Topoi (2019) 38:811–820
DOI 10.1007/s11245-016-9447-3

The 'If' in the 'What If'

Daniele Sgaravatti[1]

Published online: 19 December 2016

Abstract In this paper, I defend the view that any good account of the logical form of thought experiments should contain a conditional. Moreover, there are some reasons to think it should be a counterfactual conditional. First, I defend Williamson's account of the logical form of thought experiments against a competing account offered by Ichikawa and Jarvis. The two accounts have a similar structure, but Williamson's posits a counterfactual conditional where Ichikawa and Jarvis' posits a strict conditional. Williamson's motivation is related to the problem of deviant realizations, and Ichikawa and Jarvis propose to take care of this problem by enriching the content of the thought experiment in the way we enrich the content of a text of fiction. However, this sort of enrichment is also compatible with Williamson's account. I then consider a different view, defended by Malmgren, on which a complex possibility claim exhausts our reasoning on typical thought experiments. I argue that this account, leaving out a conditional, fails to represent an important part of our reasoning with thought experiments. This is brought out by reflection on the relationship between thought experiments and similar actual cases and by reflection on the requirement, formulated by Malmgren herself, that our reasoning should have an adequate level of generality.

Keywords Thought experiments · Logical form · Conditionals · Williamson · Ichikawa and Jarvis · Malmgren

✉ Daniele Sgaravatti
daniele_sgaravatti@yahoo.it

[1] Arché, Philoosophical Research Centre, University of St Andrews, St Andrews, UK

1 Introduction

In the present work, I am going to defend the view that any good account of the logical form of thought experiments should contain a conditional. Moreover, there are some reasons to think it should be a counterfactual conditional. For expository reasons, I will defend the latter claim first.

Taking a step back, one way to put the general question I am going to discuss is: What is the best way of representing the logical form of thought experiments? But that question is not as clear as one might wish. In particular, it is not clear from what point of view we are to evaluate different representations of logical form. Are we trying to describe the way subjects *actually* think, at some level, when they make use of thought experiments? Or are we trying to describe the way subjects *should* think, at some level, when they entertain thought experiments? I take the task to be nearer to the latter. Of course, considerations related to human psychology will still not be irrelevant, assuming ought implies can.

I will not attempt to survey all the views of the logical form of thought experiments in the literature.[1] I will start by looking at two recent proposals by Williamson and by Ichikawa and Jarvis. I will defend Williamson's view against some objections raised against it.

In the following section, I will consider a rather different view advanced by Malmgren. I will argue that it has very serious shortcomings. In light of these shortcomings,

[1] Sorensen (1992; chap. 6) and Häggqvist (2009) contain two other well-developed proposals. They add more structure to represent the relation between the thought experiment and the theory it is supposed to test, an aspect I will not be concerned with here. Both agree, however, with Williamson's proposal in positing a counterfactual conditional. Therefore, the present work could be read as a defence of (this aspect of) those proposals as well.

the similarities between the two views considered before will turn out to be more important than their differences. In particular, it will appear crucial that both views include the presence of a (not merely material) conditional in the logical form of the thought experiment.

I will end by arguing that the shortcomings of Malmgren's proposal are instructive and can be related to a requirement for any theory of thought experiments formulated by Malmgren herself.

2 Two Conditional Views

Since much of the discussion has focused on one specific kind of thought experiments, the Gettier cases, it will be useful here to focus on those as well. There is no assumption that all different thought experiments, or even just philosophical thought experiments, will fit the proposed accounts without modifications. However, an assumption I am making (together with most people in this debate) is that, in at least a significant amount of cases, there will be enough similarity for the analysis of Gettier cases to provide some insight on the logical form of the reasoning involved in using thought experiments. Horvath (2015) puts the point nicely by saying that Gettier cases in philosophical methodology are analogous to the fruit fly *Drosophila melanogaster* in genetics.

A Gettier case is here defined as a short text intended by its author to provide a possible case of justified true belief that falls short of knowledge. Here is an example to which we will come back in what follows:

> Tom knows that Serena Williams is playing the final of Wimbledon. He turns on the TV and sees her hitting an ace on the match point of a game in the central court of Wimbledon. Tom therefore forms the belief that Serena Williams has won the title this year. In fact, Serena Williams has won, but because of a weird technical problem the game shown on TV was not this year's final but last year's final.[2]

Informally, it seems quite clear what is going on when we consider such a case as a counterexample to the theory that knowledge is justified true belief (henceforth, the KJTB theory or KJTB for short). We are supposed to reason along the following lines. Tom believes that Serena Williams has won Wimbledon. Tom's belief is justified; he has no reason to suspect that the game shown is not live, and everything he has seen is consistent with his information. His belief is also true. Yet, he does not know. Therefore, KJTB is false (as Gettier noted, justified true belief might still be necessary for knowledge, but it is not sufficient).

Nevertheless, there is widespread controversy on what is the best way to represent precisely what is going on here. I will consider, to start with, the debate between Williamson (2007) and Ichikawa and Jarvis (2009). All parties to this debate agree on a number of things. They agree, for example, that reasoning based on a thought experiment can be represented in argument form. This is a rather substantial assumption. However, it should be noted that there is no assumption that the only thing which is interesting about thought experiments is their use in arguments. As we will see, there are also some shared assumptions about logical form. To represent the use of a Gettier case in argument form, let us use the following abbreviations:

> GC$x,p$ = x is related to proposition p as described in the text of the Gettier case
> JTBx,p = x has a justified true belief that p
> Kx,p = x knows that p.

We can now represent the claim that this thought experiment is supposed to disprove in quantified modal logic, as follows:

> C: $\Box\forall x\forall p$ (JTB$x,p \leftrightarrow$ Kx,p) [Necessarily, for all x and p, x has justified true belief that p if and only if x knows that p].

Moreover, it is agreed that one of the premises will say that the case described is possible. We can represent it as

1. $\Diamond\exists x\exists p$ (GC$x,p$) [It is possible that some x stands to some p in the relation described by the Gettier text].

Now, Williamson (2007) argues that the second premise cannot be

2. $\Box(\forall x\forall p$ GC$x,p \rightarrow$ (JTB$x,p \wedge\sim$Kx,p)) [Necessarily, if any x is related to any p as described in the text of the Gettier case, then x has justified true belief that p and x does not know that p].

For, although 1 and 2 clearly entail the negation of C, he thinks that 2 is false. This is because the story described in the above text could be realized in a 'deviant' way. For example, suppose that someone is in Tom's situation, but he has a friend who was watching the match live in Wimbledon and that friend has called him and told him that Serena Williams has won. Then his belief is knowledge after all. Or imagine that he knows who is playing against Serena Williams in the final and that is not the same player he is seeing (the one who was in last year's final), but he

[2] The case is a variation of one created by Jonathan Dancy (1985, p. 25).

disregards this incongruence. Then his belief perhaps is not justified after all. Williamson claims that the ways things could go wrong, for one who is arguing against C, are so numerous that it would be impossible to present a variant of the story that prevents them all.[3] But it would also be unnecessary because we can use the following counterfactual as a second premise:

2* $(\exists x \exists p\ \mathrm{GC}x,p)\ \Box\!\!\rightarrow\ (\forall x \forall p\ (\mathrm{GC}x,p\ \rightarrow (\ \mathrm{JTB}x,p \wedge \sim \mathrm{K}x,p))$.

This is Williamson's preferred formalization of the English counterfactual: 'If a thinker were Gettier-related to a proposition, he/she would have justified true belief in it without knowledge'. The argument from 1 and 2* to the negation of C is valid, and it avoids the problem mentioned for 2 because, roughly, for the counterfactual to be true it is not necessary that the consequent is true at all possible worlds in which the antecedent is true, but it is sufficient that the consequent is true at all *close* possible worlds in which the antecedent is true; and deviant realizations will typically (though not always, as we will see) occur at distant possible worlds.

Ichikawa and Jarvis defend the aptness of (something close to) the original premise 2. They claim that its apparent inadequacy is due to an incorrect way of thinking about the story presented in the Gettier case. We should not identify the story with the text; rather, the story is what we get when the text has been interpreted in the way we typically interpret works of fiction. They also argue against Williamson's alternative proposal. I will discuss two objections that they raise against Williamson, and I will argue that they both fail. I will then briefly go back to consider the proposal advanced by Ichikawa and Jarvis.

The first objection that Ichikawa and Jarvis raise is that Williamson's way of representing the argument makes it a posteriori, for evaluating a counterfactual such as 2* requires one to employ various pieces of world knowledge which are supposed to be empirically justified, if justified at all. As a result, the crucial premise would not come out a priori, and they regard this as a disadvantage. However, Williamson has objections to the epistemological significance of the a priori/a posteriori distinction which are independently motivated (Williamson 2007, ch. 6, 2013). But taking for granted the distinction, and taking for granted we wish that premise to come out a priori, Malmgren (2011) argues convincingly that this is compatible with the premise being a counterfactual: it might be that counterfactuals involved in philosophical thought experiments involve a specific sort of cognitive capacity which gives a priori justification (Malmgren 2011, pp. 307–319). So I will leave aside this objection, since it seems to depend on large issues that are not directly dependent on logical form. The second objection they raise is more interesting in the present context. The objection is that premises of the form of 2* will sometimes be false for intuitively irrelevant reasons (exactly how often is not completely clear). They will be false when the nearest (in terms of possible worlds) realization of the Gettier text, which is the antecedent of the counterfactual, is one which falsifies the consequent. In particular, this will happen whenever the Gettier text is actually true of someone, but it is not actually true that the person lacks knowledge, or it is not actually true that she has justified true belief. In other words, they will be false when a deviant realization of the Gettier case happens to be actual. See Ichikawa's (2009) example:

> Suppose that one's thought experiment is given thus:
> At 8:28, somebody looked at a clock to see what time it was. The clock was broken; it had stopped exactly twenty-four hours previously. The subject believed, on the basis of the clock's reading, that it was 8:28.
> This should be recognizable as a prototypical Gettier description.
> Now consider a world in which that description is true, but where the subject knew in advance that the clock had stopped exactly twenty-four hours previously. In that world, the Gettier text is true but misleading: its subject knows. So (the relevant counterfactual) is false in that world [footnote attached: this is so on a standard Lewis-Stalnaker semantics for counterfactual conditionals, and on any account on which A $\wedge\sim$C entails the falsity of A $\Box\!\!\rightarrow$ C]. Someone running the Gettier argument in that world, then, relies on a falsehood, even if he is innocently ignorant of the person who happens to render his counterfactual false. Relatedly, in running the Gettier argument, one commits oneself to being in a world not positioned in a way that falsifies [the relevant counterfactual]. I take these implications to be implausible. (Does one fail to know the Gettier conclusion by virtue of there being someone in his world who satisfies the text in the wrong way?) (Ichikawa 2009, p. 437)

As Ichikawa (2009) notes, Williamson (2007) already contains some discussion of this kind of objection. Williamson has two lines of reply. He says that the quantifier in the antecedent of 2* might be restricted by the context so that it excludes some cases from the domain of quantification, even though they are actual or occur in nearby worlds. I will develop a somewhat similar suggestion below. Still, Williamson admits that in some cases the counterfactual

[3] Grundmann and Horvath (2014) challenge this claim and therefore defend the aptness of a necessary conditional. I have no space here to discuss their proposal properly, so I am merely assuming, for the present discussion, that their proposal does not work.

will turn out to be false. He argues that this is not a problem for his view. He claims that when someone is shown that the text of a thought experiment is actually realized in such a way that it makes her intuitive judgment false, the correct reaction is simply to modify the example to take care of the problem. Williamson concedes that we might often be tempted to insist that this is not necessary, but he puts that down to a general tendency to fail to admit mistakes. That we have this tendency is certainly true, but something more needs to be said. Williamson also gives an example that suggests an additional reply:

> ..suppose that someone says 'Every man in the room is wearing a tie'; I look around, see a man not wearing a tie, misidentify him as Dave (who is in fact wearing a tie), and say 'Dave isn't'. When it is pointed out to me that Dave is wearing a tie, I deceive myself if I insist that my original reply was correct because the man whom I had in mind was not wearing a tie; that was just not the 'counterexample' which I actually presented. I spoke falsely when I said 'Dave isn't'. (Williamson 2007, p. 201)

The example not only suggests a way in which one might be tempted to resist the need to modify one's claims but also illustrates, I think, a certain sort of value that there can be in a false claim. The subject's claim in the example can be informative, drawing the attention of the party to the relevant counterexample to the generalization under discussion, even though it is false. As a result of the false claim that Dave is not wearing a tie, people may realize that someone else, the man I point to, is not wearing a tie. We should say something similar, I will argue, for Gettier cases which are realized in an unintended way. I will argue that there is plenty of room to explain why the judgements about those cases still have a lot of epistemic value and why we are reluctant to think they are false. To illustrate this point, it will be useful to consider two more thought experiments about thought experiments I devised:

Case 1 Suppose a subject, call him John, has his first encounter with a Gettier case reading Williamson's description of a real-life Gettier case. Williamson claims that he gave his students the false information that the only PowerPoint presentation he has given in his life was a failure, while in fact he never gave a PowerPoint presentation, successful or otherwise. He then made sure that the students could clearly see that if his only PowerPoint presentation was a failure, then he has never given a successful PowerPoint presentation.[4] On the basis of Williamson's testimony John judges that the "victims" of this machination did not know that Williamson has never given a successful PowerPoint presentation, although they had a justified true belief to that effect. However, suppose also that Williamson's description of the actual Gettier case is mistaken in the following way: the day before the lecture, Williamson had told of his intention to create a real-life Gettier case to a colleague; unbeknownst to Williamson, the colleague told of this plan to some of the students, and the whole plan, with its details, became common knowledge. So after all the students knew that Williamson has never given a successful PowerPoint presentation.

Case 2 Suppose a subject, call her Jane, is introduced to Gettier cases by the following example: Smith is told by an apparently reliable and honest mathematician that Fermat's last theorem is false. The mathematician actually wishes to deceive Smith. But at the same time, another mathematician, unbeknownst to both Smith and his informant, has just proven that Fermat's last theorem is false. Jane also mistakenly believes, based on an apparently reliable testimony, that Fermat's last theorem is false. She judges that Smith, in the story, has a justified true belief which is not knowledge.

There are four points which I want to make about these two cases. I take all four points to be entirely uncontroversial, but it is important to keep them in mind.

1. Everyone, or at least anyone who thinks thought experiments can figure in an argument, should admit that John and Jane form some false beliefs and make use of them in their reasoning. John has a false belief about the way Williamson's students formed a certain belief, and Jane has a false belief about the possibility of someone truly (and justifiably) believing the negation of Fermat's last theorem.
2. It seems both John and Jane are justified in their false beliefs. Therefore, if they go on to infer that the KJTB theory is false they will form a justified true belief (effectively, they might be Gettiered, forming a justified true belief which falls short of knowledge that the KJTB theory is false; of course, I am assuming KJTB is false). A justified true belief, even if it falls short of knowledge, has epistemically valuable features.
3. It is actually controversial, in contemporary epistemology, whether one cannot gain knowledge by inferring from a justified false belief. [5] Therefore, it is not obvious that John and Jane could not come to know that the KJTB theory is false on the basis of those counterex-

[4] Williamson (2007), p. 192.

[5] See, for example, Unger (1968), Klein (1996, 2008), Hawthorne (2004, p. 57) and Coffman (2008, pp. 188–194).

amples involving falsehoods. In particular, it is often held that one can gain knowledge by inferring from a (justified) false premise when, if one were to realize that the premise is false, one could easily replace it with a true (and justified) one. The next point should suggest that John and Jane might meet this condition.

4. Last but not least, it is clear that John and Jane, supposing they are willing to spend time thinking about the issue, can easily come to generate more Gettier cases not involving falsehoods and thereby come to know (if they did not already) that KJTB is false. If this happens, the initial cases will play a crucial role in the production of such knowledge. The cases will have a cognitive value for John and Jane, allowing them to see the structure of a possible different case.

Given that these points about the two cases I provided are to be conceded, it seems they must be conceded also to the defender of Williamson's account about the Gettier cases in which, according to such an account, our judgement is false. In particular, points 2–4 seem to give a rich account of what is still valuable about these 'deviant' cases, which helps to explain our reluctance to think that the deviant realization matters at all. The counterfactual is false because its antecedent is true and the consequent false, but we have justification to think it is true. Moreover, we could easily replace the premise with one consisting of a similar but true counterfactual. Of course, Ichikawa and Jarvis will still want to insist that, in such cases, there is literally a counterexample and no false belief is involved. But it is hard to see how this insistence is motivated. At one point, Ichikawa (2009) indicates the disagreement as one about what someone commits herself to when she says or thinks '"the subject has justified true belief but does not know", i.e., when she expresses the Gettier intuition' (p. 438). The foregoing considerations helped to explain some good features of that judgment. But whether the judgment is literally true depends, of course, on what particular Gettier case the person is thinking about. The matter cannot be less complex than providing a content for the definite description 'the subject' involved in the sentence 'the subject has justified true belief but does not know'. Clearly, we are talking about the subject described in a certain text. According to Williamson's view, if there is a subject in the actual world satisfying every element of the text, the description refers to her. Everyone agrees that in some sense the presence of this actual case is irrelevant. But there are theoretical reasons to think it should matter to the literal meaning of 'the subject'. According to the view defended by Ichikawa and Jarvis, however, one gets the content of 'the subject' right only if one enriches the literal meaning of the description of the scenario in significant ways. Whether we should stick to the letter of the text, as Williamson suggests, or should enrich it is not a straightforward pre-theoretical matter.

Let me move to consider briefly some problems for the view defended by Ichikawa and Jarvis (partly drawing from the discussion in Williamson 2009). The story we build out of the text, in order to make the strict conditional true, has to be extremely rich, for the number of situations which are compatible with the text but do not support the consequent is extremely large. However, having this extremely rich story in the content of the judgment will present a number of drawbacks. First, it will turn out that the cases differ quite a lot between any two subjects considering the thought experiment, since there are many different reasonable but incompatible ways of enriching the scenario. Sometimes the differences might matter to the philosophical issues involved. Even if they do not matter, it seems that it becomes surprisingly difficult for two people to consider the same thought experiment and even more difficult to know that they do. This is certainly counterintuitive because we say of two people that have read, for example, Gettier's paper, that they have considered the same thought experiment. Second, even for a single subject, there might be different ways she is disposed to enrich the story in response to different questions and in different contexts, so the content of the case might be indeterminate, even as interpreted by a single reader. Again, it might therefore be indeterminate whether the thought experiment serves its purpose. Third, the first premise of the argument will become very hard to know, for there might be hidden inconsistencies in the story (this point is also raised in Malmgren 2011, pp. 305–6). All these consequences seem very far from philosophical practice; the advantage of using thought experiments over real cases is often that they present us with short, simple stories in which all the relevant factors can be surveyed.

The problems I sketched here might be not insurmountable.[6] However, if they have any force, Williamson has a strong dialectical advantage. The advantage is given by the fact that most or all the difficulties for the suggestion that a thought experiment is a sort of fiction would be assuaged if we did not require the way we enrich the text to be strong enough to sustain a *strict* conditional. It is very plausible that the scenario we consider is not completely expressed by the text of the thought experiment. But accepting this insight is compatible with Williamson's view of the logical form of the reasoning typically applied to the scenario itself. We might enrich the text of our Gettier case somewhat, as we do with fiction, and then employ a counterfactual premise, saying roughly that if the story were true then there would be a case of justified true belief which is not

[6] Cfr. Ichikawa and Jarvis (2013, pp. 209–13).

knowledge (and of course we might do the same, *mutatis mutandis*, for other thought experiments). In comparing the two proposals, we should keep in mind that the strict conditional entails the counterfactual conditional, and it seems safer to employ a premise which is weaker, given that we are reaching the same conclusion. Of course, if we know that the stronger premise is true, then we might as well use it. But there is no guarantee that the enriched story can sustain a strict conditional, since we cannot simply intend our story to be such that no deviant realizations are allowed, unless we can give an informative characterization of 'deviant case'. However, the suggestion that the thought experiment be treated as a piece of fiction and its text supplemented accordingly can actually help Williamson's account to get rid of at least some of the problem cases. At least some actual deviant realization can plausibly be excluded by an appropriate interpretation of the text. This should be judged case by case, if for some reason it became relevant.

It seems enough has been said by now to warrant taking Williamson's view of the logical structure of the reasoning involved in (at least some) thought experiments as a preferred option, supposing the choice is restricted to the views considered here.

3 The Possible, the Actual and the Conditional

Malmgren (2011) advances a view of the logical form of thought experiments very different from both the views we have considered in the foregoing and, as far as I know, from any other view defended before. In Malmgren's view, the content of our judgement on a Gettier case should not be expressed as a conditional. Rather, the content is a complex possibility claim that she expresses as follows: 'It is possible that someone stands to p as in the Gettier case (as described) and that she has a justified true belief that p but does not know that p' (Malmgren 2011, p. 281). She also suggests the following formalization (the interpretation of the non-logical symbols is the same we are using):

Possibility: $\Diamond \exists x \exists p(\mathrm{GC}(x,p) \wedge \mathrm{JTB}(x,p) \wedge \sim \mathrm{K}(x,p))$.

The reasoning we need here is of course very simple to get to the conclusion. We simply need to carry out conjunction elimination on the instantiation of the existential quantifiers. I will argue that this model is too simple. It leaves out an important part of the structure of our reasoning about the case, namely the way in which our judgment that someone could have a justified true belief that falls short of knowledge depends on our judgement that someone could be in the situation described by the Gettier case. It is *because* we see that in our story Tom has a justified true belief which falls short of knowledge *and* we believe that the story describes a possible situation that we come to believe it is possible to have a justified true belief falling short of knowledge. I will articulate this criticism in three steps. First, I will advance some claims about the parallel between thought experiments and actual cases, and I will claim Malmgren's account has difficulty in accounting for some of them. Second, I will consider Malmgren's reply to the general worry that her account misrepresents the structure of our reasoning. Third, I will draw a connection between the shortcomings of Malmgren's account and a requirement she discusses in the same work on any adequate account—what she calls the requirement of implicit generality. In developing the second and third points, I will also compare my criticisms of Malmgren with similar (but, I believe, different) ones advanced in Horvath (2015).

The comparison between thought experiments and actual cases is illuminating, I think, in various ways; therefore, I will digress here to make some remarks on this matter. It is not hard to build actual Gettier cases; Williamson, as mentioned above, claims to have done so more than once, and there is no reason to think he actually failed to do it. Other thought experiments differ, but in this matter the difference is often superficial. There are often actual counterparts, so to speak, to many well-known thought experiments. I will just mention a few. Kripke claims that his Gödel–Schmidt case probably has a real-life counterpart in Peano and Dedekind. Trolley cases certainly cannot be constructed on purpose, but real life unfortunately provides similar situations. Many cases on causation discussed in the literature are absolutely realistic; it is hard for two people to throw a rock at a bottle so that the bottle would break if either rock hit it but not so hard as to be practically impossible. Often, philosophers look at hypothetical cases not because no actual case would serve their purpose but simply because it is quicker to present a hypothetical case or because a hypothetical case might be neater, screening out various kinds of noise coming from the actual ones. Think, for example, of Goldman's 'fake barn county' case or Burge's arthritis case.

The point I wish to make here is about the relation between judgements about hypothetical cases and judgements about actual situations which are identical or relevantly similar. Clearly, these judgements will have to coincide; if I judge that in a hypothetical case a certain causal relation holds or that a certain term refers to something or that a certain action is permissible, then I should give the same judgement about a corresponding actual case. In Gettier cases, we will make the judgement that someone in the described scenario lacks knowledge of a certain proposition, although they have a justified true belief. If the scenario is merely possible, then we will conclude that it is possible for someone to have justified true belief falling short of knowledge. However, if the scenario is actual, it would be peculiar and slightly irrational not to reach the

stronger conclusion that someone has a justified true belief falling short of knowledge. All this is predicted by the idea that we reason through some kind of conditional linking the description of the case and the obtaining of a certain state of affairs. So the idea of a conditional logical form offers a simple explanation here. What are the alternative explanations? Of course, Malmgren can say that whatever competence grounds the transition from the scenario to a possibility judgement also grounds the transition from the actual case to the actuality judgement. But the question is why it is always so. Why could not a competence sometimes allow one transition but not the other? I see no hope for a good answer here, unless we turn our attention to logical form.

It might be useful here to compare the way we reach the conditional judgement involved in hypothetical reasoning with the way the proof of a conditional proceeds in natural deduction.[7] In attempting to prove a conditional, one assumes the antecedent, temporarily treating it as proved, and then one attempts to prove the consequent using the antecedent together with all the resources allowed by the logic (and the previous stages of the overall proof, if there are any). If one succeeds in thus proving the consequent, one has proved the conditional, and the proof of the conditional does not rely on the assumption of the antecedent. The point of the analogy, for my present purposes, is that there is no difference between the sub-proof and a proof which could be conducted if the antecedent had been in fact proved, although of course the latter would then constitute a proof of the consequent and not of the conditional. Similarly, there is a stage in our reasoning about hypothetical scenarios which seems to be independent of whether the scenario is merely hypothetical or rather actual.[8] Parallel conclusions follow in the two cases.

Malmgren, however, claims that her account—although it does not contain a conditional—has the resources to articulate the structure of the reasoning involved in thought experiments. If this were correct, the account could probably also explain the connection between actual and hypothetical cases. The discussion of the problem occurs in a long footnote, which I will quote in full:

> On my account, the claim that a subject is related to a proposition as stipulated in the case description is, not surprisingly, paraphrased as the claim that someone could stand to a proposition in that way. [In symbols: ∃x∃pGC(x,p)]. One might wonder how this could be among our reasons for believing POSSIBILITY—how a claim of the form 'possibly p' could be a reason for believing a claim of the form 'possibly p & q'. It may even be argued that I am proposing an absurd justificatory structure. (Compare: How could the claim that someone broke in last night be a reason to believe that someone broke in last night and stole all the silver?) In reply I admit that the structure looks a bit odd, but it is not absurd. And we can relieve some of the oddness. Recall that there are different ways of expressing the candidate content in English (see n. 32); e.g. as the claim that someone who stands to a proposition as stipulated in the case description has a justified true belief but does not know. (Compare: Bettie believes that the person who—or whoever—broke in last night stole all the silver in part on the grounds that someone broke in last night. We can easily tell a story in which this is true—e.g. suppose Bettie knows that all her silver is easily accessible and desirable to thieves; this justifies her in believing that if someone broke in they would steal all the silver; that—together with her justified belief that someone broke in last night—justifies her in believing that whoever broke in last night stole all the silver). (Malmgren 2011, pp. 296–7, fn. 53).

Malmgren is claiming that the possibility of the Gettier case is the grounds for the possibility of a counterexample to the KJTB theory. This is agreed by all parties, as far as I can see. The point is whether she is in a position to make this claim, given that she does not offer any representation of the link between the contents of the two claims. The example she considers to respond to the worry seems to offer little help, and in fact it might even be used to illustrate the worry. It is telling, it seems to me, that in looking for a case in which the justificatory structure she proposes looks plausible, Malmgren *herself* reaches for a conditional ('if someone broke in they would steal all the silver'). Of course, we can move from p to p ∧ q when a conditional bringing from p to q is in place. When the conditional is not, at least implicitly, part of our reasoning, the move is often absurd or at least unjustified. In those cases, if we leave that conditional out of the account of the logical form of our reasoning, we leave out a crucial premise.

To be clear, I am not claiming that arguments of the form 'p, therefore p and q' are never rationally acceptable

[7] I owe the analogy to Magdalena Balcerak Jackson.

[8] The analogy between thought experiments and conditionals could be deepened and developed in different ways depending on one's views on the logic and semantics of conditionals, a topic obviously too vast to be addressed here. One example will perhaps be useful. Horvath (2015) develops (what he calls) a *suppositional* view of thought experiments. But in some views, supposition is also a crucial aspect of the semantic of conditional statements (see Edginton 2014 for a rapid but masterful introduction to some of the complex issues surrounding conditionals, with particular emphasis on suppositional theories). Even in different views, supposition might plausibly turn out to be crucial to *reasoning with* conditionals.

unless they are enthymematic (although a previous version of this paper plausibly suggested just that).[9] It might be that lexical or logical features of p make the transition compelling. A referee offers the following example: 'Possibly, Peter is a bachelor, therefore, possibly, Peter is a bachelor and unmarried'.[10] Now, while this is a somewhat odd piece of reasoning, I agree that one does not need to make use of the implicit premise that if Peter is bachelor then Peter is unmarried in order to rationally go through that inference (even if the conditional needs to be true in order for the inference to be truth-preserving). The issue is whether philosophical thought experiments, such as Gettier cases, can be reconstructed according to this model. And it does not seem to be the case. The reason one does not need a conditional in the inference from '(possibly) bachelor' to '(possibly) bachelor and unmarried', it seems to me, is that the inference is extremely simple. The link between p and q is so obvious that it does not need to be formulated, even implicitly. However, the sort of reasoning competence sustaining the inference could also sustain acceptance of the conditional 'if someone is a bachelor, he is unmarried'. The conditional would be close to the most natural expression of that competence, which is shown in the acceptance of 'a bachelor is an unmarried male'[11] (a statement which can plausibly be given a conditional logical form). But one need not go through the most natural steps, since the reasoning is so trivial. When instead p and q involve considerable internal complexity and their connection can be hard to see, the discovery of such connection constitutes a crucial reasoning step. In those cases, without any representation of that step, the inference is not rationally explained. It might seem to us that Gettier cases are in some ways trivial, but clearly the matter was not trivial before the publication of the original Gettier paper,[12] and, importantly, in the following debate it was also controversial which feature of the described cases was relevant. I will come back to this point below. More generally, philosophical thought experiments can certainly be interestingly complex. It is an important part of the use of a thought experiment to be able to isolate a relevant part of the scenario and its consequences. Again, Malmgren's analysis is too thin to allow a reconstruction of this sort of reasoning.

As anticipated, I should compare the criticisms advanced here to some which have already appeared in print and in particular to the ones in Horvath (2015). Horvath usefully individuates three kinds of adequacy requirements for an account of thought experiments—methodological, psychological and epistemic adequacy. Roughly, methodological adequacy is a matter of being sufficiently faithful to philosophical practice; psychological adequacy is a matter of being sufficiently realistic from, unsurprisingly, a psychological point of view and epistemic adequacy is a matter of the attempted reconstruction to yield knowledge or at least justified belief. Horvath criticizes Malmgren on the grounds of psychological and methodological inadequacy. However, and this is a first difference, I believe this is a mistake. If these were the only defects of the account, Malmgren could reply by saying that her account was only meant to achieve epistemic adequacy and idealize away, so to speak, certain aspects of our psychological and disciplinary reality. As noted in the introduction, I take the main aim of offering a reconstruction of the logical form of thought experiments to be epistemological. But I believe the criticism we are making is epistemic, or at least it ties together the epistemic aspect and the other two aspects. The problem is not just that we plausibly do not reason in the way described by Malmgren but also that if we did, we would not plausibly gain knowledge of the conclusion. As for the methodological adequacy of Malmgren's account, Horvath rightly observes that one could accept *Possibility* even if one did not find the Gettier case referred to by our story to be the basis of a successful counterexample to KJTB; it is enough that one finds the situation described by the story possible and has some other reason to deny that justified true belief is necessarily knowledge. My further (broadly) methodological criticism is that Malmgren's reconstruction does not harmonize with what Malmgren herself calls the requirement of 'implicit generality'. This will be the topic of my last few paragraphs.

The basic thought of the requirement of implicit generality is that if our judgements on a thought experiment are to be rational (and if they are to constitute knowledge), they must be stable across a wide range of irrelevant variations. Consider again our initial Gettier case. Nothing is said in the text about Tom's age, the kind of TV he is using, whether he is at home or somewhere else, etc. The success of the thought experiment as a counterexample to KJTB is

[9] Thanks to two anonymous reviewers for *Topoi* here.

[10] A related example, suggested by a second referee, is the following: this is red, therefore, this is coloured. The general point here is that an inference can be compelling even if it does not instantiate a valid logical form. That, of course, is correct, but nothing I say here suggests otherwise. Inferences of the form 'p, therefore p and q' of course are invalid, but this is not the reason they can be problematic.

[11] Actually, if this is meant as a definition, it is not complete; a priest is not a bachelor, for example. So something like 'eligible for marriage' should be added. This is a familiar point; see for example Harman (1996) p. 398.

[12] Gettier (1963). Interestingly, in 1948 Russell considered in print the case of someone looking at a stopped watch coincidentally indicating the right time. He presented it (correctly, as far as this goes) as a counterexample to the view that knowledge is just true belief. But it seems that it did not occur to him that it could be considered as a counterexample to the KJTB theory. The connection was not obvious to him, although he certainly had the intellectual skills and the knowledge of epistemology which would make it intelligible (Russell 2009, p. 91).

compatible with a vast range of different scenarios in these respects. It is also clear that it is compatible with many variations of the text itself. It would not change anything for our purposes if instead of Tom the main character of the story was called Rafa, though this might bring different associations to mind. It also would not make a difference if the case was about a different tennis player (as in the original case created by Dancy; see fn. 2). We could also build similar cases using different sports or other activities. Some modifications of course would not be irrelevant, and some would create what we call deviant realizations. The point is that a subject using the thought experiment can derive no knowledge from it unless they are, at least to a good degree, capable of distinguishing irrelevant variations from relevant ones. Suppose I judge that Tom knows when it is stipulated that he is 45 years old, but I instead claim that he does not know if Tom in the story is 60 years old. Or suppose I only attribute knowledge to Tom if it is added that he has not red hair. In those cases, I am clearly not competent to make judgments about the thought experiment.

What, if anything, has the requirement of implicit generality to do with the logical form of thought experiments? It seems to be a requirement on the subject who employs the thought experiment in reasoning, namely the requirement of possessing a certain kind of competence. Moreover, as Malmgren notes, the sort of competence we are talking about is not, or at least not obviously, limited to thought experiments. In fact, it seems to be the competence to judge (relevantly) similar cases alike. I think this much is correct, and we should not ask the logical form to explain what sort of competence is involved.[13] Malmgren also claims that her view is 'compatible' with the generality requirement. This is also true, in my view, but it is not sufficient. Malmgren can of course appeal to some notion of conceptual competence, but so can the proponent of some conditional account. But Malmgren's account, leaving out a conditional step in our reasoning, does nothing to specify the point at which the competence is required; it does not provide, so to speak, a focus for the evaluation of the competence. As Malmgren notes, the generality requirement is related to the problem of deviant realizations. One needs to distinguish irrelevant variations from variations which create deviant cases. When the case is deviant, crucially, the conditional premise will be false. It will not be the case that if a subject were related to a proposition as described, they would have a justified true belief falling short of knowledge; a fortiori, the strict conditional would be false too. This is related to the point mentioned above about the complex structure of Gettier cases. Consider Gettier's original cases. Is it because Smith forms his justified belief using a false premise that it is not knowledge? Or is it because he would form it in the same way even if it were false or something else? These different diagnoses entail different verdicts on different kinds of Gettier case; for example, the view that it is inference from a false premise that explains the lack of knowledge in the original Gettier case seems to fail to predict that the case I presented at the beginning is a genuine counterexample to KJTB. It fails to predict the correctness of conditional premises like 2 or 2*. Of course, the view that the use of false premises is crucial to at least some Gettier cases is still not entirely unreasonable. However, if I am convinced by the case only if Tom has red hair, I will reject the relevant conditional if the description of the case in the antecedent contradicts this assumption. My complete lack of competence will show in my disposition to reject that conditional. I will judge, for example, that the case would be possible if Tom had red hair, but he would not be justified. In other words, we can investigate the nature and the functioning of my competence by looking at my disposition to accept or deny the relevant conditionals. So Malmgren's view is at disadvantage in this respect in relation to both views considered in the previous section. By leaving out the conditional, her view fails to provide material useful to reconstruct how we may meet the generality requirement, and in fact there is no interesting relation at all between the logical form and the requirement. If an account of the logical form of thought experiments does not help in reconstructing the justificatory structure that is involved in their use, I believe it loses its main purpose.

[13] I attempt to explain the nature of the relevant competence in Sgaravatti (2015).

4 Conclusion

I defended Williamson's account of the logical form of thought experiments against a competing account offered by Ichikawa and Jarvis. The two accounts have a similar structure, but Williamson's posits a counterfactual conditional where Ichikawa and Jarvis posit a strict conditional. Williamson's motivation is related to the problem of deviant realizations, and Ichikawa and Jarvis propose to take care of this problem by enriching the content of the thought experiment in the way we enrich the content of a text of fiction. However, this sort of enrichment is also compatible with Williamson's account. The main objection against Williamsons' counterfactual account is based on its allegedly counterintuitive consequences in some cases in which there is an actual realization of the text of the thought experiment. However, assuming there are such cases, I have argued that there are several reasons why we could give the relevant thought experiments very high epistemic value, although reasoning with them might include a false premise.

I then considered a different view, defended by Malmgren, on which a complex possibility claim exhausts our reasoning on typical thought experiments. I argued that this account, leaving out a conditional, fails to represent an important part of our reasoning with thought experiments, and this is brought out by reflection on the relation between thought experiments and similar actual cases and by reflection on the requirement, formulated by Malmgren herself, that our reasoning should have an adequate level of generality.

References

Coffman EJ (2008) Warrant without truth? Synthese 162:173–194

Dancy J (1985) An introduction to contemporary epistemology. Basil Blackwell, Oxford

Edginton D (2014) [2001] Indicative conditionals, stanford encyclopaedia of philosophy. http://plato.stanford.edu/entries/conditionals

Gettier E (1963) Is justified true belief knowledge? Analysis 23:121–123

Grundmann T, Horvath J (2014) Gettier cases and the problem of deviant realizations. Philos Stud 170:525–533

Häggqvist S (2009) A model for thought experiments. Can J Philos 39:55–76

Harman G (1996) Analyticity regained? Nous 30:392–400

Hawthorne J (2004) Knowledge and lotteries. Oxford University Press, Oxford

Horvath J (2015) Thought experiments and experimental philosophy. In: Daly C (ed) The Palgrave handbook of philosophical methods. Palgrave Macmillan, New York, pp 386–418

Ichikawa J (2009) Knowing the intuition and knowing the counterfactual. Philos Stud 145:435–443

Ichikawa J, Jarvis B (2009) Thought-experiment intuitions and truth in fiction. Philos Stud 142:221–246

Ichikawa J, Jarvis B (2013) The rules of thought. Oxford University Press, Oxford

Klein P (1996) Warrant, proper function, reliabilism, and defeasibility. In: Kvanvig J (ed) Warrant in contemporary epistemology. Rowman & Littlefield, Lanham

Klein P (2008) Useful false beliefs. In: Smith Q (ed) Epistemology: new essays. Oxford University Press, Oxford, pp 25–62

Malmgren AS (2011) Rationalism and the content of intuitive judgements. Mind 120:263–327

Russell B (2009) [1948] Human knowledge. Its scope and limits. Routledge, London

Sgaravatti D (2015) Thought experiments, concepts and conceptions. In: Collins J, Fischer E (eds) Experimental philosophy, rationalism and naturalism: rethinking philosophical method. Routledge, London, pp 132–150

Sorensen R (1992) Thought experiments. Oxford University Press, Oxford

Unger P (1968) An analysis of factual knowledge. J Philos 65:157–170

Williamson T (2007) The philosophy of philosophy. Blackwell Publishing, Hoboken

Williamson T (2009) Replies to Ichikawa, Martin and Weinberg. Philos Stud 145:465–476

Williamson T (2013) How deep is the distinction between a priori and a posteriori knowledge? In: Casullo A, Thurow J (eds) The a priori in philosophy. Oxford University Press, Oxford, pp 291–312