausts the epistemology of thought experiments, and gives up premises B) and C) of the argument for the argument view, as well as the Elimination and Empirical Psychological theses. But it is still an argument view, because it demands reconstruction of the inductive step of all inductive thought experiments into arguments, and evaluates those steps using the material theory of induction.

[11] This is not a trivial position; some philosophers would disagree with it. One such philosopher is Paul Feyerabend, who argued that "reconstruction" and "clarification" are evils to be avoided (Stuart forthcoming a).

We might wonder how well such a view stands up to scrutiny. For example, is it really an argument view in the spirit of Norton's original, given that it does not reduce thought experiments entirely to arguments? And given that it would satisfy most of Norton's critics, can we still think of it as separate contender for the "right" epistemology of thought experiments? The point of this paper was to see how well, if at all, the material theory of induction and the argument view of thought experiments could be made to cohere. I hope I have shown that they do not cohere perfectly, but that there are concessions that can be made which preserve all of the material theory, and much of the spirit of the argument view. I want to close by clarifying exactly which epistemological features of thought experiments the new argument view could explain, and which it could not.

To this end, I want to distinguish between four epistemologically relevant aspects of thought experiments. The reason is that, as we will see, neither the original nor the new argument view can capture all of them. But the new argument view can be very useful in telling *part* of the full story. In other words, there is an aspect of the epistemology of thought experiments that this view handles very well. Insofar as the material theory handles this aspect better than the competitors, it should still be considered a substantial view in the epistemology of thought experiments.

The different aspects are: the creation, cognitive performance, performer, and interpretation of a thought experiment. These are meant as different aspects from an epistemological, not an ontological, point of view. Once we separate these different targets of epistemological inquiry, we see that a material theory of thought experiments can and should only attempt to account for the last of these.[12]

7. A pluralist epistemology of thought experiments

7.1. The argument view and the creation of thought experiments

One aspect of thought experiments that the argument view cannot address is how we create an epistemically efficacious thought experiment. The "logic of discovery" is not usually discussed in the literature on thought experiments, but it should be. Neither the new nor the original argument view can provide an epistemological account of how to create an effective new thought experiment. We might think that to construct a good thought experiment, we simply construct a good argument. But thought experiments can pack a rhetorical punch that their reconstructed arguments cannot (Norton, 1991). The argument view as an epistemological position does not have the resources to explain how we should, in general, take an argument and make it into a thought experiment. Because it is retrospective and anti-psychological, the argument view cannot serve as a normative guide for the construction of effective new thought experiments. Concluding his chapter on thought experiments, Ernst Mach wrote, "It is often said that enquiry cannot be taught. In a sense this is correct: the schemata of formal and even inductive logic cannot help much, for intellectual situations never quite repeat themselves" (1905, 146).[13]

How do we provide a prescriptive, forward-looking way to construct epistemically efficacious thought experiments? I think the best way to do most new things will always be the same: by experiment. We generate a best guess for how to do something, try it, learn from our mistakes, try something else, etc., until we succeed. Equally with thought

[12] There are other aspects that also deserve discussion, but which I don't have space to address here. One is dialectical context. For an exploration of this, see Goffi & Roux (2018). For other divisions of aspects and phases of thought experiments, see, e.g., Mach (1905/1976), Gooding (1990), Nersessian (1993), Reiner and Gilbert (2000), Buzzoni (2008), El Skaf and Imbert (2013), Chandrasekharan, S. et al. (2013), and Lenhard (2018).

[13] This is not to say that the material theory is in *conflict* with the epistemology of creation or discovery. Brigandt (2010) shows that it is not.

experiments: if we want to construct a new thought experiment, we must experiment. Experimenting is an activity with its own epistemology that already provides quite a bit of guidance. For example, we can ask whether and how our mental "instruments" are isolated from disturbances, whether we have accounted for experimental bias and sources of error, whether our instruments are well-calibrated and whether we have a theory of them, etc. (Franklin, 1986; Galison, 1987; Gooding, 1990; Hacking, 1983; Weber, 2005).

In sum, one epistemologically significant aspect of thought experiments is their creation. How thought experiments have been and should be created such that they increase our knowledge or understanding of the world is a normative epistemological question that the argument view does not answer. Norton could reply that the argument view only concerns the context of justification and not the context of discovery. But elsewhere he insists that even in the context of discovery, thought experiments are (and function epistemically as) arguments (2004b, 50). Perhaps Norton could reply that he never intended his view to cover the creation of new thought experiments. But Norton claims that the argument view "does supply a complete epistemology in the sense that all there is to learn about a thought experiment's epistemic power can be recovered from considering it as an argument" (2004b, 55). He could argue that questions about how best to create a new thought experiment are not properly epistemological, but we would need a reason to believe that. It certainly does seem epistemologically interesting to ask how scientists develop successful thought experiments, and how they might improve their ways of doing this.

7.2. The argument view and the performance of thought experiments

A second aspect of thought experiments that is not covered by the argument view is their cognitive performance. To get at what I mean by this, consider a scientist who spends some time crafting a good thought experiment. When she finally succeeds, she has also performed the thought experiment. We might think that the performance is therefore the same as the creation of the thought experiment.[14] But even if these always coincide, there are different epistemological questions we can ask about the creation and performance of a thought experiment. For example, when it comes to creation, we can ask whether a scientist went about trying to create a new thought experiment efficiently. For example, perhaps a thought experiment was built upon a needlessly complicated analogy when a simpler one would do. Meanwhile, holding the thought experiment fixed, we can ask which of several different people performed the thought experiment best, and why. That is, supposing that the thought experiment was intended to clarify the empirical content of a new concept, which performer imagined things such that their understanding of the new concept was increased the most? Since these kinds of question can be separated, I claim that the cognitive performance of a thought experiment can be separated (for epistemological purposes) from the creation of the same thought experiment.

To focus on the performance of a thought experiment, we must also be able to distinguish between the performance and the interpretation of a thought experiment. This distinction has already been made in the case of laboratory experiments (Radder, 1996). Ian Hacking calls it a "truism" that "experimenting is not stating or reporting but doing" (1991, 133). This distinction seems to hold equally well for thought experiments, which share some structural similarities with laboratory experiments (Brown, 2007, 158; De Mey, 2003; Häggqvist, 1996). In both cases, we say that a scientist performs some concrete action which is then later described in sentences in the context of meetings, publication preprints, press releases, articles and textbooks. The epistemology of experimental acts and events is the epistemology of a set of actions performed on a particular system, not the epistemology of a set of claims and their relation to experience.

The epistemic quality of an experimenter's actions depends on the nature of those actions and their context. Thus, there were many performances of the Michelson-Morley experiment, some carried out underground, some on hilltops, and some in hot air balloons (Swenson, 1970), and some of these were more accurate than others. Asking questions about what was done and why, and what worked, is a kind of epistemological inquiry that is important for justifying the interpreted result of that experiment and its import for ether theory and Special Relativity.

To evaluate the performance of a *thought* experiment, we should begin by asking what the epistemic aim of the thought experiment is. If it is knowledge, we might focus on the accuracy of mental representations, or the quality of rational intuition, since these are relevant to knowledge. If the aim is understanding, we have to be clear about what kind of understanding we have in mind: semantic, objectual, explanatory, model-based, practical, etc. (see, e.g., Baumberger, Beisbart, & Brun, 2016 for distinctions between kinds of understanding; see Stuart, 2018 for this discussion applied to thought experiments). In each case, the way forward will be different. For semantic understanding, accurate representations might be less important than powerful metaphors or clear examples.

Once the epistemic aim of the thought experiment is clear, we should ask how humans produce that aim given the set up of the thought experiment. Answers to this question will depend on the epistemological framework adopted, and the sort of thought experiment being analyzed. For those who support the view that thought experiments are manipulations of mental models, we might justify the reliability of the way we mentally evolve systems in realistic ways as a result of the evolution of the human mind, combined with conscious stipulations about the accuracy of representations and dynamics of the mental model as well as the skills or know-how of the agent (Nersessian, 1992, 2007, 2018; Miščević, 1992, 2007). An experimentalist explanation would focus on the isolation of the imaginary system, accounting for sources of bias and error, experimenter know-how, and so on (Gooding, 1992a, 1992b; Sorensen, 1992; Stuart, 2016). Epistemological particularists would insist that each performance must be evaluated on its own for reliability, instead of making reference to general facts about, e.g., human cognitive evolution.

For the epistemology of performance, the argument view is again inappropriate since Norton demurs from treating inductive inferences as psychological entities. Instead, inductive inferences are treated as objective relations of epistemic support between premises and conclusion. However, just as baking a cake is not awareness of the steps in a cookbook, performing a thought experiment is not awareness of relations of logical support. Norton might deny that the cognitive performance of a thought experiment is epistemologically relevant. However, this would be to uphold the distinction between the contexts of justification and discovery (or justification and psychology), which philosophers of science have rejected for good reasons that we need not rehearse here.

So what is the epistemological significance of the cognitive performance of a thought experiment? One possibility is that the cognitive performance justifies a premise that supports the conclusion of a thought experiment. It produces a fact, phenomenon or experience, which an inductive argument then extends. Let me explain using Norton's example of "Einstein's best known thought experiment" (Einstein 1905, 37). Norton reconstructs that thought experiment into an argument as follows.

1. In the case of electromagnetic induction, through its positing of an absolute state of rest, classical electrodynamics distinguishes states of affairs which are not observationally distinct.

Therefore in this case, classical electrodynamics violates.

[14] Indeed, this is claimed more or less explicitly by Buzzoni (2008) and Arcangeli (2018).

2. Verifiability heuristic for theory construction (version 2) ["States of affairs which are not observationally distinct should not be distinguished by the theory"]: therefore
3. Absolute velocities should be eliminated from the theoretical account of electromagnetic induction.
4. Inductive step: This example is typical since (a) there are other examples of this type and (b) there is a history of unsuccessful attempts to detect this state of rest by optical experiments. Therefore
5. Absolute velocities should be eliminated from electrodynamics (Norton, 1991, 136).

This is how Norton reconstructs Einstein's "best known" thought experiment as an argument. But I assume that many readers will not be able to guess which thought experiment is being presented here. The thought experiment is the thought experiment of the magnet and conductor. Here is an alternative presentation of the same thought experiment. Maxwell's electrodynamics postulates an absolute velocity for light in a vacuum, which Maxwell takes to imply an absolute state of rest. This state was identified with properties of the aether. The thought experiment then proceeds as follows: you take a magnet at absolute rest and move a conductor through its magnetic field. This generates a measurable current. Now, you keep the conductor at absolute rest and move the magnet. You have the same relative motion, and the same measurable effects, so we should expect the same explanation for the two phenomena. But Maxwell's theory provides a completely different kind of explanation for the latter measurement, because of its stipulation of absolute rest. "For if the magnet is in motion and the conductor at rest, there arises in the neighborhood of the magnet an electric field with a certain definite energy, producing a current at the places where parts of the conductor are situated. But if the magnet is stationary and the conductor in motion, no electric field arises in the neighborhood of the magnet" (Einstein, 1905). Einstein concludes that the stipulation of absolute rest is unwarranted: it's better to do without it (Norton, 1991).

What is important to note is that the cognitive performance of the thought experiment and the reconstruction do not share much in common, crucially, not even the same epistemic goal. The cognitive performance of the magnet-conductor thought experiment ends with the conclusion that the same relative motions of a magnet and conductor will produce the same measurable current. This justifies the first premise of the reconstructed argument. The rest of the argument involves verifiability heuristics for theory construction and a history of unsuccessful attempts to detect the aether and other things that do not seem to be part of the cognitive performance of the thought experiment at all. These are two different projects: one provides evidence for a fact, and the other aims to extend this fact inductively. The justification of each is also different. One is the justification of a cognitive output in terms of the reliability or perhaps the "presentational phenomenology" (Chudnoff, 2012) of certain cognitive processes, the other is the justification of an inductive extension of that cognitive output via material principles.

7.3. The argument view and thought experimenting agents

A third epistemological approach focuses on the properties of agents who perform thought experiments. This builds on the distinction between virtue epistemology and traditional or process-based approaches to epistemology (seminal texts include Kvanvig, 1992; Sosa, 1991 and Zagzebski, 1996). According to virtue epistemology, we should be good thought experimenting agents, even if our epistemic vices escape detection and do not affect the outcome of a particular thought experiment.

With respect to the epistemic virtues of agents, the argument view and the material theory of induction are again of no help since they are silent about the qualities that make an agent epistemically responsible or reliable as a reasoner. It might be that a fact about a reasoner is what enables her to extend her knowledge inductively, but the material theory concerns the warrant of inductive claims, not agents as a whole.

To pursue this epistemological angle, we investigate the (natural or acquired) epistemic virtues, motivations or skills needed for successful thought experimenting. These might include good (modal) intuition, imaginativeness, memory, open-mindedness, intellectual courage, creativity, honesty, and integrity. Not many philosophers focus on these virtues of thought experimenters, but some do. For example, Buzzoni (2008) and Stuart (2017, 2019) examine the epistemology of the faculty of imagination in relation to thought experiments. Specifically, they ask how the imagination, a faculty utilized during all thought experiments, can be investigated such that those with strong and weak imaginations make better or worse thought experimenters. While exercises of imagination are important for the epistemology of the creation and cognitive performance of a thought experiment, the *faculty* of imagination is a feature of the agent, and thus falls under virtue epistemology. A virtue epistemology of thought experiments would be a welcome addition to the literature, but it is unlikely to come from the argument view or the material theory of induction.

7.4. The argument view and the interpretation of thought experiments

When we perform an experiment, we typically do so to answer a question. During the experiment, something "happens," which we then attempt to interpret as an answer to the original question. Equally in a thought experiment, we set up a scenario, let it unfold, and then we have to find a way to make the outcome of that unfolding bear on our theoretical concerns. In the magnet-conductor case, the outcome is evidence for the claim that it is *relative* motion that is important for explaining the current in the conductor and that the explanations in the two cases should be the same. To marshal this against the postulation of an electromagnetic ether requires extra work. That extra work may be conceived of as part of the thought experiment, though it is analytically separable. Just as there were different ways of performing the Michelson-Morley experiment, there were different interpretations of what that experiment meant for ether theory and Special Relativity, and some interpretations were better than others (e.g., Hacking, 1991, 133).

Interpretations of a thought experiment will be evaluated using the material theory of induction when the thought experiment requires inductive extension to provide evidence for a theoretical claim. In these cases, the interpretation will involve an inductive step, which Norton can claim should be identified and clarified through reconstruction, and justified by appeal to material principles. We can show that the extension does or does not follow by determining which material principles are required by the inference, and how much epistemic risk is involved in presuming those principles. This is where the new argument view shines. This is especially clear when it comes to what Norton has called "thought experiment/anti-thought experiment pairs" (2004a, 1140–1142; 2004b, 45–49) such as Newton's and Mach's bucket. What we must do, in such cases, is examine the operative material principles to see which is the more justified.[15]

8. Happily ever after?

On a pluralist epistemology of thought experiments, we can examine a thought experiment by considering how it was created. We can explain how performances of the thought experiment reliably lead to its conclusion, perhaps by looking at our own performances or the

[15] Other (unitary) accounts have trouble with such cases. If Newton and Mach both proceed by rational intuition, it isn't clear whose intuition is to be preferred, or why. Likewise, if Newton and Mach both proceed by manipulating mental models, we can ask whose model is a more accurate representation of the target system. But sometimes, the best thought experiment is not the one that represents the world most accurately. For example, if Galileo's falling bodies thought experiment was more accurate to the way things are in reality, the conclusion would not follow (Stuart forthcoming b).

performances of others. Different people have different cognitive abilities and will imagine the scenarios in slightly different ways, so the outcome of these cognitive exercises will depend on the individual. We can also examine the qualities of the agents that are relevant to their standing as good thought experimenters, for example, their receptivity to new evidence and ability to judge when and how much imagination to deploy. Finally, to explain how we properly or improperly interpret the imagined objects and events as bearing upon a theoretical claim, we may choose to employ the material theory (or deductive logic, or something else). With respect to this last aspect of the epistemology of thought experiments, Norton might be right that focusing on the *psychological* aspect of inferences veils epistemological considerations in unnecessary confusion. We don't want to know whether this or that person interprets the objects and events imagined in a thought experiment in a certain way; we simply want to know whether the inductive extension is justified.

Splitting up the epistemology of thought experiments makes their epistemology more complicated, but that's how it should be, since they are complicated things. Also, it recovers several important features of Norton's original argument view. Thought experiments should be reconstructed as arguments in the sense that the theoretical interpretation of a thought experiment should be so reconstructed. The creation and performance of thought experiments, as well as the relevant character traits of thought experiments, however, are also important for a complete epistemology of thought experiments, and these cannot or need not be reconstructed as logical arguments.

Also, Norton can still claim that inductive thought experiments amplify knowledge inductively, and that the conclusions of inductive arguments are justified in virtue of material principles.

Finally, it's important to note that a pluralist epistemology of thought experiments would not necessarily introduce anything that contradicts Norton's empiricism. For example, to investigate the cognitive performance of a thought experiment, what is required might only be empirical research into human cognition (Brendel, 2018, 291; Gendler 2004, 1161; Nersessian, 1992), and not the introduction of a faculty of rational intuition.

In sum, the suggestion is to retain intuitive aspects of the argument view for the epistemology of our theoretical interpretations of thought experiments, weaken the Identity, Reconstruction, Reliability and Epistemic theses, and drop the Elimination and Empirical Psychological theses.

9. Conclusion

There is a tension between Norton's material theory of induction and his argument view of thought experiments. In considering how to avoid this tension, minimal modifications were attempted, but found unsatisfactory. I suggested further modifications meant to accord with Norton's empiricism and the attention he pays to scientific practice. My suggestion is to allow that thought experiments have several epistemologically relevant aspects, and that a material argument view does not account for all of them. On a pluralist interpretation, a modified version of the argument view that is consistent with the material theory of induction tells only one part of the story, but an important one.

Acknowledgements

For comments on earlier drafts I would like to thank Margherita Arcangeli, Anouk Barberousse, Elke Brendel, Jim Brown, Rawad El Skaff, Yiftach Fehige, Pierre Saint-Germier, Valeria Giardino, Michal Hladky, Alice Murphy, John Norton, Sophie Roux, Guillaume Schlaepfer, Margot Strohminger, Marcel Weber, and Harald Wiltsche. I would also like to thank three anonymous referees, Elay Shech and Wendy Parker for their work on this special issue, and the Swiss National Science Foundation for funding (grant number PZ00P1_179986). Most of all I want to thank John Norton, who has been a philosophical and personal role model. Norton presents and develops historical and philosophical views that are so clear, intuitive, and well-grounded in scientific practice, that it's hard to imagine how they could be mistaken. All this makes it so rewarding to come back to his work over and over again.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.shpsa.2020.03.005.

References

Arcangeli, M. (2018). The hidden links between real, thought and numerical experiments. *Croatian Journal of Philosophy, 52*, 3–22.
Arthur, R. (1999). On thought experiments as A priori science. *International Studies in the Philosophy of Science, 13*, 215–229.
Baumberger, C., Beisbart, C., & Brun, G. (2016). What is understanding? An overview of recent debates in epistemology and philosophy of science. In S. Grimm, et al. (Eds.), *Explaining understanding: New Essays in Epistemology and the Philosophy of science*. London: Routledge.
Bishop, M. (1998). An epistemological role for thought experiments. *Poznati Studies in the Philosophy of the Sciences and the Humanities, 63*, 19–33.
Bishop, M. (1999). Why thought experiments are not arguments. *Philosophy of Science, 66*, 534–541.
Brendel, E. (2018). The argument view: Are thought experiments mere picturesque arguments? In M. T. Stuart, et al. (Eds.), *The Routledge companion to thought experiments* (pp. 281–292). London: Routledge.
Brigandt, I. (2010). Scientific reasoning is material inference: Combining confirmation, discovery, and explanation. *International Studies in the Philosophy of Science, 24*, 31–43.
Brown, J. R. (1986). Thought experiments since the scientific revolution. *International Studies in the Philosophy of Science, 1*, 1–15.
Brown, J. R. (1991/2011). *The laboratory of the mind: Thought experiments in the natural sciences*. London: Routledge.
Brown, J. R. (1992). Why empiricism won't work. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Vol. 1992, Volume Two: Symposia and Invited Papers*, 271–279.
Brown, J. R. (2004). Why thought experiments do transcend empiricism. In C. Hitchcock (Ed.), *Contemporary debates in the philosophy of science* (pp. 23–43). Malden: Blackwell.
Brown, J. R. (2007). *Counter thought experiments* (Vol. 61, pp. 155–177). Royal Institute of Philosophy Supplement.
Buzzoni, M. (2008). *Thought experiment in the natural sciences*. Würzburg: Königshausen & Neumann.
Chudnoff, E. (2012). Presentational phenomenology. In Miguens, & Preyer (Eds.), *Consciousness and subjectivity*. Ontos Verlag.
De Mey, T. (2003). The dual nature view of thought experiments. *Philosophica, 72*, 61–78.
Einstein, A. (1905). On the electrodynamics of moving bodies. In A. Einstein, et al. (Eds.), *(1952) The principle of relativity*. Mineola: Dover.
Franklin, A. (1986). *The neglect of experiment*. Cambridge: Cambridge University Press.
Galison, P. (1987). *How experiments end*. Chicago: University of Chicago Press.
Gendler, T. S. (1998). Galileo and the indispensability of scientific thought experiment. *The British Journal for the Philosophy of Science, 49*, 397–424.
Gendler, T. S. (2000). *Thought experiment: On the powers and limits of imaginary cases*. New York: Routledge.
Gendler, T. S. (2004). Thought experiments rethought—and reperceived. *Philosophy of Science, 71*, 1152–1163.
Goffi, J.-Y., & Roux, S. (2018). A dialectical account of thought experiments. In M. T. Stuart, et al. (Eds.), *The routledge companion to thought experiments*. London: Routledge.
Gooding, D. (1990). *Experiment and the making of meaning*. Dordrecht: Kluwer.
Gooding, D. (1992a). The procedural turn; or, why do thought experiments work? In R. N. Giere (Ed.), *Cognitive models of science* (pp. 45–76). Minneapolis: University of Minnesota Press.
Gooding, D. (1992b). "What is experimental about thought experiments?" PSA. *Proceedings of the Biennial Meeting of the Philosophy of Science Association, 2*, 280–290.
Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.
Hacking, I. (1991). Speculation, calculation and the creation of phenomena. In G. Muñevar (Ed.), *Beyond reason* (pp. 131–157). Dordrecht: Kluwer.
Häggqvist, S. (1996). *Thought experiments in philosophy*. Stockholm: Almqvist and Wiksell International.
Häggqvist, S. (2009). A model for thought experiments. *Canadian Journal of Philosophy, 39*, 55–76.
Kuhn, T. S. (1977). A function for thought experiments. In *The essential tension* (pp. 240–265). Chicago: University of Chicago Press.
Kvanvig, J. (1992). *The intellectual virtues and the life of the mind*. Savage, MD: Rowman and Littlefield.
Lenhard, J. (2018). Thought experiments and simulation experiments: Exploring hypothetical worlds. In M. T. Stuart, et al. (Eds.), *The Routledge companion to thought experiments* (pp. 484–497). London: Routledge.

Lennox, J. G. (1991). Darwinian thought experiments: A function for just-so stories. In T. Horowitz, & G. Massey (Eds.), *Thought experiments in science and philosophy*. Lanham: Rowman and Littlefield.
Mach, E. (1905/1976). *Knowledge and error: Sketches on the psychology of enquiry*. Dordrecht: D. Reidel. T. J. McCormack (trans).
Miščević, N. (1992). Mental models and thought experiments. *International Studies in the Philosophy of Science, 6*, 215–226.
Miščević, N. (2007). Modelling intuitions and thought experiments. *Croatian Journal of Philosophy, 7*, 181–214.
Nersessian, N. J. (1992). How do scientists think? Capturing the dynamics of conceptual change in science. In R. N. Giere (Ed.), *Cognitive models of science*. Minneapolis: University of Minnesota Press.
Nersessian, N. J. (1993). In the theoretician's laboratory: Thought experimenting as mental modeling. *Proceedings of the Philosophy of Science Association, 2*, 291–301.
Nersessian, N. J. (2007). Thought experiments as mental modelling: Empiricism without logic. *Croatian Journal of Philosophy, 7*, 125–161.
Nersessian, N. J. (2018). Cognitive science, mental modeling, and thought experiments. In M. T. Stuart, et al. (Eds.), *The Routledge companion to thought experiments* (pp. 309–326). London: Routledge.
Norton, J. D. (1991). "Thought experiments in Einstein's work.". In T. Horowitz, & G. Massey (Eds.), *Thought experiments in science and philosophy*. Lanham: Rowman and Littlefield.
Norton, J. D. (1996). Are thought experiments just what you thought? *Canadian Journal of Philosophy, 26*, 333–366.
Norton, J. D. (2004a). On thought experiments: Is there more to the argument? *Philosophy of Science, 71*, 1139–1151.
Norton, J. D. (2004b). Why thought experiments do not transcend empiricism. In C. Hitchcock (Ed.), *Contemporary debates in the philosophy of science*. Malden: Blackwell.
Norton, J. D. (2005a). A conjecture on Einstein, the independent reality of spacetime coordinate systems and the disaster of 1913. In A. J. Kox, & J. Einsenstaedt (Eds.), *The universe of general relativity. Einstein studies* (Vol. 11, pp. 67–102). Boston: Birkhaeuser.
Norton, J. D. (2005b). A little survey of induction. In P. Achinstein (Ed.), *Scientific evidence: Philosophical theories and applications* (pp. 9–34). Johns Hopkins University Press.
Norton, J. D. (2010a). How hume and Mach helped Einstein find special relativity. In M. Dickson, & M. Domski (Eds.), *Discourse on a new method: Reinvigorating the Marriage of History and Philosophy of science* (pp. 359–386). Chicago and La Salle, IL: Open Court.
Norton, J. D. (2010b). There are no universal rules for induction. *Philosophy of Science, 77*, 765–777.
Norton, J. D. (2012). Paradoxes of sailing. In P. A. Goold (Ed.), *Sailing: Philosophy for everyone* (pp. 148–163). Wiley-Blackwell.
Norton, J. D. (2013). Chasing the light: Einstein's most famous thought experiment. In M. Frappier, et al. (Eds.), *Thought experiments in philosophy, science and the arts*. London: Routledge.
Norton, J. D. (2014). A material dissolution of the problem of induction. *Synthese, 191*, 671–690.
Norton, J. D. (2018). The worst thought experiment. In M. T. Stuart, et al. (Eds.), *The Routledge companion to thought experiments* (pp. 454–468). London: Routledge.
Norton, J.D. (Forthcoming). The material Theory of induction. Draft of June 26, 2018.
Radder, H. (1996). *In and about the world: Philosophical studies of science and technology*. Albany: State University of New York Press.
Sorensen, R. (1992). *Thought experiments*. Oxford: Oxford University Press.
Sosa, E. (1991). *Knowledge in perspective*. Cambridge: Cambridge University Press.
Stuart, M.T. (Forthcoming a). Telling stories in science: Feyerabend and thought experiments. HOPOS: The Journal of the International Society for theHistory of Philosophy of Science, 10.
Stuart, M.T. (Forthcoming b). The productive anarchy of scientific imagination. Philosophy of Science.
Stuart, M. T. (2016). Norton and the logic of thought experiments. *Axiomathes, 26*, 451–466.
Stuart, M. T. (2017). Imagination: A sine qua non of science. *Croatian Journal of Philosophy XVII, 49*, 9–32.
Stuart, M. T. (2018). How thought experiments increase understanding. In M. T. Stuart, et al. (Eds.), *The Routledge companion to thought experiments* (pp. 526–544). London: Routledge.
Stuart, M. T. (2019). Towards a dual process epistemology of imagination. *Synthese*. https://doi.org/10.1007/s11229-019-02116-w.
Swenson, L. S. Jr. (1970). The Michelson-Morley-Miller experiments before and after 1905. *Journal for the History of Astronomy, 1*, 56–78.
Weber, M. (2005). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.
Zagzebski, L. (1996). *Virtues of the mind: An inquiry into the nature of virtue and the ethical foundations of knowledge*. Cambridge: Cambridge University Press.