

Thought Experiments in Current Metaphilosophical Debates

Daniel Cohnitz & Sören Häggqvist

I. Introduction

Although thought experiments were first discovered as a *sui generis* methodological tool by philosophers of science (most prominently by Ernst Mach), the tool can also be found--- even more frequently---in contemporary philosophy. Thought experiments in philosophy and science have a lot in common (cf. the discussion in Sorensen 1992, Häggqvist 1996, and Cohnitz 2006). However, in this chapter we will concentrate on thought experiments in philosophy only. Their use has been the center of attention of metaphilosophical discussion in the past decade, and this chapter will provide an overview of the results this discussion has achieved and point out which issues are still open.

In metaphilosophy, thought experimentation is also often referred to as “the method of cases”; and there sometimes seems to be an assumption in metaphilosophy that there is a common methodological role that thought experiments play throughout philosophy. In the next section, we will first argue that this assumption is mistaken, by pointing out two other methodological roles that thought experiments play in philosophy (there are, arguably, more, cf. Cohnitz in preparation). However, there is one role which seems to be the most interesting from a metaphilosophical point of view, namely the use of thought experiments as counterexamples or “alethic refuters”. In section III we will discuss several attempts at reconstructing the logical structure of such refutations by

thought experiments. As we will learn from these reconstructions, thought experimentation typically involves several *modal* judgments. This leads us to the epistemology of thought experimentation: why are we entitled to these modal judgments? A prominent answer is that we know the truth of the modal claims intuitively. This line has come under attack in the last decade by “experimental philosophers”, who empirically challenge the reliability of intuitive judgments about the kind of subject matter that philosophical thought experiments deal with. In section IV we will look at this discussion. Another prominent epistemology of modality tries to explain the modal knowledge involved in thought experimentation via the notion of conceivability. This will be the topic of section V. In section VI we will conclude with a few general remarks about the prospects of metaphilosophy and modal epistemology.

II. Roles of thought experiments

Talking of “the method of cases” or the “method of thought experimentation” suggests that there is only one (or at least one dominant) way in which philosophers put the considerations of hypothetical cases to use in philosophy. However, just as there are different functions for thought experiments in the sciences, there are different functions for thought experiments in philosophy. Let us characterize a thought experiment (in philosophy) as the consideration of a hypothetical case in the domain of things at issue. This characterization is intentionally vague. There are many things that can count as “considerations” (including, for example, being modeled in a computer simulation), and also many that count as being in a domain of things “at issue”.

Nevertheless, this liberal characterization excludes some hypothetical scenarios in philosophy from the class of thought experiments proper. For example, it excludes mere analogies, i.e. cases which are merely in some respect analogical to what is at issue (as, for instance, Plato's cave-scenario, which is supposedly analogical to our actual epistemic situation, which it is intended to illustrate).

It allows us, though, to speak of "illustrative thought experiments". These are thought experiments that are meant to illustrate the intended content of a theory or of a definition by providing a hypothetical example which exemplifies the content of that theory (perhaps in contrast with a rival theory) particularly well. Consider---as an example from physics---what would happen to the earth if the sun were suddenly to explode. According to the Newtonian theory, the earth would immediately departure from its usual elliptical orbit, while Einstein's theory predicts that the earth would stay in its usual orbit for another 8 minutes, the time it takes for light to travel from the sun to earth. Here, the hypothetical cases (the sun suddenly explodes) is used to illustrate the difference between Newton's and Einstein's theory. Thought experiments with this function are very common in philosophical texts.

The liberal characterization also allows to count "puzzle cases" as thought experiments. Puzzle cases are hypothetical scenarios which are intended to provoke a philosophical investigation. A prominent example is the trolley case-pair as it is often presented in ethics textbooks (and was originally presented as a pair in Judith Jarvis Thomson's 1985). One case of that pair is the so-called "standard trolley problem". In the standard

trolley problem you are standing by a railroad track when you notice that an uncontrolled trolley is rolling down the track, heading for a group of five railroad workers. You know that they will all be killed if the trolley continues on its path. The only thing you can do to prevent these five deaths is to throw a switch that will divert the trolley onto a side track, where it will kill only one person. When considering what you should do in such case, it seems that you should throw the switch and save the lives of five for the life of one.

The other case of the pair, the “footbridge case” is very similar. Again, the trolley is about to kill five people. This time, however, you are not standing near the track, but on a footbridge above the track. Also, this time you can’t divert the trolley by throwing a switch. You can only stop the trolley by putting something in its way that is heavy enough. You consider throwing yourself in front of the trolley, but you know that you are not heavy enough to stop it. Standing next to you, however, is a very large stranger. The only way you can stop the trolley killing the five people is by pushing this large stranger off the footbridge, in front of the trolley. If you push the stranger off, he will be killed, but you will save the other five. Again, considering what you should do, this time it seems wrong to push the stranger off the bridge even though, as before, you would save the lives of five for the life of one.

In Thomson’s presentation of the two cases, her intention was to provoke a theoretical analysis of why it is that we react differently to the two cases although they seem on a par with respect to their morally relevant aspects (it seems that in both cases you need to

decide whether you should kill one in order to save the lives of five by active intervention). Puzzle cases are cases in which it is either not clear what we should say about the case, or where our pre-theoretic judgment about the case seems incoherent (as in the trolley case), and they are typically intended to provoke or motivate a theoretical analysis.

A third way to use thought experiments is to use them as counterexamples against already existing theories. This is the use that most metaphilosophers seem to have in mind when speaking of thought experiments. Here is an example, borrowed from Jackson (1982). Physicalism in the philosophy of mind is the view that all facts (including all facts about the mind) are physical facts. Now imagine Mary, a woman that was born and raised in a black and white room and has never seen any color in all her life. However, in her black and white prison she was trained to become an expert in color perception. She now knows every physical fact that there is to know about the process of color vision and color experience. When for the first time confronted with a red tomato, Mary forms the thought “So that’s what it is like to see red.” It seems that Mary at that point learns something new; she learns what it is like to see red. This is knowledge of a fact about color experience. But by stipulation, Mary already knew every physical fact there is concerning color vision. Thus, the physical facts don’t exhaust all the facts and physicalism is false.

The use of thought experiments as counterexamples, (or---following Sorensen 1992----“alethic refuters”) gives rise to interesting epistemological and methodological

questions that we will discuss in detail below. These questions include the general puzzle of thought experimentation: how a merely hypothetical case can teach us anything interesting about the world and can even count as a counterexample to some (often well-established) theory. Moreover, how do we know what is possible and what isn't, and what would be the case in such a possible scenario?

The other uses, however, do not seem to give rise to the same questions. Whether a puzzle case is actual or merely hypothetical doesn't seem to matter for its heuristic value. Likewise, that an illustrative thought experiment presents a merely hypothetical case does not in any way diminish its pedagogical value. Also, there is no question of how we know what's the case in a hypothetical scenario. In an illustrative thought experiment, we know what's the case because we stipulated it, and in the puzzle case, there is no claim to knowledge anyway (it's the point of the puzzle that we don't know what to say about the hypothetical case).

Another observation about these different functions is of importance: Thought experiments can have histories and "lives of their own" (pace Hacking 1992). A hypothetical scenario C was perhaps once introduced with the intention of providing a puzzle case, but in the process of deliberating about it, it became clear what one should say about the case. Perhaps this judgment, let us call it (or, rather, its propositional content) J , became so convincing, that the case could then serve as a counterexample to a theory $T1$ which predicted a different evaluation for that case, while a rival theory $T2$ would predict precisely $C \rightarrow J$. At some point, textbooks explaining the content of $T2$

might cite C and its paradigmatic evaluation when explaining the content of T2. Also, a thought experiment might start as an illustration of a theoretical alternative and later become used in an argument for that alternative. Because one and the same hypothetical case might play such different functional roles, it can be difficult to determine in which function a thought experiment is intended (a thought experiment in one and the same text might even serve several such functions, either simultaneously for the same reader, or different ones for different readers).

The existence of these different functional roles is a source of confusion in contemporary metaphilosophical discussion. For example, experimental philosophers (which we will discuss below) are critical about the use of intuitive judgments when thought experiments are used as counterexamples. In his (2012) Herman Cappelen argues that this scepticism is unwarranted, because intuitions play no decisive role in thought experiments. However, Cappelen focusses in his argumentation mainly on thought experiments that are used as puzzle cases (cf. Cohnitz 2013, Chalmers 2014), while the experimental philosophers were criticizing thought experiments that play a role in theory choice, viz. alethic refuters.¹ (We will return to the matter below in section IV.)

III Content and Form of Alethic Refuters

¹ On an alternative interpretation, the debate between Cappelen and the experimental philosophers isn't confused about which thought experiments should matter for settling the dispute about whether intuitions play a role in philosophy, but Cappelen just (falsely) believes that all thought experiments function as puzzles.

In the previous section we stressed the diversity of roles that thought experiments may play within various theoretical disciplines. This section concerns the particular role we called “alethic refuters” (after Sorensen 1992): hypothetical cases intended to falsify a statement or theory by constituting a counterexample to it, in analogy with how ordinary experiments may falsify theories by providing negative instances. Viewed in this way, the use of thought experiments as alethic refuters falls under the broad rubric of hypothesis testing.

Just as philosophers of science have turned to logic in order to understand how testing of theories and hypotheses works, theorists of thought experiments have tried to use logic to study the fine mechanics of alethic refutation. In particular, an aim is to understand how certain thought experiments may be analyzed as arguments issuing in the conclusion that the theory under testing is false. Before briefly looking at some of these attempts, a few preliminary points need to be clarified.

The first concerns the tendency to focus on cases purporting to falsify theories (as hinted by Sorensen’s term). This is not due chiefly to any Popperian complaints about the notion of confirmation in general. Rather, there are many more influential instances of thought experiments used to oppose theories than there are of attempted confirmations. This is widely recognized (cf. Sorensen 1992, Williamson 2007, Malmgren 2011). Moreover, the fact that thought experiments involve hypothetical scenarios freely invented by their creators means that whatever problems surround the notion of confirmation in ordinary, non-hypothetical testing are exacerbated. Surely the claim that

I can invent a possible case which fits a certain general theory doesn't carry much weight towards showing that the theory is true. Such cases may abound, but they may often more naturally be taken as illustrations (for various purposes, see section II) of theories rather than tests of them, or as puzzle cases for which a theory that makes sense of them is sought (cf. section II).

Second, the fact that thought experiments are hypothetical means that any formal reconstruction has to be done in modal logic. Of course, this does not imply that any particular thought experimental scenario cannot also be actual (cf. Malmgren 2011, 279), or that the epistemic point it is used to make cannot also be made by an actual case (cf. Williamson 2005, 15; 2007, 192; Malmgren 2011, 273).

Third, it is important not to conflate the issues discussed in this debate (and this section) with the debate over whether thought experiments are identical to arguments, or may be replaced by them without epistemic loss. That question was prompted by John Norton's claim that thought experiments are just arguments adorned with particular details that are strictly irrelevant to their conclusions (Norton 1991). Norton offered this claim in response to Brown's suggestion that thought experiments constitute a peculiar vehicle for a priori knowledge (Brown 1986; 1991a; 1991b); prolonged debate ensued, mostly between Brown and Norton (Norton 1996, 2004a, 2004b; Brown 2004a, 2004b).

The current debate with which we are here concerned takes no stand on this question, however. Just as one may look at the argument form of a piece of reasoning for the

conclusion that a theory is false when the premises concern an experiment – and loosely talk of this is as the logical form of falsification – one may be interested in the structure of the corresponding reasoning in connection with thought experiments.

The parallel with ordinary experiments serves to amplify the difference between these debates, and to introduce a final preliminary point. When thought experiments are used to test theories, it seems natural to hold with Sorensen that they “are arguments if and only if experiments [in general] are arguments” (Sorensen 1992, 214). It also seems natural to hold that experiments in general are not arguments: the former, but not the latter, are entities with a spatio-temporal location; the latter, but not the former, have properties like validity and soundness and contain parts with truth-values. Regardless of whether experiments *are* arguments, however, they are, as we just noted, clearly connected with arguments, and one of the benefits of achieving “a fine-grained understanding of the arguments that underlie thought experiments” (as Williamson 2007, 180) puts it, is that we may get a clearer picture of the epistemic challenges posed by the premises – that is, what the experiment itself, understood as a process or event, is supposed to supply.²

Now just as an ordinary experiment in the field or a laboratory involves particulars – a particular set-up, particular unfoldings, and particular observations and reports of these – a thought experiment typically involves a particular hypothetical scenario. This

² It is not clear why Williamson here talks of arguments *underlying* thought experiments; it would seem more natural to say that certain arguments are based on certain experiments. But quibbles about directional metaphors need not detain us here.

suggests that the premise reporting the “outcome” of the thought experiment should also be particular (as in the case of ordinary experiments).

2. Motivations.

Why should we try to formalize alethic refuters? One motivation has already been mentioned above: if we are able to connect an interesting and sizeable class of such thought experiments with arguments sharing a common form, or a few forms, we may be in a better position to see what epistemic challenges they present, and the prospects of meeting them. Of course, just as individual arguments are distinct when the statements composing them are, different thought experiments will be connected with distinct arguments. But insofar as these arguments share logical form, they already have an interesting (if abstract) property in common. This standing motivation for seeking logical structure and form extends to the theory of thought experiments partly for the same reason as in other domains.

An interest in the epistemology of alethic refuters may take different forms. It may also be informed by various methodological and meta-philosophical preconceptions. The debate between Norton and Brown mentioned above was largely driven by Brown’s suggestion that the importance of thought experiments in science shows science to have a partly a priori character. The current debate concerning the form of alethic refuters in philosophy is sometimes informed by the goal of demonstrating the methodological integrity of philosophers’ use of such thought experiments without thereby construing this use as peculiarly a priori (Williamson 2007). Sometimes, however, a stated aim is

precisely to make room for a distinctively aprioristic view of alethic refuters and, with them, philosophy (Ichikawa & Jarvis 2009, 2012, Malmgren 2011). A subsidiary aim for some writers has been to provide an analysis reflecting psychologically real or possible routes to the argument's premises.

3. Three recent proposals

Early proposals concerning the form of alethic refuters employed propositional modal logic (Sorensen 1992, Häggqvist 1996). They also sought to apply proposed formal schemata to several influential alethic refuters in philosophy. By contrast, recent proposals tend to employ quantified modal logic. They have concentrated almost exclusively on one sort of closely related alethic refuters, *viz.* Gettier cases used as counterexamples to the JTB theory of knowledge. But the authors of these proposals also express the claim or hope that the proposal generalizes. Williamson says, for instance: "The discussion can be generalized to many imaginary counterexamples that have been deployed against philosophical analyses and theories in ways more or less similar to Gettier's." (2007, 180). As he immediately concedes, much work remains in order to show his proposal to be generally applicable. Similarly, Malmgren says that "the aim is to capture an argument form that is common to at least a core set of negative experiments" (2011, 272). But neither proceeds to actually attempt application to other cases.

To fix what we are talking about, let us consider a specific instance of a Gettier-like vignette [we may want to change this eventually]:

(S) Suppose that Smith believes that Jones owns a Ford, on the basis of seeing Jones drive a Ford to work and remembering that Jones always drove a Ford in the past.

From this, Smith infers that someone in his office owns a Ford. Suppose furthermore that someone in Smith's office does own a Ford – but it is not Jones, it is Brown. (Jones's Ford was stolen and Jones now drives a rented Ford.)

Now, does Smith know that someone in his office owns a Ford? The expected judgement is that while having a justified and true belief in this proposition, Smith doesn't know this.

Williamson suggests that the argument connected with a case like this has the form

- (W) (i)_w $\Diamond \exists x \exists p GC(x,p)$
(ii)_w $\exists x \exists p GC(x,p) \Box \rightarrow \forall x \forall p (GC(x,p) \supset (JTB(x,p) \& \neg K(x,p)))$
 $\therefore \neg \Box \forall x \forall p (K(x,p) \leftrightarrow JTB(x,p)),$

where the variables “x” and “p” respectively range over subjects and propositions, “GC(x,p)” says that x stands to p in the relation specified by (S), “JTB(x,p)” that x has justified true belief in p, and “K(x,p)” that x knows that p. (We will use this lexicon throughout this subsection.)

On this rendering, as on the others shortly considered, the conclusion is that the JTB theory of knowledge, here understood as a metaphysical claim rather than conceptual analysis, is false. A thinker coming to this conclusion on the basis of Gettier's alethic refuter does so because she accepts, after contemplating the case as presented in (S), the premises of (W). The first premise asserts the metaphysical possibility of someone being related to some proposition in the same way as (S) specifies. The second premise expresses the claim that if someone were thus related to a proposition, she would have justified true belief in it without knowledge.³ The latter judgement is of particular interest, since it is intended to capture what is usually called the intuition, or intuitive judgement, concerning Gettier cases. On Williamson's proposal this is a counterfactual conditional. But it also highlights the claim that the scenario described in the vignette is possible. For most Gettier cases, this claim is trivial. But without it, the argument wouldn't be valid.

It is important to note that the hypothetical scenario of the vignette – the “GC” predicate of (W) – is itself neutral with respect to epistemic properties: a person's and a proposition's satisfying “GC(x,p)” is itself compatible both with her knowing and with her not knowing that proposition: the scenario itself holds no prejudice with respect to the outcome. This is analogous to the distinction between set-up and observation (or data) in an ordinary experiment: one may know the former without being in any way committed about the latter.

³ It does not quite say this, a point we'll return to shortly. Williamson discusses the merits of (ii)_w at some length (2007, 195-199).

Williamson's proposal has been criticized on various grounds, all (notably) concerned with its internal viability as an analysis of Gettier cases, rather than with any difficulties of generalization. Specifically, Ichikawa (2009) and Malmgren (2011) both complain that the major premise

$$(ii)_w \quad \exists x \exists p GC(x,p) \square \rightarrow \forall x \forall p (GC(x,p) \supset (JTB(x,p) \& \neg K(x,p)))$$

may be false for reasons apparently unrelated to what it is aiming to capture. Consider that someone may satisfy the antecedent in the nearest world where it is true (perhaps even the actual world) but also, as it happens, have “good reasons to believe that he is prone to hallucinate people driving Fords to work and prone to misremember what cars people drove in the past” (Malmgren 2011, 279). Since such a person will not have justified true belief, the judgement will be false if construed as $(ii)_w$. Someone might also satisfy the antecedent but have grounds for believing the proposition, additional to and independent from those mentioned in (S). $(ii)_w$ may then be false because such a person does have knowledge. But – the objection continues – such instances are *deviant* in relation to what the intuitive judgement about Gettier cases should be. Hence $(ii)_w$ does not capture the intuitive judgement made about such cases.

Moreover, since the consequent of $(ii)_w$ is a universal statement, $(ii)_w$ may be false simply because among the persons in the closest antecedent-world standing in the relation to a proposition specified by (S), only some, but not all, happen to have justified true belief in this proposition without knowing it. This case arises because $(ii)_w$ does not

exactly match the anaphoric binding suggested in Williamson's own version of the judgement in English: "If a thinker were Gettier-related to a proposition, he/she would have justified true belief in it without knowledge" (Williamson 2007, 195).

Williamson's chief reply is that we should accept that our judgements about alethic refuters may be mistaken.⁴ In cases like those broached by Ichikawa and Malmgren, we should admit that our judgement was wrong (should we discover this), and amend it accordingly by strengthening the stipulations in the vignette so as to rule out the unwanted instantiations of the "GC" predicate (Williamson 2009). His response is thus, in effect, to chalk such unwanted instantiations up to general fallibility, rather than admitting them as deviant: "We cannot realistically expect that the method of thought experiments in philosophy will turn out to be much more reliable than the methods of the natural sciences." (2009, 469).

Just what the "GC" predicate expresses depends on what the stated scenario of a given Gettier cases stipulates, of course. Hence, Ichikawa notes, the risk of (ii)_w being false due to unintended, accidental realizations seems inverse to the specificity of the scenario. This is why strengthening the scenario is a way of doing away with such realizations. It may seem odd, however, that a thought experiment should be better – in the sense of running less risk of such deviant instantiations – simply for mentioning more specific facts even if these seem entirely irrelevant to its point. Take (S) again. Clearly, stipulating not only the make of the car (a Ford) but also its production year

⁴ He also hints at domain restriction, but notes its limits as remedy (2007, 200). Cf. Ichikawa (2009, 437 and 440-442).

diminishes the risk of deviant instantiations of the antecedent. Ichikawa complains that such specificity should not be viewed as improvement (Ichikawa 2009, 440).

Ichikawa's concern with deviance is largely epistemological: "Williamson's account renders it much too difficult to know the Gettier intuition" (Ichikawa 2009, 440).

Malmgren emphasises what she sees as a semantic mismatch between what $(ii)_w$ expresses and what is actually judged concerning (S), and that this is shown by the fact that $(ii)_w$ doesn't rule out deviant realizations (2011, 279). She also notes that her complaint appears parallel to the considerations that lead Williamson himself to reject e.g.

$$\Box \forall x \forall p (GC(x,p) \supset (JTB(x,p) \& \neg K(x,p)))$$

as capturing the judgement (Malmgren 2011, 275-280; Williamson 2007, 184-185).⁵

These criticisms are connected with a concern that a formal account of alethic refuters should allow them to be epistemically successful. On this view, a proposal should render knowledge of the "outcome" premise practicable; on Ichikawa & Jarvis's (2009) and Malmgren's (2011) proposals are motivated by a claim that it should be knowable a priori. Ichikawa & Jarvis (2009) and Malmgren (2011) worry that Williamson's account makes the central premise too hard to know, since knowing the counterfactual requires

⁵ Deviance considerations also appear to lead Williamson to reject $\forall x \forall p (GC(x,p) \Box \rightarrow (JTB(x,p) \& \neg K(x,p)))$ as unviable (see Williamson 2007, 196-197).

knowing that there are no counterinstances to it. They also offer alternative proposals which eschew construing the judgement as a counterfactual.

Ichikawa & Jarvis (2009) argue that we may reasonably take contemplators of a thought experiment to “fill out” the scenario with propositions beyond those explicit in a vignette. Thus a more complete scenario is contemplated; they suggest that the judgement concerning such a scenario of a Gettier case may be construed as a strict conditional:

$$(ii)_{IJ} \quad \Box(g \supset \exists x \exists p(JTB(x,p) \ \& \ \neg K(x,p))),$$

where “g” denotes the enriched scenario, identified in thought by a demonstrative “things are like *that*”. Together with the possibility premise

$$(i)_{IJ} \quad \Diamond g$$

entail the sought conclusion (as understood by Williamson, i.e.: $\neg \Box \forall x \forall p(K(x,p) \leftrightarrow JTB(x,p))$). In contrast to Williamson’s (ii)_w, Ichikawa & Jarvis’s (ii)_{IJ} renders the outcome judgement as a necessity judgement, so as to safe-guard it against the vicissitudes of contingency (2009, 223).

Williamson rejoins that the enrichment supposedly blocking deviant instantiations cannot consist in explicit consideration of various alternative ways of filling out the

fiction, hence must consist in dispositions to enrich the scenario beyond the stated text; that such dispositions are likely to vary between different contemplators of a case, thereby threatening public debate about it; and that competing dispositions to enrich may co-exist within a subject, thereby threatening the plausibility of the first premise (i)_{II}. Moreover, he claims, even a richer scenario constructed along the lines envisaged by Ichikawa and Jarvis (2007) and Ichikawa (2009) may be subject to deviant realizations; thus, even if construed as (ii)_{II}, the judgement risks unintended falsity (Williamson 2009, 466-468). If Williamson is right about the intrasubjective coexistence of competing dispositions to enrich the scenario, this presumably also threatens a contemplator's demonstrative thought ("things are like *that*") with failure to refer to a determinate proposition *g*.

Malmgren argues that construing alethic refuters as fiction lets in deviant realizations because what is a legitimate enrichment for a fiction need not be one for a scenario in the vignette of a thought experiment, and vice versa: "What is true/false/indeterminate in a problem case does not line up (across the board) with what is true/false/indeterminate in a fiction" (Malmgren 2011, 304).

Her own proposal is that the Gettier judgement is a possibility judgement (Malmgren 2011, 281):

$$(i_M) \quad \diamond \exists x \exists p (GC_{x,p} \ \& \ JTB_{x,p} \ \& \ \neg K_{x,p}).$$

As in the other proposals, the modality here is metaphysical. (i_M) yields the same conclusion as the other accounts without additional premises. It also, Malmgren argues, meets the demand of being knowable a priori (though it will presumably not be necessary unless S5 is assumed).

4. A few comments

Williamson's account has been criticized on the grounds that it makes judgments about alethic refuters contingent. The felt force of such objections will vary depending on one's general views about what philosophy is. To someone for whom the subject matter of many branches of philosophy is not well characterized as conceptual analysis, but rather to be regarded as continuous with the attempts of science to learn about various domains, these objections may be less pressing; whereas they will carry force with adherents of what Ichikawa and Jarvis call "the traditional view of thought experiments and intuitions" (2009, 223). Clearly one's attitude will also depend on whether one takes the distinction between a priori and a posteriori knowledge (or justification) to be valid and important, or not. Williamson has repeatedly expressed the view that this distinction "is a superficial one, of little theoretical interest" (2013; see also Williamson 2007).

Although in a spirit very different from the gradualism between philosophy and science endorsed by Quine, he has also repeatedly stressed that the target of alethic refuters such as Gettier's is not an analysis of the concept of knowledge, but a theory about the phenomenon of knowledge.

Parties to these debates tend to be non-skeptics about alethic refuters (at least within philosophy). They also operate within what might loosely be called a non-Platonist epistemology. Hence, they recognize that the premises of their accounts should be non-mysteriously knowable to contemplators of cases in at least some instances (and the focus on Gettier cases is probably motivated partly by the recognition that such cases appear to be successful alethic refuters). Williamson (2007) appeals to a general human capacity for evaluating counterfactuals, which, while fallible and not distinctly a priori, is reliable enough to allow such knowledge. Ichikawa and Jarvis claim that “whatever explains our capacity for everyday knowledge should also be able to explain knowledge of thought-experiment intuitions” (2009, 235-236), arguing that this should make a priori knowledge of the the premises of alethic refuter – construed as metaphysical possibility and necessity claims – non-mysterious. While both Williamson and Ichikawa & Jarvis do elaborate on these suggestions, it is fair to say there these issues are unresolved at present.

IV. Intuitions and Metaphysical Modality

Whatever the exact genesis of our knowledge of the premises of an alethic refuter may be, many philosophers hold that the major premise of a thought experiment is in any case judged *intuitively*. Philosophers disagree about what exactly they mean by “intuitive”, but it seems that they at least hold that it means that the major premise is not arrived at via any conscious inference from other knowledge.

1. The Experimental Challenge

In the last decade, so-called “experimental philosophers” (or “*xphiles*”) have criticized the use of thought experiments in philosophy on the basis of the idea that the crucial premise of thought experiment arguments are established by intuitive judgment. In a series of empirical experiments they try to show that non-inferential judgments about philosophical thought-experiment scenarios vary with philosophically irrelevant factors and hence should not be trusted. For example, the experiments show that a particular version of a Gettier case (as described above) will elicit (on average) different spontaneous evaluations depending on whether the person making the judgment is (a non-philosopher) from a Western or from an East-Asian cultural background (Weinberg, Nichols & Stich 2001; Machery et al. 2004). Similar studies show that judgments may vary with socio-economic status, gender, the order of the cases presented, or whether or not the judgment was made in a clean or a messy environment.

Xphiles draw different conclusions from their results. The most modest conclusion is that philosophers shouldn’t simply assume that their intuitive judgments are widely shared, but should empirically check whether they really are (Weinberg, Nichols, and Stich 2001). More radical *xphiles* demand a moratorium on the use of intuitions in philosophy until we know under which conditions our intuitions are not subject to such variation (Weinberg 2007), and the most radical suggest that we should give up on whole branches of theoretical inquiry, since we don’t at present possess reliable methods for finding out the truth in these areas of inquiry (Mallon et al. 2009).

Experiments typically involve a description of a thought experiment scenario, followed by a forced choice question. Such a “vignette”, as used by Weinberg, Nichols, and Stich in their 2001, reads as follows:

Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

REALLY KNOWS

ONLY BELIEVES

When Weinberg, Nichols and Stich tested the probe on undergraduates at Rutgers university, it turned out that the majority of students with an East Asian cultural background as well as the majority of students with a cultural background from the Indian subcontinent would judge that Bob *knows* that Jill drives an American car, while the majority of students with a Western cultural background judged that Bob *only believes* it.

If one does not want to conclude from such results that knowledge is culturally relative, one seems forced to conclude that intuitive judgments about hypothetical cases are not providing us with reliable evidence. There are several ways, however, to endorse the second of these disjuncts, without giving up on the method of thought experimentation.

2. The Expertise Defense

One way is to agree that intuitive judgments are not reliable, if the judgments are made by non-experts. As we said above, early x-phi studies tested the judgments of undergraduate students with little or no training in philosophy. The so-called “Expertise Defense” against the challenge coming from x-phi holds that the intuitive judgments of undergraduate students might well be unreliable, however the intuitive judgments that play a role in theory choice in philosophy are the judgments of philosophical *experts*. That *their* intuitions vary has, however, not been shown (and they presumably don’t, because otherwise thought experiments could play no role in theory choice).

There are (at least) three ways in which this defense can be fleshed out, only one of which seems somewhat promising. One way to argue for the superior expertise of philosophers and its relevance for the evaluation of hypothetical cases is to point out that intuitive judgments, just as all judgments, are theory-laden. If so, then we should trust those judgments more that are more likely to be influenced by correct theories rather than incorrect theories. Since professional philosophers are likely to be informed by correct theories, and undergraduates are not, we should trust the judgment of philosophers, and discard the judgments by undergraduates as unreliable (for a defense along these lines, see Devitt 2006, Devitt 2011).

The problem with that strategy is that this particular view on the epistemology of intuitive judgments undermines the kind of evidential role they are supposed to play. If intuitive judgments are a mere expression of one’s theoretical commitments, then intuitive judgments about hypothetical cases can never provide us with new evidence

against our currently held theories, since we already need to be committed to a better theory in order to generate recalcitrant judgments that are informed by good theory.

An alternative way of fleshing out the Expertise Defense is to point out that philosophers are experts about the subject matter at issue because of having paid a lot of attention to cases of, say, knowledge and mere belief. On this view, philosophers are more reliable in making correct judgments about hypothetical cases because they are familiar with a bigger data set from which to generalize (Devitt 2011, Devitt 2012).

The problem with this version of the Expertise Defense is that it makes a certain *empirical* claim. It claims that people who are professionally more exposed to cases of certain kinds should also become more reliable in their judgment. Unfortunately, this claim doesn't seem to hold up. In a study that tested this empirical assumption for an Expertise Defense of semantic intuitions, Machery 2012 found that experts of different fields of semantics had significantly different judgments about hypothetical cases in philosophical semantics than their philosophical colleagues, although they are professionally just as much confronted with the same linguistic phenomena as their colleagues. Thus, the homogeneity in philosophical judgments about hypothetical cases is not explained by the expertise of philosophers that stems from exposure and attention to relevant cases.

A third way of fleshing out the Expertise Defense would highlight that philosophical evaluations of hypothetical cases are more likely to be reliable because (i) they are more

familiar with the specific use of thought experiments in their area of inquiry and (ii) they know better which kind of intuitive reaction is of relevance and which terminology properly reports that reaction. (i) is based on the observation that the description of a hypothetical scenario in a philosophical text does not settle every detail of the case. We already discussed above that a hypothetical scenario can be enriched to “deviant instances” which wouldn’t any longer support the same intuitive evaluations. This is a feature that thought experiment descriptions share with other fictional texts.⁶ In the latter case it is a matter of familiarity with the relevant *genre* to know how the open details of a story may or may not be filled out by the reader in order to draw conclusions from the text that are merely implicitly contained in it. Likewise, it is a matter of familiarity---not just with the genre “thought experiment”---but with the genre *thought experiment in this particular theoretical context* that enables the philosopher to fill out the details of the hypothetical case in the way relevant for the function the thought experiment is supposed to play in a particular theoretical inquiry (Cohnitz 2006, Ichikawa & Jarvis 2009). (ii) is based on the observation that philosophers typically know better what is at stake, and how to use the relevant terminology. To see what we mean, consider the use linguists make of grammaticality judgments. Instead of asking lay persons whether a certain string seems “grammatical” to them, they sometimes ask about whether the string seems “acceptable”, in order to prevent whatever views on grammar the subject might have to taint the judgment. However, there is a danger that this “fix” does more harm than good because a string of words can seem acceptable or unacceptable in many

⁶ For example, in a Sherlock Holmes story it will be safe to assume that the standard laws of nature and biology hold (at least to the extent they were known to the author of the story); on the other hand, when reading a fantasy novel this is not safe to assume.

ways, while the linguist is only interested in its *grammatical* acceptability. A linguist, unlike a lay person, would know that, and is hence less likely to misunderstand the intended sense of “acceptability” if she were given the same task. Moreover, since the linguist knows what is at issue, she would also be able to answer questions framed in terms of “grammaticality” directly in the intended way. Likewise, philosophers might be better placed than lay persons to evaluate thought experiments, since they might know better how to apply the relevant theoretical terminology in reports of their intuitive judgments (Cohnitz & Haukioja 2006).

Of course, also this last version of fleshing out the Expertise Defense is based on certain empirical assumptions.⁷ They might seem plausible, but there is still work to do to show that is really what explains the difference in intuitive judgment between the test subjects that participated in the x-phi studies and the professional philosophers.

3. The Irrelevance of Intuitions for the Method of Cases

Herman Cappelen (2012) and Max Deutsch (2015) argue for another response to the Experimental Challenge. According to them, the challenge is simply misplaced, because intuitive judgments do in fact not play any evidential role in contemporary analytic philosophy. The main premises of thought-experiment arguments aren't presented and aren't established as based on intuitions, but based on argument and reasoning. Hence,

⁷ And, at least for linguistics, the empirical assumptions seems not to hold, as is argued in Culbertson & Gross 2009.

the empirical result that intuitions vary a lot with irrelevant factors is an interesting result in itself but of no relevance for philosophical methodology.

Both, Cappelen and Deutsch, support their claim by showing that philosophers typically back up their evaluation of a hypothetical case by arguments. Now, surely, in order to establish that a certain claim holds in generality for philosophy, this kind of empirical analysis is the way to go. However, there are some issues that have not been sufficiently addressed in Cappelen's or Deutsch's study.

The first issue has to do with the selection of cases. As we explained above, thought experiments may serve different functions in philosophy. At least Cappelen's selection of cases has been criticized for being somewhat idiosyncratic and for focussing on puzzle cases, rather than on alethic refuters, which is where the metaphilosophical action is (Cohnitz 2013, Chalmers 2014).

The second issue is that it might be difficult to determine whether something is an argument for a conclusion (where the premises of the argument carry the evidential weight for the conclusion), or whether something is a retrospective explanation of an intuitive judgment (where the evidential weight is carried by the judgment). Perhaps the

apparent argument is just cited for heuristic purposes or a mere rhetorical maneuver, as

Julia Langkau explains:

[T]here could be other, not purely epistemic reasons why an author may give arguments. First, the arguments could serve a purely heuristic purpose, i.e., they could help us to get to the truth without actually playing a justificatory role.

Usually, one would think that intuitions are used for heuristic reasons, and that the arguments in support of their contents bear the epistemic burden. However, a view according to which our intuitions have Rock status and our arguments merely serve to rationalize them is not completely implausible. Second, an author might give arguments merely for conversational or rhetorical reasons.

While it is obvious to the author that the intuition has Rock status, it is not to their audience. In order to convince the audience, it might be easier to give arguments for the truth of the content of the intuition than for the claim that the intuition has Rock status. Third, arguments might be required in order to get a paper published. Surely there are certain presentational requirements on philosophical text which favour certain methods over others. Maybe it is not sufficient in academic philosophy to simply give the thought experiment without

support from arguments. In all three cases just mentioned, intuitions could still provide the relevant justification [...]. (Langkau, forthcoming)

In a rational reconstruction of an argumentation, it might well be possible to argue that the apparent argument for the relevant claim is not intended as an argument. For example, if the premises of the proposed argument are question begging in the dialectical situation at issue, it might be more reasonable to suppose that this was in fact a retrospective explanation of the intuitive judgment. However, neither Cappelen⁸ nor Deutsch seem to pay sufficient attention to the problem.⁹

Finally, there are several philosophical projects in which the targeted subject matter is explicitly conceived as grounded in psychological competences which manifest themselves in certain intuitive reactions (for example, concept applications, interpretations or productions of utterances, attributions of mental states, etc.). Given the aim of these projects, it makes total sense to “psychologize” their evidence, because psychological processes is what these projects are after. Cappelen and Deutsch seem to assume that they are a clear minority or that they are methodologically confused, but---as these projects often make their intentions explicit---we know they do exist in considerable number, and it’s not clear (and not addressed sufficiently by Deutsch or

⁸ Cappelen briefly discusses the issue in his 2012, and dismisses it by shifting the burden of proof to the other side.

⁹ As Ole Koksvik argues in his 2013, a judgments that is intuitively true can be arrived at through conscious reasoning. But its intuitiveness might still be what determines its evidential status, and an author might describe a reasoning-route towards the judgment to help others see the truth of the judgment.

Cappelen) why one should regard all such projects as confused or flawed or unphilosophical. However, for philosophical projects in which the targeted subject matter is not supposed to be grounded in psychological competences, but rather concerned with metaphysical truth, it seems that intuitions should not and probably also do not play a substantial evidential role in philosophy. But that still leaves the matter open how we know that major premise of a thought experiment argument is true.

IV. Conceivability and metaphysical modality

As mentioned in sections II and III, reconstruing thought experiments as (valid) modal arguments pinpoints the question of how we are to know the premises of such arguments. The previous section dealt with the prospects of appealing to intuition. In this section, we will briefly canvass appeals to conceivability. In line with the main proposals mentioned in section III, it will be understood here that the modality at issue is so-called metaphysical modality (i.e. modality *de re*).

Metaphysical modality is usually contrasted with so-called conceptual modality. As its name suggests, however, conceivability appears to concern in the first instance what is conceptually possible. Since it is widely thought that what is conceptually possible need not be metaphysically possible, and that what is metaphysically necessary needn't be conceptually so, the conceivability theorist needs to somehow account for the difference. Currently popular versions of appeals to conceivability do this by appealing to collateral, a posteriori, information about the actual world.

Adherents of two-dimensional semantics – to wit, Jackson (1996) and Chalmers (1996) – prefer to gloss such information rather as information about which world is actual. On their view, the applicability of a term at a world w may be judged either in accordance with the assumption that w is the actual world, or in accordance with whatever *is* the actual world. A term thus has both what they call a primary extension (and intension) and a secondary extension (and intension).¹⁰ For some terms, these diverge. Since it is the secondary intension that determines the metaphysically modal profile of a term, this is what a thinker needs to know in order to know what is metaphysically possible, impossible, or necessary. However, although the secondary intension may be knowable a priori *conditional* on what world is the actual one, the antecedent – that is, what is the actual world – can only be known a posteriori. Thus, Jackson, says, “our best reasons for concluding that certain claims which aren’t conceptually necessary are metaphysically necessary derive from claims that are about what is or is not conceptually necessary conjoined with a posteriori claims based on experimental results in some broad sense” (Jackson 2009, 106). To illustrate, the claim that water is H_2O is metaphysically necessary is supposed to flow from conceptual knowledge that *if* water is H_2O , it is so necessarily, together with empirical knowledge that water is actually H_2O .¹¹

Another recent exploitation of this idea is what Ichikawa & Jarvis (2012) call their moderate modal rationalism (or MMR). On their suggestion, a claim is metaphysically

¹⁰ Jackson (1998, 48) uses the labels “A-extension” and “C-extension” respectively.

¹¹ This is perforce sketchy. For a fuller exposition of two-dimensionalism’s modal epistemology, as well as elaborations of it, see especially Chalmers (1996) and (2002).

possible just in case it is conceptually possible and no falsity about the actual world is conceptually necessitated by it. Thus (with the subscript “M” indicating metaphysical and “C” conceptual modality, and “@” an actuality operator):

$$\text{MMR: } \Diamond_{MP} \leftrightarrow (\Diamond_C p \ \& \ \neg \exists q (\neg @q \ \& \ \Box_C (p \supset @q))). \text{ (Ichikawa \& Jarvis 2012, 147)}$$

Since what is conceptually possible and conceptually necessitated may plausibly be within epistemological reach of conceivability, MMR holds out hope for a route to knowledge of metaphysical modality based on conceivability together with broadly speaking empirical knowledge, just as two-dimensionalism.¹²

Space considerations prevent extensive discussion of these proposals here. We’ll just make two brief remarks. A dialectical oddity about two-dimensionalism, in this context, is that the schema or grid specifying secondary intensions conditional on which world is (taken as) actual appears itself to depend on consideration of various thought experiments and their outcomes. Concerning Ichikawa and Jarvis’s MMR, one may note that the second clause of its right-hand side requires only absence conceptual necessitation of potential defeaters. It seems fair to ask whether this is enough.

¹² Ichikawa & Jarvis (2012, 135) explicitly reject two-dimensionalism. Although its commitments are weaker, however, their proposal retains the element under discussion here. The strategy of invoking conceivability, broadly speaking, informed by or conjoined with a posteriori knowledge of the world is of course exemplified by many more accounts, including ones as diverse as Armstrong (1989) and Yablo (1992).

Although very different in outlook, Williamson's own modal epistemology might (perhaps with some strain and surely against his wishes) be regarded as a species of conceivability approach. This epistemology makes central appeal to counterfactuals and our capacity to evaluate them. Williamson suggests that they are evaluated by supposing the antecedent and adding further judgements by using imagination, logic, and constraining background knowledge. A counterfactual is to be accepted just in case such development eventually leads one to add its consequent; it is rejected when the consequent robustly fails to emerge after suitably diligent and varied development (Williamson 2007, 152-153). Insofar as "supposing" is kin with "conceiving", this approach is perhaps at least a relative of conceivability accounts. Like others, it allots a role to factual knowledge about the world; ineliminably so since counterfactuals (often enough) express worldly connections and dependencies (and so are "metaphysical"). As Williamson says, evaluation of them "can in principle exploit all our background knowledge" (2007, 143).

Williamson is happy to let knowledge of metaphysical modality be fallible, and not distinctively a priori. As we saw in section III, other theorists, including Ichikawa and Jarvis, insist that this makes thought experimental knowledge too hard to come by, and complain that "Williamson's account leaves intuitions [i.e., judgements about thought experiments] as mere judgments about contingent matters of fact." (Ichikawa & Jarvis 2007, 226). However, it should be noted that none of the conceivability-based accounts of knowledge of metaphysical modality broached in this section eliminates dependence on contingent matters of fact entirely.

V. Theory of thought experiments and the prospects of metaphilosophy

As (we hope) this survey shows, the discussion of thought experiments is very lively in current metaphilosophy. However, it seems to us that there are at least two issues that the current debate is not sufficiently taking into account.

One issue we already alluded to above. Philosophy is a very diverse field with different areas of inquiry, each of which containing a wide variety of views of how the subject matter of the area should be thought of, and how we can have knowledge of that subject matter. This by itself isn't in the way of there being one particular methodological tool that could still serve all of philosophy. For example, pretty much regardless of how you conceive of the subject matter you think you are dealing with and the epistemic access you believe you have, to argue for your view in a logically sound way seems almost always to be a good methodological advice. But as our survey has probably shown, the method of thought experimentation isn't quite the neutral arbiter that logic is.¹³

Reconstructing the method of cases involves substantial assumptions about the subject matter that philosophy is dealing with, the nature of the relevant evidence, and an epistemology concerning that evidence. This will make it unlikely that there is one reconstruction of the method of cases that fits all its instances in philosophy. We noted already that there are different functions for thought experiments in philosophy. We suspect that the functional characterisation 'alethic refuter' might be still be too broad, and that there are different ways in which a thought experiment can serve as an alethic

¹³ And even logic isn't a completely neutral arbiter either, cf. Williamson 2014.

refuter which would require different stories about how the crucial major premise of an alethic refuter gets established. Perhaps metaphilosophy could make better progress in narrowing its scope, and confining its claims only to certain subareas of philosophy, which plausibly have a common epistemology.

In another sense, metaphilosophy should perhaps broaden its scope. Currently thought experiments in philosophy are discussed in isolation from thought experiments in the sciences, despite the fact that several authors have pointed out strong similarities between the use of hypothetical cases in both (Sorensen 1992, Häggqvist 1996, Cohnitz 2006). And, indeed, one might learn from the analysis of thought experiments in the sciences for the case of philosophy. As Chalmers points out in his 2014, the special epistemic status of philosophical intuitions should perhaps best be thought of as their *dialectical* justificatory status. As he puts it, “what is distinctive about appeals to intuition is that intuitive claims are taken to have a dialectical justification that is broadly noninferential” (537). This is not necessarily far from the view of Cappelen (2012), who holds that what matters about the crucial premises in thought experiment arguments is not so much their intuitiveness, but their being in the common ground.

This observation matches a view that has been defended for thought experiments in the sciences (Gendler 2000, Kühne 2005, Cohnitz 2006). These authors argue that thought experiments are best analyzed when considered in the dialectical context in which they were presented. They also emphasize that considering a thought experiment in its dialectical context provides us with the constraints a hypothetical scenario has to satisfy

in order for it to constitute a relevant counterexample. Moreover, this perspective on thought experiments enables us to understand the very point of thought experimentation: why scientists use an imaginary *concrete* case in their argumentation. The idea is that a thought experiment manages to present a case for which a certain judgment is in the common ground between all parties to the debate, even though no generalization of that judgment is available (at the time) that would likewise be in the common ground. For example, as Tamar Gendler (200) argued, in his famous falling bodies thought experiment Galileo Galilei would have needed a claim like “Entification is not physically determined” in order to argue with general premisses against the Aristotelian. However, there would have been various ways for the Aristotelian around accepting such claim. On the other hand, for the concrete case of a cannon ball chained to a musket ball there is no room to claim that how tightly they are connected (and thus, whether you should consider them as one physical body or two) will matter for how they fall. That this was true for this concrete case was in the common ground - although the general principle was not.

Something very similar can be said for some philosophical thought experiments.

Consider Frank Jackson’s thought experiment with Mary the color scientist again. For the concrete case it is in the common ground (for most participants to the debate) that Mary learns something new, when seeing the red tomato for the first time. On the other hand, it is safe to assume that no general premise establishing learnability of something new despite full knowledge of physical details would have been dialectically accessible to Jackson (cf. Cohnitz 2006, forthcoming). Again this would explain why Jackson chooses to argue with a thought experiment in the first place. We believe that it is here

that metaphilosophy could gain important insights from philosophy of science, further clarifying the dialectical function of thought experimentation.

References:

- Brown, J.R. (1986) Thought Experiments Since the Scientific Revolution, *International Studies in the Philosophy of Science* 1, 1-15.
- Brown, J.R. (1991a) *The Laboratory of the Mind*. Routledge.
- Brown, J.R. (1991b) Thought Experiments: a Platonic Account. In T. Horowitz and G. Massey (eds), *Thought Experiments in Science and Philosophy*, pp. 119-128. Rowman & Littlefield.
- Brown, J.R. (2004a) Why Thought Experiments Transcend Empiricism, in C. Hitchcock (ed), *Contemporary Debates in Philosophy of Science*, pp. 23-43. Blackwell.
- Brown, J.R. (2004b) Peeking Into Plato's Heaven, *Philosophy of Science* 71, pp. 1126-1138.
- Chalmers, D. (1996) *The Conscious Mind*. Oxford UP.
- Chalmers, D. (2002) Does Conceivability Entail Possibility? In Gendler, T. S. and J. Hawthorne (eds), *Conceivability and Possibility*, pp. 145-200. Oxford UP.
- Chalmers, D. (2014) Intuitions in philosophy: a minimal defense, *Philosophical Studies* 171 (2014), 535-544.
- Cohnitz, D. (2006) *Gedankenexperimente in der Philosophie*. Mentis.
- Cohnitz, D. (2012) Philosophy without Intuitions, by Herman Capellen. *Disputatio* 33, 546-553.

- Culbertson, J & Gross, S. (2009) Are Linguists Better Subjects? *British Journal for Philosophy of Science* 60, 721-736.
- Devitt, M. (2011) Experimental Semantics. *Philosophy and Phenomenological Research* 82, 418-435
- Devitt, M. (2012) Whither experimental semantics? *Theoria* 27, 5-36.
- Gendler, T. (2000) *Thought Experiments: On the Power and Limits of Imaginary Cases*. Garland Publishing.
- Häggqvist, S. (1996) *Thought Experiments in Philosophy*. Almqvist & Wiksell International.
- Ichikawa, J. (2009) Knowing the intuition and knowing the counterfactual. *Philosophical Studies* 145, 435-443.
- Ichikawa, J. & B. Jarvis (2009) Thought-experiment intuitions and truth in fiction, *Philosophical Studies* 142, 221-246.
- Ichikawa, J. & B. Jarvis (2012) Rational Imagination and Modal Knowledge. *Noûs* 46, 127-158.
- Jackson, F. (1998) *From Metaphysics to Ethics*. Oxford UP.
- Jackson, F. (2009) Thought Experiments and Possibilities. *Analysis Reviews* 69, 100-109.
- Koksvik, O. (2013) Intuition and Conscious Reasoning. *Philosophical Quarterly* 63, 709-715.
- Kühne, U. (2005) *Die Methode des Gedankenexperiments*. Suhrkamp.
- Langkau, J. (forthcoming) Experimental Philosophy: Against Undermining the Challenge.

- Machery, E. (2012) Expertise and Intuitions about Reference. *Theoria* 27, 37-54.
- Malmgren, A.-S. (2011) Rationalism and the Content of Intuitive Judgments. *Mind* 120, 263-327.
- Norton, J. (1991) Thought Experiments in Einstein's Work. In T. Horowitz and G. Massey (eds), *Thought Experiments in Science and Philosophy*, pp. 129-148. Rowman & Littlefield.
- Norton, J. (1996) Are Thought Experiments Just What You Thought?, *Canadian Journal of Philosophy* 26, 333-366.
- Norton, J. (2004a) Why Thought Experiments Do Not Transcend Empiricism. In C. Hitchcock (ed), *Contemporary Debates in Philosophy of Science*, pp. 44-66. Blackwell.
- Norton J. (2004b) On Thought Experiments: Is There More to the Argument?, *Philosophy of Science* 71, 1139-1151.
- Sorensen, R. (1992) *Thought Experiments*. Oxford UP.
- Williamson, T. (2007) *The Philosophy of Philosophy*. Blackwell.
- Williamson, T. (2009) Replies to Ichikawa, Martin and Weinberg. *Philosophical Studies* 145, 465-476.
- Williamson, T. (2013) How Deep Is the Distinction Between A Priori and A Posteriori Knowledge? In A. Casullo and J. Thurow (eds), *The A Priori in Philosophy*. Oxford UP.
- Williamson, T. (2014) Logic, Metalogic, and Neutrality, *Erkenntnis* 79, 211-231