

## “Thought-Experiment Intuitions and Truth in Fiction”

Jonathan Ichikawa, Rutgers University

Benjamin Jarvis, Brown University

*Philosophical Studies*, forthcoming. Penultimate draft.

**Abstract.** What sorts of things are the intuitions generated via thought experiment? Timothy Williamson has responded to naturalistic skeptics by arguing that thought-experiment intuitions are judgments of ordinary counterfactuals. On this view, the intuition is naturalistically innocuous, but it has a contingent content and could be known at best *a posteriori*. We suggest an alternative to Williamson’s account, according to which we apprehend thought-experiment intuitions through our grasp on truth in fiction. On our view, intuitions like the Gettier intuition are necessarily true and knowable *a priori*. Our view, like Williamson’s, avoids naturalistic skepticism.

### I. Introduction

We start (where else?) with a story.

Professor Muthos was teaching an epistemology course. “It’s surprisingly difficult,” he said, “to provide an analysis of knowledge.” A student in the back row raised his hand. “Yes, Brian?”

“I don’t see why we should think it’s so hard to provide an analysis of knowledge,” Brian said. “I think I know what knowledge is—it’s justified true belief.” Professor Muthos marveled privately at how conveniently the dialectic was progressing. “Let me tell you a story.” The class leaned forward attentively. “Listen to this, and see if you think that knowledge is justified true belief.

Joe had left his watch at home, and he wanted to know what time it was. He saw a clock on the wall. As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. ‘Good,’ thought Joe, ‘I still have fifteen minutes until Mr. Pumbleton will be expecting me.’ Joe had arranged an important meeting at 10:30.

However, things were not as they appeared, for Joe had been looking at an inaccurate clock—it was 15 minutes slow. It was already 10:30. But fate smiled on Joe that day—for due to a careless error on the part of Mr. Pumbleton’s secretary, Mr. Pumbleton thought Joe’s appointment was at 10:45. So Joe’s belief

about how much time he had until Mr. Pumbleton would consider him late was true after all.

Think about Joe's belief about how much time he had," Professor Muthos suggested. "It was justified, and it was true, but we can see that it was not knowledge. So knowledge isn't justified true belief." Brian and the rest of the class thought about the story and conceded that Professor Muthos was right. Justified true belief wasn't the same as knowledge.

Philosophers often construct thought experiments to test theories; Professor Muthos's presentation is a fairly typical example; this is standard philosophical methodology. Sometimes, such intuitions play significant roles in philosophical analysis. But there is apparent reason to be skeptical about these intuitions; one may worry that to posit a faculty that can accurately judge non-actual thought-experiment situations is unduly mysterious. How could we be justified in trusting mere *intuitions* in these matters? From a naturalistic point of view, to suppose that we have such abilities can look like vain superstition. Consequently, there is now an extensive literature challenging the role of intuitions in philosophical inquiry. Skeptical arguments threaten the whole project of philosophical analysis.<sup>1</sup> If the critics are right, then Professor Muthos is presupposing a problematic philosophical methodology. Far from helping them see into the nature of knowledge, he is confusing and indoctrinating his students with philosophical dogma.

More recently, Timothy Williamson has offered a sort of a defense of Professor Muthos's strategy.<sup>2</sup> On Williamson's view, there is no special faculty of intuition that we must invoke in order to judge that Joe doesn't know. Instead, the 'intuition' we form is just a straightforward counterfactual judgment, no different in kind from the naturalistically innocuous judgments we make every day. (We discuss the details of Williamson's account below.) Although Williamson is aptly read as defending Muthos's invocation of a thought experiment, it is not right to think of Williamson's project as a defense of a traditional understanding of philosophical methodology, for the judgments with which he identifies 'intuitions' are scarcely recognizable as the things traditional philosophers had in mind. In particular, on Williamson's view, the propositional contents of the intuitions in question are contingent. This is problematic, from a traditional standpoint, for two reasons. First, tradition has it that intuitions like the Gettier intuition have necessarily true contents; Williamson's counterfactuals, however, will vary in truth value between possible worlds. Second, the standard view has it that intuitions like the Gettier intuition can be known *a priori*. As we will later explain, Williamson's contingent counterfactuals could be known at best empirically—in our view, a disappointing result.

The skeptics charged that traditional philosophical methodology should be abandoned as superstition; Williamson's formulation made some progress in de-mystifying the intuitive judgments, but left them as mere *a posteriori* counterfactuals. Our project is to find a middle ground for the traditional view of thought experiments and intuitions. We think that Williamson is right that we can give a naturalistically innocuous account of thought-experiment intuitions, and one that draws a close connection between our judgments about thought experiments and our everyday judgments about the world around us, but wrong in thinking that we must give up

---

<sup>1</sup> For a sampling, see the various papers in DePaul & Ramsey (1998).

<sup>2</sup> Williamson (2005) and Williamson (2007).

claims of necessity and *a priori* to do so. On our view, intuitions such as the Gettier intuition are judgments of necessity, and we see no in principle objection to the idea that we can know them *a priori*. We suggest that what is missing from Williamson's view is a proper understanding of the way that we engage with fictional texts. Once this is supplied, we will have no compelling reason to understand intuitions like the Gettier intuition as having merely contingent content.

We will begin with a restatement of Williamson's argument, then attempt to establish a suitable alternative. In the final sections of the paper, we argue that our formulation, like Williamson's, avoids naturalistic skepticism about intuition. We follow Williamson in using the Gettier intuition as a paradigm.

## II. A Williamsonian Argument

How shall we formalize the Gettier argument? Here is a first stab:

$K(x, p)$ :  $x$  knows that  $p$ .

$JTB(x, p)$ :  $x$  has a justified true belief that  $p$ .

$GC(x, p)$ :  $x$  stands to  $p$  as given in the (interpreted) text of the Gettier story<sup>3</sup>

(1)  $\Diamond \exists x \exists p GC(x, p)$

(2<sub>n</sub>)  $\Box \forall x \forall p [GC(x, p) \supset (JTB(x, p) \ \& \ \sim K(x, p))]$

$\therefore$  (3)  $\Diamond \exists x \exists p (JTB(x, p) \ \& \ \sim K(x, p))$

In English:

(1) It's possible for some  $x$  to stand to some  $p$  as given in the text of the Gettier story.

(2<sub>n</sub>) Necessarily, if  $x$  stands to  $p$  as given in the text of the Gettier story, then  $x$  has a justified true belief in  $p$  that isn't knowledge.

Therefore, (3) it's possible for someone to have a justified true belief that isn't knowledge.

In this formulation, (2<sub>n</sub>) represents the Gettier intuition. And this Gettier intuition is a necessity claim, just as traditional philosophers think it is. But Williamson argues that this cannot be the correct formulation of the Gettier intuition. For (2<sub>n</sub>), says he, is false. The key to seeing this is that the texts of Gettier stories are inevitably underspecified. It is possible, consistent with

---

<sup>3</sup> Williamson (2007), p. xxx puts it thus: *Someone could stand in the relation described in the Gettier story to some proposition*. We characterize Williamson's GC relation in terms of *texts* of Gettier stories, which is clearly what Williamson has in mind, because we believe there is an important difference between stories and texts. See section IV below.

the text of the story, to insert pieces that will prevent the story from presenting a case of justified true belief that isn't knowledge. We can describe cases in which the given Gettier text is true, but there is no case of non-knowledge justified true belief (hereafter NKJTB), as **bad** Gettier cases. Gettier cases with NKJTB are **good** ones.

Following are two ways we could understand Professor Muthos's story from our introduction as a bad Gettier case by adding consistent bits to it:

(a) Joe had left his watch at home, and he wanted to know what time it was. He saw a clock on the wall. As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. **Joe ignored the dozens of other clocks decorating the wall, each of which showed a different time.** "Good," thought Joe, "I still have fifteen minutes until Mr. Pumbleton will be expecting me."

(b) As a proficient reader of clocks, Joe had no difficulty in determining that the clock read 10:15. "Good," thought Joe, "I still have fifteen minutes until Mr. Pumbleton will be expecting me." **Joe knew that Mr. Pumbleton was watching him via closed-circuit television and would accurately predict that it would take him fifteen minutes to reach his office from his current location.**

Both (a) and (b) describe bad Gettier cases—in each case, the literal text is satisfied, but Joe does not have NKJTB. In (a), Joe's belief fails to be justified; in (b) his belief qualifies as knowledge. Williamson points out that, given the possibility of bad Gettier cases, (2<sub>n</sub>) cannot be right. Sometimes, when x stands to p as in the Gettier text, there is no NKJTB.

A philosopher's natural move here is to tweak the story. We'll come up with a new Gettier text, and do so with extra care to 'close the loops.' If one is clever and careful enough, one *might* be able to generate texts that are not susceptible to bad satisfaction. Even if this were true, however, it misses the force of Williamson's observation. The problem is not that it is impossible to construct 'invincible' Gettier texts—it's that most of the texts we actually use, *and appropriately use*, are not invincible. Professor Muthos might have been able to construct an airtight story, but he didn't have to. The text given at the start of the paper works perfectly well to generate the Gettier intuition, but the first stab formulation of that intuition, (2<sub>n</sub>), is false, because the text can be satisfied in bad ways.

So (2<sub>n</sub>) must be weakened. Here is Williamson's version of the argument:

(1)  $\Diamond \exists x \exists p \text{ GC}(x, p)$

(2<sub>cf</sub>)  $\exists x \exists p \text{ GC}(x, p) \square \rightarrow \forall x \forall p [\text{GC}(x, p) \supset (\text{JTB}(x, p) \ \& \ \sim \text{K}(x, p))]$

$\therefore$  (3)  $\Diamond \exists x \exists p (\text{JTB}(x, p) \ \& \ \sim \text{K}(x, p))$

In English:

(1) It's possible for some x to stand to some p as given in the text of the Gettier story.

(2<sub>cf</sub>) If some *x* were to stand to some *p* in a way satisfying the (literal interpreted) text of the Gettier story, then anyone who satisfied the text of the story with a proposition would have NKJTB.<sup>4</sup>

(3) So, it's possible to have NKJTB.

This formulation of the Gettier argument, like the first, is valid. It also avoids Williamson's objection to the first argument, because the mere possibility of bad cases does not falsify the counterfactual. It's *possible* to stand to a proposition in a way matching the text's description of Joe's relation to *I have fifteen minutes* without NKJTB—but, Williamson says, the relevant counterfactual does not pick out the bad cases. Rather, he suggests, *if someone were to stand to a proposition in the way matching the text, it wouldn't be bad, even though it's possible to be in such a position in a bad way. If someone were to stand to a proposition in a way matching the story, it would be good.*

### **III. Worries for Williamson's Interpretation**

There are at least two reasons to be uneasy with Williamson's treatment. The first is one we've been focusing on already: Williamson's interpretation of the Gettier argument renders the intuition as a judgment of an *a posteriori* counterfactual; (2<sub>cf</sub>), which is meant to correspond to the Gettier intuition, can be at best contingently true. This is so because, in the terms of a Lewis/Stalnaker semantics for counterfactuals, a counterfactual's truth depends on characteristics of the *nearest* possible worlds where the antecedent is true, and which worlds are the nearest worlds depends on what the actual world happens to be like. (Any plausible alternative semantics will also allow such counterfactuals to vary in truth-value between worlds.) So warrant for the Gettier intuition, on Williamson's view, will involve a great deal of empirical knowledge about the actual world. If the Gettier intuition is like this, then it is not the sort of thing that traditional philosophy takes it to be. Williamson's account leaves intuitions as mere judgments about contingent matters of fact. As such they cannot be knowledge *a priori*.

This is, we think, sufficient reason to look further for a treatment of thought-experiment intuitions. But there is a more direct difficulty with Williamson's formulation: not only can (2<sub>cf</sub>) not be known *a priori*—for many, perhaps most, Gettier thought experiments, it cannot be known at all. Williamson's version of the Gettier intuition is, *if some person were to stand to some proposition in a way satisfying the text of the Gettier story, then anyone who satisfied the text of the story with a proposition would have NKJTB*. This counterfactual is false when the nearest worlds where someone satisfies the text include someone doing so in a bad way. Suppose that (perhaps unbeknownst to the Professor) in the actual world, there is someone who satisfies the text of Professor Muthos's story—and that he does so in one of the bad ways specified above.

---

<sup>4</sup> Williamson is trying to formalize something like (CF) "if a thinker were Gettier-related to a proposition, he/she would have justified true belief in it without knowledge." It can be surprisingly challenging to formalize such 'donkey anaphora' counterfactuals. See §4 of Williamson (2007)'s ch. 6, and Appendix 2. Our objections to Williamson here will transcend the question of whether this formalization faithfully captures the English counterfactual – they will apply to (CF) and (2<sub>cf</sub>) alike, even if they are not equivalent.

Then, in the nearest world where someone satisfies the text—the actual world—it is not the case that everyone who satisfies the text has NKJTB. Williamson’s  $(2_{cf})$  is false.<sup>5</sup>

This cannot be right. The appropriateness of Muthos’s thought experiment does not depend on there being no one in the actual world who satisfies the text in a bad way. On Williamson’s formulation, if there is such a person, Muthos’s students are misled, and *falsely* believe  $(2_{cf})$ . If they go on properly to reason to the conclusion that knowledge is not justified true belief, then they have come to a true belief by competently deducing it from a false belief. As we well know, deducing a true belief from a false belief, even a justified one, does not confer knowledge. So Williamson’s view implies that in this case, Muthos’s students do not know that knowledge is not justified true belief—a defective thought experiment has generated a Gettier case about the analysis of knowledge! This, we take it, is unacceptable.

A natural move is to weaken Williamson’s counterfactual; we replace the universal quantifier in the consequent with an existential:  $(2_{cfw}) \exists x \exists p GC(x, p) \square \rightarrow \exists x \exists p [GC(x, p) \supset (JTB(x, p) \ \& \ \sim K(x, p))]$ . This does not avoid our objection, for this counterfactual is false in worlds where exactly one person satisfies the story, and does so in a bad way. Furthermore, an additional objection arises as this weakening doesn’t allow for straightforward disagreement of intuitions. Someone who has the Gettier intuition,  $(2_{cfw})$ , won’t disagree with some (nonstandard) person who has the opposing intuition about the case and hence accepts  $(*) \exists x \exists p GC(x, p) \square \rightarrow \exists x \exists p [GC(x, p) \supset (JTB(x, p) \ \& \ K(x, p))]$  as  $(2_{cfw})$  and  $(*)$  are consistent.<sup>6</sup>

#### **IV. Truth in Fiction**

Williamson gives us the predicate  $GC(x, p)$ ,  $x$  stands to  $p$  as given in the text of the Gettier story. The way he treats it,  $GC(x, p)$  holds any time that  $x$  stands to  $p$  in a way that makes each sentence of the text true. But there are more resources available here. We suggest that what is missing from Williamson’s analysis is the recognition of the difference between a text and a story. There is more to a story than the literal claims of the sentences used to tell it. In particular, we can invoke the notion of *truth in fiction*.<sup>7</sup> Some, but not all, fictional truths are explicitly stated in the text. It is true in our fiction that Professor Muthos was teaching an epistemology course; this is so because our text included the sentence, “Professor Muthos was teaching an epistemology course.” Other fictional truths are not so directly identified. It’s true in our fiction that Brian had exactly two eyes, even though we didn’t say so. It’s true in our fiction that

---

<sup>5</sup> Williamson recognizes this worry in his (2007), pp. xxx. Incredibly, he seems willing to bite the bullet on this point, and to admit that in such cases, our thought experiments are defective. We find this result implausible, both on its face and in light of the particular objection mentioned in the main text below.

<sup>6</sup> Richard Heck has proposed (in conversation) another move on behalf of Williamson. Instead of a counterfactual, one could render premise (2) as a generic like ‘Fire engines are red’, which is true not because all fire engines are red (we’ve seen some green ones), but because fire engines typically are red. Premise (2) then becomes: in cases that satisfy the Gettier text, people (typically) have justified true belief without knowledge. The conclusion we arrive at, then, is that typical cases that satisfy the Gettier text are ones in which people have justified true belief without knowledge. The generic version of (2) would seem to avoid the objections that plague counterfactual renderings while keeping the spirit of Williamson’s counterfactual approach.

<sup>7</sup> The seminal piece on truth in fiction is Lewis (1978). See also Walton (1990) and Currie (1990). For recent developments in truth in fiction, see Hanley (2004) and the papers cited therein.

Professor Muthos was not a wizard, even though it's consistent with the text we wrote that he was. The challenge of a theory of truth in fiction is to explain all of this.

Truth in fiction may appear to be just what we need to rule out the bad Gettier cases. Although it's consistent with Muthos's text that Joe was in a bad Gettier case, it is *true in the fiction* that Joe was in a good Gettier case. (Compare: although it's consistent with our text that while relating the story about Joe, Muthos grew a third eye, it is true in the fiction that he did not do so.) So we might try using truth in fiction in a reformulation of the Gettier argument.

The 'first stab' formulation from §2 was unsound; (2<sub>n</sub>) is false, because there are bad stories that fit the Gettier text. But suppose we fix the fictional truths themselves—not merely the text of the fiction. Let GC<sub>tf</sub>(*x*, *p*) stand for *x stands to p in the relation in which it is true in the fiction that Joe stands to I have fifteen minutes*. Then we have:

$$(1_{tf}) \diamond \exists x \exists p GC_{tf}(x, p)$$

$$(2_{tf}) \Box \forall x \forall p [GC_{tf}(x, p) \supset (JTB(x, p) \& \sim K(x, p))]$$

$$\text{Therefore: } (3) \diamond \exists x (JTB(x, p) \& \sim K(x, p))$$

On this formulation, both premises invoke truth in fiction: it's possible for a person and a proposition to be in the relation that it's true in the fiction that Joe is in to his proposition. And necessarily, any such person has NKJTB. This formulation is sound, and we need no counterfactual; (2<sub>tf</sub>) is a necessary universal, just as we wanted it to be. So is this a satisfying account of the Gettier argument?

Not quite, for at least three reasons. First, on this suggestion, the Gettier intuition stands on the notion of truth in fiction. One might worry, then, that everything would depend on the correct theory of truth in fiction; the Gettier intuition would seem to be *a priori* only if we can access fictional truths, given texts, *a priori*; and this is far from clear. Indeed, on one of David Lewis's views a truth in fiction claim *is* a counterfactual claim: *p is true in the fiction* is analyzed as, *were the story told as known fact, p would be true*.<sup>8</sup> Our view would then be similar to Williamson's, with respect to the epistemic status of the Gettier intuition.<sup>9</sup> Of course, there are good reasons to think that the Lewis view is not right, and some of them locate the problem just with the too-contingent nature of the counterfactual.<sup>10</sup> Nevertheless, the correct theory of truth in fiction will very likely share this feature with Lewis's: it will have it that fictional truths cannot

<sup>8</sup> Lewis (1978). This is (paraphrased) Lewis's *Analysis I*, the simplest of several theories of truth in fiction he suggests.

<sup>9</sup> But the views would still not be identical. Consider the modal status of the propositional content of the Gettier intuition. On the Lewis account of truth in fiction, the account above would imply that in order to reach the Gettier conclusion, one relies on knowing (*q\**): *necessarily, if the fictional truths (for this actual fiction) were true, Joe would have NKJTB*. On the Williamson, one relies on knowing (2<sub>cf</sub>): *if some x were to stand to some p as in the Gettier text, then anyone who stood to any proposition in a way matching the text would have NKJTB*. Williamson's (2<sub>cf</sub>) is false in possible worlds where nearby Gettier worlds include bad cases. But *q\** may be true in all worlds, as 'actually' rigidifies the embedded counterfactual. A given story has the same fictional truths in every possible world; a given text picks out different fictional truths depending upon the circumstances in which it is told.

<sup>10</sup> One of the best sorts of objection to this view is raised by Lewis himself in (1978), p. 274. For additional arguments against Lewis's account, see Currie (1990), §2.3, pp. 62-70.

be known *a priori*. If we cannot know fictional truths *a priori*, this formulation is not one that obviously meets our criteria.<sup>11</sup>

A second problem for the  $GC_{tf}$  formulation involves the invocation of *truth in fiction* in the content of the Gettier intuition. Many philosophers of fiction think that our ordinary sentences about fictional characters include an unnoticed elliptical ‘it is true in the fiction that’ operator,<sup>12</sup> so it is perhaps not terribly worrying that our Gettier intuitions don’t *feel* like judgments about fictions. Nevertheless, it seems wrong to suppose that invocation of fiction is an essential part of Gettier reasoning. Non-fictional Gettier cases work just as well as fictional ones to set up the Gettier conclusion. If Muthos had presented his story to his class as fact rather than as fiction, they would have gone through the Gettier reasoning in just the same way. So it does not seem as though the concept *truth in fiction* can play a role at this level.

A third challenge: we have presented our formulation using the  $GC_{tf}$  relation, the *relation in which it is true in the fiction that Joe stands to I have fifteen minutes*. Which relation, however, is *the* relation? Joe stands in many relations to the proposition in question. Which of those is the  $GC_{tf}$  relation? Specifying one is difficult, but without specifying one we have not fully offered a formulation.

### **V. From Truth in Fiction to Entertaining Propositions**

There is something attractive about the  $GC_{tf}$  formulation given above, but it will not do as it stands. We suggest that a better application of truth in fiction in a Gettier formulation will make use of the notion less directly. On our view, the fictions that are thought experiments are useful for *picking out* and *thinking about* propositions that are key to the Gettier argument. As we see it, one listening to a thought experiment can consider the set of propositions that are true in the fiction, and in particular, the proposition that every member of that set is true, and then subsequently reason with that proposition to the conclusion of the thought experiment. Let us explain.

Begin with the set of fictional truths. Given a fiction, normal speakers recognize that some propositions are true in the fiction, and others are not. So, when Brian hears Muthos’s story, he may consider the set of propositions that are true in the fiction. Brian cannot, of course, entertain each proposition of this infinite set individually, but he can grasp and entertain the set containing all of them, and refer to it. Suppose Brian names the set *STORY*. Now, Brian may go on to entertain *p*, the proposition that *every element of STORY is true*. This proposition will play the crucial role in the Gettier reasoning. Now, Brian may reason thus:

---

<sup>11</sup> In an earlier version of this paper, Ichikawa defended a view related to the one explored in this section. That account was: (1) It’s possible for someone to be in a position like Joe’s position in this fiction I’m actually engaging with. (2) Necessarily, anyone in a position like Joe’s position in this fiction I’m actually engaging with would have NKJTB. (3) Therefore, it’s possible to have NKJTB. The rigidification of the fiction was to allow for us to misinterpret authors and engage with our own private Gettier fictions, which would underwrite our fictions. On this view, the Gettier conclusion here will be *a priori* if one knows *a priori* that Joe’s position in the fiction one’s actually engaging with comes to justified belief without knowledge. If (and only if) introspective knowledge is *a priori*, this is plausible. Our eventual account will not depend on the apriority of introspection.

<sup>12</sup> See especially Lewis (1978) and Walton (1990).

(1<sub>p</sub>)  $\diamond p$ <sup>13</sup>

(2<sub>p</sub>)  $\square (p \supset \text{Someone has NKJTB})$

Therefore: (3)  $\diamond \exists x (x \text{ has NKJTB})$ .<sup>14</sup>

Here, our invocation of fictional truth explains how we come to entertain the proposition  $p$ , but the concept FICTION does not enter into the Gettier reasoning itself. Competence with truth in fiction is an important step in engaging with the thought experiment, but its role is exhausted before the actual invocation in reasoning of (2<sub>p</sub>), the Gettier intuition. Put another way, one's ability to grasp a story told through a text serves only as a *means to grasp* a certain proposition, which will figure into the intuition; the content of the intuition itself makes no use of the notion of truth in fiction.

Of course, the ability to grasp stories through texts involves (perhaps tacit) *a posteriori* knowledge, albeit knowledge that even small children possess to some degree.<sup>15</sup> However, this *a*

---

<sup>13</sup> Premise (1<sub>p</sub>) isn't *quite* right. Joe is a fictional character, and as Kripke says, he may be essentially so. (Although it is possible for there to be some person that matches the description of Sherlock Holmes, we might doubt that any such possible individual would *be* Sherlock Holmes.) If this is right, then it is not possible for all of the members of STORY to be true, for it is not possible for any person to be Joe. The solution to this worry is technical and peripheral to the larger project, and the simpler version we give above is clearer and more intuitive, so we present the more accurate version in this footnote.

Although it is not strictly possible for fictions involving characters picked out by proper names (who are not inhabitants of the actual world) to be true, there are available modified versions of the fictions which are possible. Professor Muthos's story began with the metaphysically impossible, "Joe had left his watch at home." A modified, metaphysically possible version of the story would insert, at the start of the story, something like the computer scientist's variable declaration: existential introductions to all the elements in the story that would have had proper names, along with instructions as to how to refer to them while engaging with the story. So we'd have something like: "There was some guy, whom we will refer to as 'CHARACTER1'. CHARACTER1's name was 'Joe'. Joe had left his watch at home...." If every object with a proper name is introduced thus, then the story is metaphysically possible.

On the more technical account, then, we replace the story with a new story (whose truths describe possible worlds that are qualitative duplicates to those described in the old story), taking care to introduce each character in the way described. Then we can reason as described above, using the set of truths in the new fiction.

See pp. xxx of Williamson's (2007) for a similar treatment.

<sup>14</sup> Ernest Sosa has suggested in conversation that these two premise argument accounts of Gettier reasoning don't do justice to the simplicity of the Gettier intuition. Really, in having the Gettier intuition, one just comes to see that one way to satisfy the text of the Gettier story involves someone having justified true belief without knowledge. Sosa's suggestion may amount to the proposal that people directly apprehend (3). Moreover, while (3) is no doubt necessarily true, necessity doesn't enter into its content. (Hence, the discussion in Section X is irrelevant.) The propositional content of the Gettier intuition, then, is much simpler in form than our (2<sub>p</sub>).

After some thought on the issue, we're not confident that this suggestion represents a genuine alternative to our view. After all, we need some account of how the thought experiment generated (3). Surely, it did so through guiding one to consider a fairly determinate possibility—something like the possibility in which  $p$  is true. But it's not enough merely to have in mind this determinate possibility; one has to see that if this possibility obtains, then there would be an instance of justified true belief without knowledge. Of course, this process is very close to (if not identical with) what we've suggested. We don't mean to insist that such a process be entirely explicit, conscious, or deliberate, but it's hard to see how thought experiments work without some such process. See also Williamson's (2007), pp. xxx.

<sup>15</sup> For some psychological studies on the way humans think about fictional worlds, see Skolnick and Bloom, 2006a and 2006b. Weisberg (*née* Skolnick) and Bloom are currently working on a more directly-related question: how children and adults come to recognize fictional truths that are not explicitly stated. Results are not yet published.

*posteriori* knowledge, tacit or not, does not prevent the thought-experiment reasoning from being *a priori*. Consider an analogy. Suppose that Professor Muthos utters the following sequence of sentences:

(i) If Julius Caesar is the successor of the number one, then Julius Caesar is identical to the number two.

(ii) Two is a prime number.

Hence, (iii) If Julius Caesar is the successor of the number one, then Julius Caesar is a prime number.

For his students to understand Professor Muthos's verbal reasoning, they need *a posteriori* knowledge: they need whatever *a posteriori* knowledge is required for interpretation of these sentences. The students' *a posteriori* knowledge of English allows them to come to grasp the propositions *if Caesar is the successor of the number one, then Caesar is identical to the number two*, *two is a prime number*, and *if Caesar is the successor of the number one, then Caesar is a prime number* from his English utterances, which, in turn, leads them to run through that bit of reasoning. The exchange student in Muthos's class may not have sufficient *a posteriori* knowledge of English; as a result he may fail to run through the reasoning. The reasoning, however, is not *a posteriori*, because the *a posteriori* knowledge the students deploy is not deployed in the content of their reasoning, but rather as a means to get to that content. The reasoning is *a priori* because of the *a priori* status of the premises and the *a priori* entitlement to move from the premises to the conclusion.

Likewise, even though Muthos's students need competence with fictions in order to grasp the relevant propositions for the Gettier reasoning, this does not imply that the knowledge they come to receive is *a posteriori*. Only if the *a posteriori* knowledge were deployed as *warranting* could we so conclude. But it is not so deployed. It is merely deployed *as a means to come to grasp the propositions involved in reasoning involving the thought-experiment intuition*. The *a posteriori* knowledge serves as a sufficient (not a necessary) *causal enabler* for the reasoning process; it does not play a warranting role within the reasoning itself.

It is crucial here to note the way in which we introduced the proposition *p*. It can be tempting—and, in some contexts, harmless—to think of *p* descriptively as the proposition *that all of the propositions that are true in the Gettier fiction are true*. However, *p* does not have that descriptive (Fregean) sense; it is merely true in all the same possible worlds. Otherwise, *truth in fiction* would be a constituent in the thought-experiment intuition, and one's *a posteriori* knowledge of what is true in the Gettier fiction would stand in the warrant relation to the conclusion that knowledge is not necessarily justified true belief. But this is not so in our formulation. Knowledge of truth in fiction plays only the role of a (merely sufficient) causal enabler; our account is compatible with someone going through the Gettier reasoning without any knowledge of truth in fiction whatsoever.

Here is an example. Suppose that one of Professor Muthos's students, Brian, is linguistically incompetent with respect to truth in fiction. So he lacks the *a posteriori* knowledge we typically rely on to engage with fictions. In many cases, the stories he generates from texts

differ radically from those that most people generate from the same texts. If we suppose, however, that on the occasion that Professor Muthos relates the Gettier text, Brian generates, merely by luck, the Gettier story most people generate, Brian may come to the Gettier conclusion via the reasoning given above—this, in spite of the fact that Brian does not have even tacit knowledge of the principles of generation for truth in fiction. It makes no difference whether Brian really is in touch with truth in fiction, or even whether he considers the text to be a presentation of a work of fiction at all; if he treated the text as assertive testimony, he could still reason in just the way we describe.<sup>16</sup> That people regularly come to have a (tacit or explicit) representation of the same story explains how it is that we can so easily *share* thought-experiment intuitions; but it is inessential in explaining how some *particular* person comes to a conclusion on the basis of a thought-experiment intuition on a particular occasion.

## **VI. The Psychology of Thought-Experiment Intuitions**

One might object to our formulation of the Gettier reasoning on phenomenological grounds. Does it really *seem* that people go about naming the collections of propositions constituting a story as a precursor to engaging in the Gettier reasoning as we have suggested? Plausibly not. Nonetheless, it does seem that in many cases the Gettier intuition may come to many people with a demonstrative in the content; the Gettier intuition might come to many as:

(2<sub>that</sub>) Necessarily, if things are like *that*, then someone has NKJTB.

where ‘that’ refers (directly) to how things are according to Professor Muthos’s story. As demonstratives in a context and proper names share many of the same characteristics, it’s not too far of a stretch to think of a demonstrative in a context as a (temporary) proper name—at least for our purposes. If the Gettier intuition comes as something like (2<sub>that</sub>), then again as with (2<sub>p</sub>) the concept of truth in fiction makes no appearance in the content of the Gettier intuition. Rather, one’s competence with truth in fiction goes towards fixing and apprehending the reference of the demonstrative. Thinking of the demonstrative as a proper name, it’s not too hard to see that in many cases something very much like the account given in the last section may actually be the reasoning process people go through when they run into Gettier cases. In these situations, people aren’t explicitly baptizing with a proper name, but the tacit reference-fixing of a demonstrative amounts to pretty much the same thing in the end (at least for our purposes).

Of course, the question of how well our account—or any other, for that matter—matches what actually goes on in people ultimately must be answered through empirical investigation. While it may be a virtue of our account (over one like Williamson’s) that it better matches the way people reason with thought experiments, we don’t want to rest much on that point. We are content to establish the plausibility of two claims: first, that people have *a priori* justification for believing the conclusions traditional philosophers often take thought experiments to substantiate, and second, that they are *a priori* justified in believing such conclusions by going through a process similar to the one that many people actually go through. Whether it is the case that people are actually *a priori* justified in believing such conclusions we leave, to some extent, as an open question whose answer could only be determined by a more thorough examination of

---

<sup>16</sup> Williamson (2005) also emphasizes that Gettier cases need not be fictional (p. 15).

how people reason, and how similar their reasoning is to the ideal practices that would yield a *a priori* justification.

## **VII. Apriority and Possibility**

When rejecting the naïve account summarized in Section II, Williamson considers (in lecture and conversation) a modified account in which a ‘things are as we would otherwise expect’ clause is appended to the text of the thought experiment. He makes two complaints against this modification. The first is that what *we* expect is still a contingent matter. One way to see this is to recognize that we’re introducing another counterfactual: ‘what we expect’ will have to be evaluated as something like ‘what the relevant members of the community *would* expect if *they were given* the story’. So this move won’t help the modal intuitions team.

The second objection Williamson raises to this modified account is that invocation of a clause like this begins to threaten the status of premise (1) in the Gettier reasoning. As we originally understood it (1) states (roughly) that the Gettier situation—that situation given in the literal Gettier text—is possible; with the modification it would state that the Gettier situation—that situation given by the Gettier text appended by the ‘things are as we would otherwise expect’ clause—is possible. It’s somewhat obvious that it’s possible for the literal text Professor Muthos gave to have been true; Williamson points out, however, that it is much less obvious the text could have been true if it had concluded, “And things were otherwise as you’d expect.” We probably would expect a great number of things if a case like Joe’s were related to us, and it’s not immediate that all these things taken individually are possible and less immediate still that they are compossible.

This modified things-are-as-you-would-otherwise-expect account Williamson considers is in many ways similar to ours—but, of course, not equivalent. Williamson’s first objection demonstrates an advantage of our view; *things are otherwise as we expect* looks again like a contingent counterfactual, but our truth-in-fiction account does not. On the other hand, the second criticism Williamson makes reveals a point of commonality between the two views. There *does* seem to be parallel worry on our view: how can we know that the story we’re engaging with in a thought experiment describes a possible situation?

We are sanguine. To be sure, Kripkean considerations demonstrate that sometimes empirical knowledge rules out the possibility of propositions we took to be *a priori* possible—*Hesperus is not Phosphorus*, for instance.<sup>17</sup> Nevertheless, broad skepticism about possibility does not seem to be in order. It is in some difficult-to-spell out sense *conceivable* that the members of STORY are all true.<sup>18</sup> Kripke-style considerations have prompted some skepticism about the idea that this kind of conceivability is a good guide to metaphysical possibility, but there does seem to be a useful notion of *conceptual* possibility to which this conceivability is an excellent guide. Conceptual possibility is closely tied to what one can rationally and coherently

---

<sup>17</sup> See Kripke (1980). For some of the considerable discussion in this area, see for example the papers in Gendler & Hawthorne (2004).

<sup>18</sup> For a few attempts to spell out this conceivability, see Chalmers (2004) and Yablo (1993). For a very different kind of strategy, see Bealer (2004). We do not commit to the success of any of these attempts here; we merely express confidence that there is a relevant kind of conceivability according to which our *p* is conceivable.

conceive.<sup>19</sup> If a proposition is a conceptual possibility, then an ideal rational agent can coherently conceive of it as true, and hence have no reason to conclude *a priori* that it is false. Understood as a claim of conceptual possibility, then, our premise (1<sub>p</sub>) should not be controversial. While there may be some reason to be skeptical about conceptual possibility—as we are not ideal rational agents—its close ties to our own cognitive abilities that approximate ideal rationality at least some of the time make radical skepticism with regards to it implausible.

Of course, one pressing the Williamson worry would not find this initial reassurance particularly reassuring. After all, one of the occasions in which we generally fall short of ideal rationality is when large amounts of computation are required. Yet, surely large amounts of computation are required to check that all of the propositions true in the Gettier story are coherently co-maintained and consequently conceptually co-possible. As such checking would seem to be prerequisite to knowing that our premise (1<sub>p</sub>), understood as a claim about conceptual possibility, is true. So it would seem that our rational failings prevent us from knowing as much.

So, how can we know that the presumably infinite propositions true in the Gettier story could be coherently maintained by an ideal rational agent given our limited cognitive abilities? To answer this question, we need to note another difference between our truth-in-fiction account and the things-are-otherwise-as-we-would-expect account Williamson rightly rejects. To know that what we would expect given the interpreted text of the Gettier case is conceptually possible, we would have to check that all of our expectations are collectively coherent—a genuine problem. To know that the story given by the Gettier text is possible, on the other hand, one does not need to check that all of the fictional truths are collectively coherent; one merely needs to recognize that fictional truths are (at least typically) generated so as to maintain their collective coherency. What is true in the fiction is not identical to what we would expect it to be given the interpreted Gettier text. If one were to discover that some expected fictional truth were incoherent with the other expected fictional truths taken together, one would not have discovered that the fiction is incoherent—and *a fortiori*—impossible; rather one would have discovered that the expected fictional truth in question was not in fact a fictional truth. Fictional truths are closer to what an ideal rational agent with our background knowledge and other cognitive abilities would judge them to be given the interpreted Gettier text. Given the coherency constraint on generating fictional truths, so far as we can see, there's no reason to be especially skeptical as to whether the Gettier story is conceptually possible. One's grasp of the proposition *p* comes with an understanding that the propositions it entails are collectively coherent.

This would settle the matter if we were understanding (1<sub>p</sub>) as a claim about conceptual possibility. But we are not; the conclusion of the Gettier argument is that NKJTB is possible – *metaphysically* possible. To establish this using the bit of reasoning we claim the Gettier argument relies on, we need *p* to be metaphysically possible. Thus, one might object that although we can know *a priori* that *p* is *conceptually* possible, it could only be known *a posteriori* that *p* is *metaphysically* possible. We do not find this objection compelling. Although Kripke has shown us some propositions that are conceptually but not metaphysically possible, these trade either on rather specific features of proper and natural kind names or on supposed essentialisms of various sorts, e.g. origin or material constitution. So, for instance, *Hesperus is not Phosphorus* is metaphysically impossible because 'Hesperus' and 'Phosphorus' refer rigidly

---

<sup>19</sup> See Yablo (2004) for one interesting discussion of conceptual possibility.

to the same object—something that requires some empirical investigation to establish. Likewise, *the Empire State Building is made out wood* is metaphysically impossible given the putative essentialism of material constitution because the Empire State Building is actually made using steel—again something we would have to verify through empirical investigation. It is difficult to see how something similar could interfere with the move from our story’s conceptual possibility to its metaphysical possibility. Are we to be worried that there is an empirically-discoverable fact establishing that ‘know’ and ‘believe truly and justifiedly’ rigidly refer to the same relation? Given our uses of the two terms, such a discovery seems inconceivable.<sup>20</sup>

Indeed, the preceding discussion suggests the following *a priori* test for metaphysical possibility, assuming *a priori* access through conceivability to conceptual possibility:

For any proposition *q*, if *q* is conceptually possible but metaphysically impossible, then there is some conceptually possible empirical discovery which would show that *q* is metaphysically impossible.

Why think this? For one thing, any such *q* is an *a posteriori* necessity, and *a posteriori* necessities can be discovered. If all possible discoveries are conceptually possible, as they must be if we have an adequate conceptual scheme, then this is sufficient to prove the principle.<sup>21</sup> If this test is right, then the conceivability of *p*, along with the inconceivability of learning that *p* is metaphysically impossible, provides an *a priori* sufficient condition for *p*’s metaphysical possibility. As Joe’s story is conceptually possible and it is inconceivable that an empirical test could show it metaphysically impossible, it would seem we have very good reason to think the story is metaphysically possible. Thus, we don’t see that Kripkean considerations offer us any reason for worry about the *a priori* status of premise (1<sub>p</sub>).

Of course, to actually establish that we have nothing to worry about would require far more work than we have attempted here. For one thing, even given our test, we would need *a priori* access to conceptual possibility. We are optimistic that this is plausible, but it is clear that the notion of conceptual possibility requires more attention than our current constraints permit.

### **VIII. Conceptual Role, Meaning, and Justification**

Thus far, we have pointed out that because people have a grasp on truth in fiction, we can use fictional texts to communicate far more than their literal contents. This in turn allows people to come to grasp propositions too difficult easily to express in literal speech. As a result, the challenges Williamson raises to a traditional understanding of philosophical methodology fail. But what of the naturalistic challenges that motivated Williamson? We have so far paid little heed to whether propositions like the content of (2<sub>p</sub>) (our version of the Gettier intuition) that we grasp through apprehending thought experiments can be known *a priori*—or even whether they can be known at all. In these next three sections, we will argue that on our account, naturalistic skepticism about our ability to learn through thought experiments is not warranted. Whatever it

---

<sup>20</sup> Brian Weatherson (2003) argues that given David Lewis’s theory of meaning (the “reference magnetism of natural properties” view—see Lewis (1984)), we might just discover that ‘know’ and ‘believe truly and justifiedly’ are co-intensional. We find this suggestion implausible and will ignore it. (If Lewis’s theory of meaning does entail as much, all the worse for his theory of meaning.)

<sup>21</sup> Thanks to Richard Heck for helpful discussion here.

is that explains our capacity for everyday knowledge should also be able to explain knowledge of thought-experiment intuitions. We suggest, consequently, that there is nothing mysterious about the invocation of thought-experiment intuitions. Since there must be a naturalized explanation of everyday knowledge, there is a naturalized explanation of our knowledge of thought-experiment intuitions. Our argument begins with some basic truisms of metasemantics.

Let's explore our first truism. Suppose that we were trying to teach you a new adjective, 'hopta', and we went about 'explaining' it in the following manner:

It might be the case there some things are actually hopta. Then again, maybe nothing is hopta. Indeed, it might be impossible for a thing to be hopta—or it might not. Perhaps people are the sorts of things that are hopta. On the other hand, perhaps small particles are the proper bearers of hoptatuity. Maybe being hopta is having a certain shape. Alternatively, it could be having a disposition to hate high-pitched musical instruments. Or it could be none of those things.

We take it that you have no idea what 'hopta' means. You don't understand the word; you haven't grasped some concept *hopta*. Why not? Our 'explanation' gives you no way to associate a robust conceptual role with 'hopta'. You don't really know what you would be supposing in supposing that something was or wasn't hopta. You might associate a conceptual role with 'hopta' that involves deferring to our expert judgment. However, this only works if 'hopta' plays some conceptual role for us, which in fact it doesn't. Consequently, there is no concept *hopta* to grasp. The word 'hopta' has no meaning at all.

There is an important point not to be overlooked: we all can disagree on how tight a connection there is between concepts and conceptual roles. Still, we can (almost) all agree with the Wittgensteinian point that to have a meaning requires having some or other conceptual role.<sup>22</sup> This is our first truism: *Terms without a use have no meaning.*

There is another point of common ground here regarding the connection between meaning and conceptual roles. If we are to explain everyday knowledge that *p*, it's going to have to be the case that the conceptual roles or 'uses' associated with the constituent concepts of *p* are capable of delivering knowledge under some everyday conditions about certain everyday cases. Thus, the metasemantic account we give has to permit that at least some beliefs formed in accordance with the conceptual roles associated with the constituent concepts are—at least on certain occasions and under certain suppositions—not only true and correct but also justified. This is our second truism: *If we have knowledge with a concept as a constituent, we must have some ability to justifiably apply that concept, at least in some favorable circumstances.*<sup>23</sup>

Our second truism is not redundant; without it, even if all meanings had conceptual roles, we couldn't rule out their having conceptual roles that never offer correct deliverances: a word 'ghacklin' could have the same conceptual role as that of the word 'squirrel' in an English idiolect and yet mean (and have the same referent as) *baby bird*; we would consistently form false beliefs about ghacklins. This is implausible; we can't have this sort of situation for any

---

<sup>22</sup> The only recent dissension we know of from this truism is Bonjour (1998).

<sup>23</sup> It's worth noting that neither of these truisms commit one to anything but a very thin (and Fodor friendly!) version of conceptual role semantics. See Fodor (1998) and Greenberg & Harman (2006).

concepts constituent in propositions that we know. Clearly so much should follow from the correct metasemantic account of meanings/concepts. Indeed, not only should so much follow, the correct metasemantic account also needs to allow for an explanation of how it is that when one forms beliefs in accordance with conceptual role of constituent concepts, the resulting beliefs are not merely often true, but also justified. Without an explanation of this sort, there will be no explanation of how we have any knowledge at all.

Notice, however, that once we have an explanation of how one possessing concepts can in certain circumstances track the actual truth of propositions deploying those concepts using concomitantly associated conceptual roles, there's no mystery about how one could track the truth in fiction of propositions deploying those concepts using those same associated conceptual roles. How does one track whether Joe has knowledge in the Muthos story? One uses the same conceptual role associated with *know* that one can use to track whether people actually know or not. If one actually encountered a case like Joe's, the conceptual role one associates with 'know' would give one the resources to know that it wasn't an instance of knowledge. One uses those same resources when one encounters the case through fiction.

Consequently, there's nothing non-natural or mysterious about thought-experiment intuitions. All concepts have associated conceptual roles. These conceptual roles deliver warranted beliefs about certain cases under the right conditions. Sometimes these cases are presented to us through perceptual faculties; sometimes they are presented to us through thought-experiment fictions. Presumably there's nothing unduly mysterious about how conceptual roles deliver knowledge about actual cases; whatever naturalized explanation we can give there should work just as well for the non-actual cases presented to us via thought-experiments.

### **IX. Tracking Truth in Near and Distant Worlds**

These results shouldn't be surprising, as they are harmonious with much that we should expect for other reasons. It's hard to see how one could recognize correct or incorrect deployment of a concept one possesses in actual cases without being able to recognize its correct or incorrect deployment in nearby non-actual cases. So any explanation that explains our ability to apply the concept correctly in actual cases will also explain our ability to apply the concept in many counterfactual cases. To take advantage of our ability to recognize correct deployment of a concept in non-actual cases, one merely needs a medium through which non-actual cases might be presented. Thought experiments give us such a medium.

The foregoing considerations suggest that *any* thought-experiment intuitions attributing concepts (whether 'conceptual analysis' intuitions or moral intuitions about trolley cases, whether *a priori* or not) should be accurate so long as, first, our actual faculties for recognizing correct or incorrect deployment of the concept are genuinely good, and second, the counterfactual cases under consideration are relevantly similar to what we might encounter in the actual world.<sup>24</sup> Hence, we have some reason to think that any thought-experiment intuition should allow for knowledge so long as we are sufficiently good at deploying the concepts

---

<sup>24</sup> For a good discussion of a particular sort of case in which we may go awry in distant worlds, see Gendler (2002). Gendler's suggestion that we are subject to certain kinds of systematic mistakes with distant thought experiments is, of course, consistent with our invocation of our general reliability.

constituting the propositional content of the intuition in actual cases and so long as the thought experiment isn't so very far-fetched. Gaining knowledge through thought experiments isn't mysterious; it's done continuously with the way that we gain knowledge every day. And this is sufficient for our project about the Gettier intuition; most Gettier stories do not invoke distant worlds. But in fact, we also believe that in many cases, we can be reliable even in distant worlds.

The preceding considerations offer us an *argument* for the conclusion that we should be able to track truths in proximal possible worlds (assuming that we track them in the actual one); they don't, of course, offer an *explanation* of how it could be that we are able to track those truths. If we can track truths in proximal possible worlds, however, there must be some explanation of that fact. That sort of explanation just would be an explanation of how it is that the conceptual roles concomitant with the concepts we possess yield justified true beliefs when confronting actual and other similar circumstances. This is, of course, the very sort of explanation we pointed to in the last section.

Importantly, however, the explanation linking conceptual roles and justification could easily extend to an explanation of how conceptual roles concomitant with the concepts we possess can yield justified true beliefs even when confronting some far-fetched circumstances. Indeed, it would be surprising for this not to be the case. While (rather infamously) people disagree as to their judgments as to what's true in some far-fetched scenarios, there is also very widespread agreement about many ordinary cases.<sup>25</sup> Suppose for instance that someone on a planet whose oceans are made of orange juice, someone becomes submerged in the masses of orange juice and dies when his lungs are filled up with juice so that he suffocates. Did that person drown? Most English speakers will answer that he did. In this way we might find evidence that the particular liquid a person suffocates in is immaterial to whether he or she drowns—it would seem we have succeeded in doing some conceptual analysis of drowning. Yet, how do we come to track whether 'drowning' applies in this world that must be so very different from ours if indeed the oceans there are made of orange juice? Why think that are abilities tailored for a world like ours will work for a world like that one?

There need be no real mystery here either. We see no reason not to think that thought-experiment intuitions should allow for knowledge even in far-fetched cases; after all, on many occasions, the conceptual role associated with a person's grasp of a concept is sufficient to guarantee that the person be able rationally to track the truth of propositions deploying that concept—even in cases far into conceptual space. To return to our example, it would seem that a conceptual role necessarily concomitant with grasping *drowning* (as opposed to some other related concept) allows us to track drowning in some instances very far into conceptual space—a person apparently voicing the opinion that the unfortunate fate doesn't constitute a drowning in this far-fetched scenario would provide evidence that 'drowning' is his idiolect meant something somewhat different than it does in ours.

---

<sup>25</sup> There are numerous ways that a correct metasemantic account could explain disagreement. An easy explanation is provided if the relation between concept and conceptual role is one-to-many. Different conceptual roles could lead to different judgments even if they coincide when it comes to more quotidian cases, and concepts might apply indeterminately in such cases. Alternatively, if for every concept there is one conceptual role, it might be that people possess different but very closely related concepts.

While it is, of course, a matter for debate how much our judgments on these matters have to agree in order for us to possess the same concepts, without question, the tolerance of departure in such judgments has a limit before one begins to suspect that disagreement stems from possessing different (often related and nearly actually co-extensive) concepts—this even when judgments concern possibilities far distant from those we actually face. If agreement on such judgments did not have to extend into distant conceptually possible worlds, it's hard to see how our concepts could be so finely individuated as they sometimes are.

Put simply, our point is this: we know that any explanation of ordinary knowledge must include an explanation of how a conceptual role concomitant to grasping a concept can yield knowledge, at least in ordinary cases. And there is every reason to expect that this explanation will also apply in more distant cases. There has to be an explanation for why a person grasping *red* is justified in judging in good light that red surfaces are red. Someone who wasn't justified in applying the word 'red' in those instances wouldn't mean *red* by 'red'. In analogous fashion, there has to be an explanation available for why a person grasping *drowning* is justified in judging that the person submerged in orange juice in our story drowned. Someone who wasn't justified in applying the 'drowning' in those instances wouldn't mean *drowning* by 'drowning'. An explanation of how we know conceptual truths is thus continuous with an explanation of how we know anything at all. To explain either, we must explain the connection between possessing a concept and using its associated conceptual role rationally to track whether propositions with it as a constituent are true. Any non-skeptic about ordinary knowledge should expect that there is an explanation for how we go about so-called 'conceptual analysis.' Disagreements in intuitions over some thought experiment may indicate reason for local skepticism about some particular conceptual analysis, particularly when the thought experiment is far-fetched or when there is good reason to think we'd be prone to systematic error.<sup>26</sup> However, the foregoing considerations strongly suggest that general skepticism is unwarranted.

### **X. Knowledge and Necessity**

Our suggestion in the last two sections has been that the competence with concepts needed to explain everyday knowledge is sufficient to explain our warrant for thought-experiment intuitions. So, for example, the faculty underwriting our ability to pick out cases of knowledge in everyday circumstances is the same ability at work in picking out cases of knowledge as presented to us through fictions. This line seems somewhat plausible. Still, one might think that there's a special problem not with knowing the proposition, but with knowing that necessarily the proposition is true. How can we explain our knowledge of necessity?

We can't hope to give an adequate modal epistemology here; instead, we prefer to point out that there's no special problem here. The concept *necessarily* has associated with it a particular conceptual role. Indeed, it is not too difficult to give a rough characterization of that role. If, in imaging an arbitrary possible scenario (i.e. one free of any particular or special features), one concludes that *p* is the case, then one infers *necessarily, p*. From *necessarily, p*, one can conclude that in any possible scenario one can imagine, it is the case that *p*. To give an adequate modal epistemology is to show how this conceptual role tracks necessary truths, i.e. it is to demonstrate how the conceptual role produces justified beliefs about what is necessary.

---

<sup>26</sup> See Gendler (2002).

Moreover, presumably there is an adequate modal epistemology, because presumably we have as much everyday knowledge about necessity as we do about knowing, believing, or justification. It's widely known that necessarily, if something is a dog, it is an animal. It's also widely known that it's not necessarily the case that it won't rain, even if the weatherman said it won't. (Of course, different notions of necessity are at work here, but we have parallel conceptual roles at work, the difference in conceptual role being the scenarios we're counting as possible.) These pieces of common modal knowledge provide ample evidence that there is an explanation of how the conceptual role of *necessarily* produces justified true modal beliefs. That we philosophers have not yet come up with a conclusive explanation provides a challenging avenue for future investigation, not a reason for broad skepticism.

Incidentally, one might well accept Williamson's own explanation on this count. Williamson points out that we can define metaphysical necessity in terms of counterfactuals.<sup>27</sup> Consequently, an adequate account of counterfactual knowledge (which Williamson gestures towards) is sufficient for an adequate account of knowledge of necessity. Although we reserve some skepticism as to this approach, we have no reason to take issue with it for the purposes of this paper.

### **XI. Intuitions and Apriority**

Even if we can explain our thought-experiment intuitions in a naturalistic way, why think that thought-experiment intuitions like the Gettier intuition are justified *a priori*? There seems to be a burden of proof argument in our favor. Certainly, we generate fictional truths from the text via our *a posteriori* knowledge. As we have emphasized, however, all of this *a posteriori* knowledge is non-warranting for an intuition like ( $2_p$ ); its role is merely to apprehend and help fix a story given a certain text. In a Gettier case like Joe's, our *a posteriori* knowledge helps deliver the 'facts' of the case. Given these 'facts'—that Joe looked at the clock, that it was this time, that he formed this belief, etc.—what *a posteriori* knowledge does one rely on to arrive at the conclusion that he has a justified true belief without knowledge?

This burden-of-proof style argument can hardly settle the question. To really determine whether thought experiment intuitions are *a priori* requires, in the first place, some criteria for apriority—such have proven quite difficult to give. Contributing our own attempt to provide adequate criteria is not our project, but it's worth considering some lines that philosophers often pursue when trying to make the *a priori/a posteriori* distinction. These considerations should strengthen our burden of proof argument by giving us some further indication of which side of the distinction we might expect thought experiment intuition to come down on.

One way people have tried to carve up the *a priori/a posteriori* distinction is via perceptual or sensory experience.<sup>28</sup> Roughly, one has essentially *a posteriori* justification for believing that *p* if essentially one's warrant for believing that *p* rests at least partly on perceptual

---

<sup>27</sup>  $\Box A =_{\text{def}} \forall p (p \Box \rightarrow A)$ . Note that, given that one can define metaphysical necessity in terms of counterfactuals, the existence of certain counterfactual knowledge is also sufficient to show there must be some modal knowledge, and thus, that some explanation as to how we come to have modal knowledge must ultimately succeed. See Williamson (2005), Williamson (2007) ch. 5, Hill (2006).

<sup>28</sup> See, for example, Kitcher (1980) and (2000) for one attempt to make the *a priori/a posteriori* distinction along these lines.

or sensory experiences. One has some *a priori* justification for believing that *p* if one has justification that is not *a posteriori*. If the distinction is to be made out on these lines, it's hard to see how one's justification regarding thought-experiment intuitions would be in all cases *a posteriori*. No doubt coming to have a thought-experiment intuition may require perceptual or sensory experience in a variety of ways—for instance, one often has to listen to or read the thought-experiment text. But beliefs arrived at through deductive reasoning often show similar dependencies, and yet it's hard to feel that the *a priori* justification we generally associate with such reasoning processes is threatened by these dependencies. Whatever role perceptual and sensory experience plays in deduction, this role is not warranting, at least with respect to the apriority of the resulting beliefs.

Another way philosophers have tried to make the *a priori/a posteriori* distinction makes reference to one's ability to entertain a proposition. One has *a priori* justification for believing that *p* if and only if full grasp of the proposition *p* is concomitant with accepting it (or perhaps having an inclination to accept it) or *p* follows logically from other beliefs for which one has *a priori* justification.<sup>29</sup> If the distinction can be made out in this way, it would seem that many thought-experiment intuitions do indeed qualify for *a priori* justification. It does appear, for instance, that merely entertaining the propositional content given in the Gettier intuition is sufficient for having some inclination to believe it.

Of course, Williamson has recently argued against the possibility of making out the *a priori/a posteriori* in this way. He argues effectively for the thesis (T) that there is no unique set of cognitive/inferential skills and dispositions required to grasp the concepts of a particular proposition, so that a subject possessing the relevant concepts might lack any particular cognitive/inferential skills or dispositions vis-à-vis those concepts so long as the subject has other cognitive/inferential skills and dispositions vis-à-vis those concepts to compensate.<sup>30</sup>

In fact, Williamson contends on the basis of this thesis that although there is an intuitive distinction between the *a priori* and *a posteriori*, there is no principled distinction whatsoever here.<sup>31</sup> More specifically, he contends that there is no principled distinction here because there is no principled distinction as to when perceptual/sensory experience plays a warranting role rather than the role of merely enabling one to grasp the proposition in question.

Williamson's argument seems to be that since there are no particular cognitive/inferential skills that a subject need have in order to possess the relevant concepts, there's no saying which of those skills are merely enabling the subject to grasp the proposition in question and which are additional. As there's no principled distinction between the merely enabling and additional skills, he concludes there's no principled distinction between the perceptual/sensory experience that has gone towards acquiring the skills merely enabling one to grasp the proposition in question and the perceptual/sensory experience that has gone towards acquiring additional skills vis-à-vis the

---

<sup>29</sup> While they have not necessarily made the *a priori/a posteriori* distinction in this way, a number of prominent philosophers have tried to defend the *a priori* by showing that one's ability to entertain the proposition in question is sufficient for having the inclination to believe it. See, for example, Bealer (1999), Boghossian (1996), Peacocke (1993), and Sosa (2004).

<sup>30</sup> Williamson (2003), (2006), and especially chapter 4 of (2007).

<sup>31</sup> (2007), ch. 5.

proposition in question, and hence, that should count as playing a warranting role when those additional skills are executed.

Even accepting Williamson's thesis (T), we see no reason to accept his conclusion.<sup>32</sup> His step from (T) to the conclusion that there's no principled distinction between merely enabling and additional skills is not valid. (T) may well be true. Nonetheless, it may also be true that associated with any particular concept is some idealized set of cognitive/inferential skills, and moreover, it is by (jointly?) approximating these skills that members of a community possess the concept. Indeed, that at least some cognitive/inferential skills are required for fully possessing concepts roughly follows from our two truisms from Section VIII (which are, incidentally, compatible with (T)). Moreover, even Williamson apparently concedes that these skills must be *similar* in the sense that they dispose us to draw a large number of the same *token* inference, e.g. particular instances of *modus ponens*, regardless of whether they dispose us to follow the same *fully general* pattern of inferences, e.g. the general rule of *modus ponens*. That for any concept there is a special idealized set of cognitive/inferential skills that we must all approximate in order to possess that concept would effectively explain those two facts.

Given that for any concept there is a special idealized set of cognitive/inferential skills that any possible thinker must approximate in order to (fully) possess that concept, we may well make a principled distinction between those merely enabling and additional skills according to whether they contribute towards that approximation. There should be a principled distinction between those skills which drive us towards drawing the token conclusions that we would be driven to draw using only the special idealized set of skills and those that lead to additional conclusions.<sup>33</sup> Those skills that contribute to the approximation of the special idealized set of skills are merely enabling. The others are additional. As there is now a principled distinction between merely enabling and additional skills, there should also be a fairly principled distinction between when perceptual/experience goes towards acquiring merely enabling and when it goes towards acquiring additional skills. Thus, the principled distinction between perceptual/sensory experiences in their role as enabler versus provider of warrant is retained, so that the principled distinction between the *a priori* and *a posteriori* is retained as well.

Of course, this somewhat speculative response to Williamson is not fully developed. Nonetheless, it should give some indication that his dismissal of a principled *a priori/a posteriori* distinction is not mandatory; there is space for an alternative. We continue to maintain that the distinction is principled. Given this position, we see good reason to think that judgments like the Gettier intuition are paradigms of *a priori* knowledge.

## **XII. Apriority and the Quotidian**

A hypothetical objector might balk at our suggestion that thought-experiment intuitions are *a priori*, given our earlier explanation of their source of warrant. If the explanation for how

---

<sup>32</sup> Incidentally, we are not so fully convinced of (T), but we assume it for the sake of argument.

<sup>33</sup> The flexible individuation of cognitive/inferential skills is probably essential in that regard. We can think of the same person both as simultaneously having the skill of drawing valid instances of *modus ponens* and having the skill of drawing some more restricted set of the valid instances of *modus ponens*. This flexibility should allow one to isolate completely those skills which contribute towards approximating the idealized set of cognitive/inferential skills associated with a concept.

we are warranted in going about traditional philosophical inquiry through thought experiments is indeed the same as the explanation for how we are warranted in quotidian rational belief formation, why think that conceptual analysis is *a priori*? When someone tracks everyday actual instances of knowing or not knowing, they invoke *a posteriori* knowledge. Why think, then, that tracking Joe's knowing or not knowing results in *a priori* knowledge if the explanation for our ability to track knowing here is the same as for tracking everyday actual instances?

There are two mistakes our hypothetical objector makes in pursuing this tension. The first comes in thinking that somehow the conceptual role associated with a concept will produce only *a priori* or only *a posteriori* knowledge. This isn't the case. Take as an example the conceptual role of 'and', which is presumably exhausted by a disposition to make certain logical inferences. The deductive inferential patterns tied to 'and' will yield *a priori* conclusions when the inputs are *a priori* (say, tautologies) and *a posteriori* conclusions otherwise (when the inputs are say, pieces of knowledge about World War II). Whether carrying out the conceptual role associated with a concept leads to *a priori* or *a posteriori* knowledge depends on what the process uses as input—and in particular whether the warrant for these inputs ultimately depends upon the reliability of *a posteriori* conduits like perception. Knowing that your wife doesn't know that the car is parked in the garage requires being in a position to know that she exists. One's warrant for that proposition, however, comes through perceptual experience. As a result, this piece of knowledge is *a posteriori*. Production of the Gettier intuition, on the other hand, does not seem to take as an justificatory input anything similar whose warrant ultimately traces back to perceptual experience. (Indeed, production of the Gettier intuition doesn't seem to take anything whatsoever as *justificatory* input—whether a belief or perceptual or sensory experience.)

The second mistake he makes in pressing the above point, however, is thinking that all pieces of everyday knowledge must be *a posteriori*. Knowing that your wife doesn't know that the car is parked in the garage requires not only being in a position to know that she exists, but also being in a position to know that the mental state that her apparent behavior would seem to indicate she is in is not an instance of knowing that the car is parked in the garage. One's perceptual experiences allows you insight into your wife's mental states via her behavior, but in order for you to know that this mental state doesn't count as knowledge that the car is in the garage, you must be justified in categorizing that particular mental state as one which excludes such knowledge. One's warrant *here*, however, would not appear to rest on any perceptual experience whatsoever, nor would it seem to require anything more than grasping the concept *knowing*. As a result, it would appear that the *categorization* of the mental state in question—the one picked out via perception—as nonknowledge—is *a priori*.

Further, that the mental state was had on this occasion does not appear relevant to your warrant for believing that the particular mental state your wife is in falls short of knowledge at all. Thus, it would seem that your warrant should generalize so as to give warrant for believing, of any arbitrary case of that particular mental state type, that it was not an instance of knowing. In other words, it would appear you have warrant to believe that *necessarily*, that a state of that type is not an instance of knowing. The propositional content of this potential belief available from everyday interactions would seem to have a form very similar to the Gettier intuition. They both involve conditionals embedded in a necessity operator—we can think of them as roughly of the form: *necessarily, if the situation is like **that**, then it does not present an instance of knowing*. (Of course, the Gettier intuition goes further: it says that the situation presents an instance of

justified true belief.) In the everyday case, the demonstrative ‘that’ picks up its reference via perceptual experience. In the thought experiment case, ‘that’ picks up its reference via truth-in-fiction. In neither case does perceptual experience play a warranting relation—it merely provides resources for one to consider a certain kind of case. In both cases, merely mastering the use of ‘know’ puts one in a position rationally to discriminate between knowledge and non-knowledge. For these reasons, both would seem to be knowable *a priori*.

This parallelism exemplifies and underscores the earlier point. If one can give an explanation of how one is justified in believing in the everyday case, one should be able to offer a parallel explanation for thought experiments. In both cases, one needs to explain the connection between mastering ‘know’ and being in a position justifiably and rationally to discriminate cases of knowing. The only difference between how we form everyday judgments and judgments stemming from thought experiments is in how we come into contact with the cases we’re judging; this difference is incidental to the justification of our judgments on cases.

### **XIII. Conclusion**

We have here offered a sketch of a defense of traditional philosophical methodology and its use of thought experiments. Our defense relies, in the first place, on our demonstration that ordinary apprehension of truth-in-fiction can provide the means for grasping thought-experiment intuitions that are full-blooded true necessities—just as they have been traditionally conceived. Furthermore, there is no obvious obstacle to knowing such intuitions *a priori* on the account developed.

Although we disagree with Williamson on other counts, the view defended here fits well into what we take to be his broader project: that of ‘de-mystifying’ philosophical practices, grounding them in our ordinary knowledge about the actual world. Williamson writes that “[s]o-called intuitions involve the very same cognitive capacities that we use elsewhere.”<sup>34</sup> We agree. We do not see, however, that this offers compelling reason to abandon a traditional understanding of philosophical methodology that embraces both the apriority of thought-experiments intuitions and their form as judgments of necessity.<sup>35</sup>

### **References**

- Bealer, G. (1999). A Theory of the A Priori. *Philosophical Perspectives*, 33, 29-55
- \_\_\_\_\_. (2004). Modal Epistemology and the Rationalist Renaissance, in Gendler, T. and Hawthorne, J. (2004).
- Boghossian, P. and Peacocke, C., eds. (2000). *New Essays on the A Priori*. Clarendon Press.
- Boghossian, P. (1996). Analyticity Reconsidered, *Noûs*, 30.3 (1996), pp. 360-91.

---

<sup>34</sup> Williamson (2004), p. 152.

<sup>35</sup> We are grateful to Tamar Szabó Gendler, Richard Heck, Kelby Mason, Joshua Schechter, Peter Godfrey-Smith, Alvin Goldman, Deena Skolnick Weisberg, Ernest Sosa, John Turri, and Timothy Williamson for invaluable comments and discussion. An earlier version of this paper was presented at the 2006 Harvard-MIT Graduate Philosophy Conference; we are grateful for helpful discussion from the conference participants.

- \_\_\_\_\_. (2003). Blind Reasoning, *The Aristotelian Society Supplementary Volume*, 77.1 (2003), pp. 225-48.
- Bonjour, L. (1998). *In Defense of Pure Reason*. Cambridge University Press.
- Chalmers, D. (2004). Does Conceivability Entail Possibility?, in Gendler T. and Hawthorne, J. (2004).
- Currie, G. (1990). *The Nature of Fiction*. Cambridge University Press
- DePaul, M. and Ramsey, W., eds. (1998). *Rethinking Intuition*. Lanham. Rowman & Littlefield.
- Fodor, J. (1998). *Concepts*. Clarendon Press.
- Gendler, T. (2002). Personal Identity and Thought Experiments. *The Philosophical Quarterly*, 52, no. 206 (2002), pp. 34-54, January 2002.
- Gendler, T. and Hawthorne, J., eds. (2004). *Conceivability and Possibility*. Oxford University Press.
- Greenberg, M. and Harman, G. (2006). Conceptual Role Semantics in Lepore, E. and Smith, B., eds., *The Oxford Handbook of Philosophy of Language*, Oxford University Press, 2006.
- Hanley, R. (2004). As Good as it Gets. Lewis on Truth in Fiction, *Australasian Journal of Philosophy*, 82 (2004), pp. 112-128, March 2004
- Hill, C. (2006). Modality, Modal Epistemology, and the Metaphysics of Consciousness in Nichols, S., ed., *The Architecture of the Imagination. New Essays on Pretense, Possibility, and Fiction*. Oxford. Oxford University Press.
- Kitcher, Philip (1980). A Priori Knowledge, *The Philosophical Review*, 89.1 (1980), pp. 3-23.
- \_\_\_\_\_. (2000). A Priori Knowledge Revisited in Boghossian, P. and Peacocke, C. (2000).
- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
- Lewis, D. (1978). Truth in Fiction, *American Philosophical Quarterly*, 15 (1978), pp. 37-46. Reprinted in Lewis, D. (1983).
- \_\_\_\_\_. (1983). *Philosophical Papers, Vol. I.*, Oxford University Press.
- \_\_\_\_\_. (1984). Putnam's Paradox, *Australasian Journal of Philosophy*, 62 (1984), pp. 221-36.
- Peacocke, C. (1993). How Are A Priori Truths Possible? *European Journal of Philosophy*, 1.2 (1993), pp. 175-99.

- Skolnick, D., and Bloom, P. (2006a). The Intuitive Cosmology of Fictional Worlds, in Nichols, S., ed., *The Architecture of the Imagination. New Essays on Pretense, Possibility, and Fiction*. Oxford. Oxford University Press.
- \_\_\_\_\_. (2006b). What does Batman think about SpongeBob? Children's understanding of the fact/fiction distinction, *Cognition*, 101(1), B9-B18.
- Sosa, E. (2004). Reliability and the A Priori, in Gendler, T. and Hawthorne, J., eds., *Conceivability and Possibility*. Oxford University Press.
- Walton, K. (1990). *Mimesis as Make-Believe*. Harvard University Press.
- Weatherson, B. (2003). What Good are Counterexamples?, *Philosophical Studies*, 115 (2003), 1-31.
- Williamson, T. (2003). Understanding and Inference, *The Aristotelian Society Supplementary Volume*, 77.1 (2003), pp. 249-93.
- \_\_\_\_\_. (2004). Philosophical 'Intuitions' and Scepticism About Judgment, *Dialectica* 58.1 (2004), pp. 109-153.
- \_\_\_\_\_. (2005). Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking, Presidential Address, *Proceedings of the Aristotelian Society* 105 (2005). 1-23
- \_\_\_\_\_. (2006). Conceptual Truth, *The Aristotelian Society Supplementary Volume*, 80 (forthcoming).
- \_\_\_\_\_. (2007). *The Philosophy of Philosophy*, (forthcoming).
- Yablo, S. (1993). Is Conceivability a Guide to Possibility?, *Philosophy and Phenomenological Research* 53 (1993). 1-42.
- \_\_\_\_\_. (2004). Coulda, Woulda, Shoulda, in Gendler, T. and Hawthorne, J., eds., *Conceivability and Possibility*. Oxford University Press.