

# Rationalism and the Content of Intuitive Judgements

ANNA-SARA MALMGREN

University of Texas at Austin  
annasara@austin.utexas.edu

It is commonly held that our intuitive judgements about imaginary problem cases are justified a priori, if and when they are justified at all. In this paper I defend this view — ‘rationalism’ — against a recent objection by Timothy Williamson. I argue that his objection fails on multiple grounds, but the reasons why it fails are instructive. Williamson argues from a claim about the semantics of intuitive judgements, to a claim about their psychological underpinnings, to the denial of rationalism. I argue that the psychological claim — that a capacity for mental simulation explains our intuitive judgements — does not, even if true, provide reasons to reject rationalism. (More generally, a simulation hypothesis, about any category of judgements, is very limited in its epistemological implications: it is pitched at a level of explanation that is insensitive to central epistemic distinctions.) I also argue that Williamson’s semantic claim — that intuitive judgements are judgements of counterfactuals — is mistaken; rather, I propose, they are a certain kind of metaphysical possibility judgement. Several other competing proposals are also examined and criticized.

## 1.

### 1.1 Introduction

What demarcates philosophy from other academic disciplines, specifically from (other) sciences? One striking difference is that in philosophy we typically do not subject our hypotheses and theories to empirical testing — somehow it is supposed to be sufficient to test a theory in *thought*. Where the chemist sets up a lab experiment and the sociologist conducts a survey, the philosopher sits back and runs a thought experiment.<sup>1</sup> How could that be enough? Indeed, how could an experiment performed in thought tell us *anything* about the nature of knowledge, consciousness, time, moral value, or any of the other things that philosophers are interested in?

<sup>1</sup> Thought experiments are used in other disciplines too, but there they have a much less central role (and arguably a different structure). In this paper I am exclusively concerned with ‘philosophical’ thought experiments.

To get clear on this, we first of all need a model of the relevant test procedure—the thought experiment, as used in philosophy. In brief outline, it has the following structure: the hypothesis or theory that is under evaluation states or entails some modal claim (typically a necessary bi-conditional or one-way implication) and in a thought experiment we check that modal claim against our intuitive verdict on an imaginary problem case. If the claim conflicts with our intuitive verdict, this is treated as strong evidence against the theory—indeed the theory may be abandoned as a result. We say that we found a *counterexample* to it. If not, this is treated as at least some evidence in support of the theory. We say that it *accommodates our intuitions* about the case.

As familiar as this method is (the ‘method of cases’), we still lack a good account of how it works—a good explanation of how, if at all, it provides the presumed evidence. In particular, we lack a plausible epistemology for our intuitive judgements about cases. The success of the method seems to depend on the epistemic status of these judgements: other things equal, the method cannot supply us with a justified and reliably true ‘output’ belief (i.e. a belief that the theory under evaluation is true/false) unless our intuitive judgements about the given problem case are both reliable and justified in turn.<sup>2</sup> But there is no good explanation available of how they *could* be—of what, if anything, makes such judgements justified and what, if anything, makes them reliable.

Take my intuitive judgement about one of Edmund Gettier’s well-known problem cases—loosely put: the judgement that Smith has a justified true belief without knowledge (Gettier 1963); my intuitive judgement about one of Hilary Putnam’s ‘Twin-Earth cases’—loosely put: the judgement that Oscar<sub>1</sub> and his Twin Earthian counterpart Oscar<sub>2</sub> do not mean the same thing by ‘water’ (Putnam 1975); or my intuitive judgement about one of Judith Jarvis Thomson’s ‘Trolley cases’—loosely put: the judgement that George may not shove the fat man onto the track, even though doing so would

<sup>2</sup> It is plausible that what goes for justification and reliability goes for knowledge too: that the method only supplies us with knowledge (that the target theory is true/false) if our intuitive judgements constitute knowledge in turn. But nothing here turns on this. Another issue on which nothing turns is whether belief is gradable. For convenience I mostly write in terms of outright belief (and outright justification) but the important points can all be restated in terms of partial belief. Last, nothing turns on the exact relationship between reliability and justification—I do distinguish the question what makes intuitive judgements justified from the question what makes them reliable, but I leave open that the answers to these questions may be closely related (or even coincide).

bring down the death toll (Thomson 1973). In virtue of *what* are judgements like these justified, and how—by what mechanism—do they ‘track’ what goes on in the respective problem cases? Without good answers to these questions, our heavy reliance on intuitive judgements can come to seem like a mere matter of faith. This is especially problematic in light of the mounting sceptical literature on thought experiments.<sup>3</sup>

The focus in this paper is on the question of justification, indeed on a specific type of answer to it: that intuitive judgements are justified *a priori*—roughly, independently of experience—to the extent that they are justified at all. Let us call this view ‘rationalism’. (But note that you can be a rationalist in this sense and still be a sceptic about the *a priori*.) It is clearly the received view among participants in the current debate: both proponents and critics of the method of cases tend to hold that intuitive judgements could not be empirically justified.<sup>4</sup> Rationalism is also integral to a common conception of philosophical methodology—as distinctively *a priori*.<sup>5</sup> And it is a view with a great deal of *prima facie* plausibility: it is very hard to see how empirical considerations could help me determine the question at issue in a thought experiment—say, whether Smith has a justified true belief without knowledge. What empirical evidence or grounds could I possibly have for passing one verdict rather than another on the given problem case?

This question is particularly pressing on the assumption that intuitive judgements ‘really’ express claims of metaphysical necessity or possibility. (More on this shortly.) But the basic puzzle does not

<sup>3</sup> Just to mention a few contributions: Cummins 1998; Devitt 1994; Hintikka 1999; Kornblith 2002, 2005, 2006; Machery, Mallon, Nichols, and Stich 2004; Stich 1988; Weinberg, Nichols, and Stich 2001; Nichols, Stich, and Weinberg 2003; Swain, Alexander, and Weinberg 2008; Unger 1983, 2002. For some recent attempts at providing a non-sceptical epistemology for intuitive judgements, see Goldman 2007; Goldman and Pust 1998; Graham and Horgan 1998; Margolis and Laurence 2003; Weatherson 2003; and the first set of references in the next footnote. (I lack the space to criticize those attempts individually here.)

<sup>4</sup> Rationalist proponents include Bealer 1998, 2000, 2002; BonJour 1998; Jackson 1998; Pust 2000; Sosa 2007a, 2007b, 2009. Rationalist critics include Devitt 1994; Fodor 1989; Kornblith 2002, 2005, 2006.

<sup>5</sup> The thought is this: ‘the crucial contrast between philosophical methodology and scientific methodology lies in the nature and role of thought experiments. Not only do philosophers rely on thought experiments to a larger extent than scientists do, but the *kind* of thought experiment they use can deliver *a priori* knowledge, whereas the kind that scientists use can only deliver empirical knowledge.’ This line of thought presupposes that the intuitive judgements that figure in philosophical thought experiments are capable of being *a priori* justified (since otherwise those experiments too would at best deliver empirical knowledge).

depend on that assumption—there just does not seem to be any empirical evidence at hand that could do the job, even if we take these judgements at face value. For one thing, it is highly implausible that my visual or auditory perception of a description of the case is (or provides) even part of my justification for believing that Smith has a justified true belief without knowledge—at most the perceptual state enables me to *access* a justification.

We could of course describe a route such that, if I *were* to take that route, I *would* end up with an empirically justified belief about the case—say, if I relied on good testimony, or on induction from past experience with relevantly similar problem cases. But for present purposes the route must also be *available* to me—qua thought experimental subject—and the testimonial and the inductive route are typically not (in fact we try to screen for them).<sup>6</sup>

The above consideration constitutes a strong *prima facie* case for rationalism. But it does not yet support *non-sceptical* rationalism: for all that has been said, intuitive judgements are never justified, a priori or otherwise. To my mind, the main challenge facing non-sceptical rationalism is to explain *what it is* (or what it would be) for intuitive judgements to be justified independently of experience: to give a positive account of the purported a priori justification. This, in turn, is an instance of a perfectly general challenge: to fully vindicate the claim that some category of judgements are capable of being a priori justified, we must explain what it is for such judgements to be a priori justified. (I do not attempt to meet this challenge here.) Second, a *prima facie* case is only *prima facie*—but a *prima facie* case stands in the absence of good reasons to the contrary.

In this paper I defend rationalism against a recent objection, due to Timothy Williamson, that threatens to undermine the *prima facie* case. In the course of so doing, I discuss the formal structure of thought experiments in more detail; in particular, how to analyse intuitive judgements—what their ‘real’ content is. To give a brief preview of the issue: intuitive judgements (about imaginary problem cases) appear to commit us to the actual existence of certain objects and states of affairs that we do not believe to be actual. Hence—on pain of having all these judgements come out untrue—it seems we cannot take them at face value. But that is to say that we are not *really*

<sup>6</sup> If, say, it turned out that my judgement were based on someone else’s testimony that would normally—and on the face of it rightly—be taken to invalidate the thought experiment. (Likewise for the envisaged inductively based belief.)

asserting what we seem to be asserting when we utter a sentence like ‘Smith has a justified true belief but does not know’. What, then, are we asserting (if indeed we are asserting anything)?

Williamson defends a particular answer to this question — that we are asserting a certain subjunctive conditional — and uses that to argue against rationalism. I will question both his answer and the use he makes of it. But notice that the question is of interest quite apart from its implications, if any, for rationalism. The correct answer is likely to at least constrain the range of available accounts of what makes intuitive judgements reliable and justified. It is also crucial to understanding the role of such judgements in thought experiments: *why* their status matters to the status of the output beliefs. A promising initial thought is that they matter because the output belief is *inferred* from the intuitive judgement — the content of that judgement serves as a premiss in an inference to the conclusion that the theory under evaluation is true/false. What could the content be, such that the judgement is able to confer justification on the output belief? (This is one among several distinct concerns that will guide my discussion of the ‘content problem’.<sup>7</sup>)

A final preliminary remark: I have set up the issue in terms of intuitive *judgements*, rather than *intuitions*. This is simply to avoid confusion, and to secure an uncontroversial starting point. The term ‘intuition’ is often used to denote a special type of mental state, a conscious representational state with a ‘quasi-perceptual’ phenomenology — a state not reducible to ordinary propositional attitudes such as belief (including occurrent belief, i.e. judgement). It is frequently argued, or just taken for granted, that states of this sort play a crucial role in the method of cases; that they are used as the primary evidence ‘against which candidate philosophical theories are evaluated’

<sup>7</sup> On an alternative approach, this problem is better described as an ‘attitude problem’: the challenge is to explain what we are really *doing* — not what we are really saying or what belief we are really expressing — when we utter ‘Smith has a justified true belief but does not know’. (See e.g. Eagle 2007; Nichols 2004; Walton 1990; Yablo 2001.) Details aside, the key idea is that our intuitive ‘judgement’ is not a genuine judgement or belief at all, but some other cognitive attitude — one that does not carry the problematic existential commitments of a genuine belief (perhaps it is a *pretend-* or a *suppositional* judgement). But I mention this approach just to set it aside — interesting as it may be, it raises some large and difficult issues that I lack the space to comment on here. (See e.g. Davies and Stone 1998, 2001; Nichols, Stich, Leslie, and Klein 1996; Stich and Nichols 1998; Stanley 2001.) Moreover, it is hard to see how the choice between these two approaches (to the content/attitude problem) could substantially affect the result of our overarching epistemological inquiry. For a related discussion, see Sect. 3 below.

(Pust 2000, p. 11).<sup>8</sup> On that assumption, the natural way to frame the debate is to ask how, if at all, intuitions (thus conceived) could be evidence and what, if anything, makes them reliable. But this framework excludes too many options at the start: in effect, it excludes any non-sceptical account of the method of cases on which there are no intuitions, in the intended sense, and/or on which intuitions do not play the designated epistemic role. The present framework allows for a broader range of options — including accounts on which intuitions *do* play that role. Such accounts are best regarded as attempts to *answer* the justification and reliability question about intuitive judgements, not as part of some alternative approach on which these questions lack importance.<sup>9</sup>

A more complicated issue concerns how to demarcate the relevant class of judgements or beliefs. What *counts* as an intuitive judgement, for the purpose of this inquiry? On the face of it, we have five broad options: we can identify the intended judgements in terms of their propositional content, their causal antecedents, their epistemic ground, their role in the method of cases, or by reference to examples. The last option seems to me vastly preferable, given the desire for a neutral starting point — all the others are bound to be controversial, and to limit the theoretical possibilities too much in advance. For our purposes, then, an intuitive judgement is any judgement relevantly similar to certain paradigms or examples (where it is left open what exactly makes for relevance); for instance, my judgement about Gettier's Smith and my judgement about Oscar<sub>1</sub> and his twin.

### 1.2 An argument against rationalism

Williamson's argument is of particular interest because it seems to be an argument 'from on high' — an argument that rests on *principled* considerations against rationalism. (Contrast the strategy of assessing specific rationalist explanations one by one and arguing that none of

<sup>8</sup> Other people who take this approach include Bealer (1996, 2000), Kagan (2001), Sosa (2007a, 2007b).

<sup>9</sup> To see this, consider what so-called 'intuitions' are most plausibly regarded as (direct) evidence *for*. Surely the only viable proposal is that an intuition that *p* is evidence that *p*. But on this understanding of the view, it is clear that it is just a candidate account of the epistemology of intuitive judgements: a candidate explanation of what makes such judgements justified and reliable. This also seems to be the best way to understand the view that beliefs *about* our intuitions — or about 'what we would say' — are the primary evidence (against which philosophical theories are evaluated).

them work.) It is an argument, then, that threatens to rule out *any* attempt to construe intuitive judgements as a priori. (At least, any ‘theoretically interesting’ attempt—see below.) That result would considerably simplify our discussion: it would rid us of the great majority of extant accounts of intuitive judgements, and it would provide us with some solid guidelines in our search for the correct account.

Williamson’s basic idea is that intuitive judgements are a certain kind of counterfactual judgement—judgements of certain counterfactuals. This is supposed to dispense with the need for a special causal-psychological explanation of intuitive judgements, but without a special causal-psychological explanation there is no way to sustain (any interesting form of) rationalism.

More precisely, he argues that the content of an intuitive judgement is a certain counterfactual or subjunctive conditional (step 1). He writes as if this entails that our capacity to make intuitive judgements is just ‘an application of our general cognitive capacity to handle counterfactuals’ (2). That, in turn, is supposed to show that these judgements are not formed by means of any special-purpose capacity or mechanism—for example, a faculty of rational intuition (3). Some rationalists invoke such a faculty to explain how intuitive judgements are formed, and what makes them reliable and justified.<sup>10</sup> Williamson’s point is that the motivation for this move falls away once we accept that intuitive judgements are counterfactual judgements (albeit in disguise): there is *no need* to postulate a faculty of rational intuition—or any other special psychological machinery—to account for the existence and the epistemic status of intuitive judgements, since there is a perfectly ordinary cognitive capacity already in place to do the job: our general capacity to handle counterfactuals.

Next, he argues that this general capacity is not ‘exclusively a priori’ (4), by which he just means that not every justified judgement that the capacity delivers is a priori justified. This should be uncontroversial—many, perhaps most, of our counterfactual judgements are not capable of being a priori justified (e.g. my judgement that if I had made the supper it would have been inedible). But he then goes on to argue that there is no principled way to single out even *some* such judgements as a priori—at any rate, there is no such way that ‘cuts at the cognitive joints’ (5). Last, he claims that any a priori/a posteriori distinction that fails to cut at these joints is of little

<sup>10</sup> See e.g. Gödel 1964, and—more recently—BonJour 1998; Brown 1991; Eagle 2005; Katz 1998.

theoretical interest (6). Ergo: there is no (theoretically) interesting sense in which intuitive judgements are a priori (7).<sup>11</sup>

In summary:

- (1) Intuitive judgements are counterfactual judgements. [Premiss]
- (2) The cognitive capacity by which intuitive judgements are formed is ‘simply an application of our general cognitive capacity to handle counterfactuals’.<sup>12</sup> [From 1]
- (3) Intuitive judgements are not formed by means of any special-purpose capacity or mechanism—e.g. a faculty of rational intuition. [From 2]
- (4) Our general capacity to handle counterfactuals is not ‘exclusively a priori’. [Premiss]
- (5) There is no principled way to single out only some of the judgements this capacity delivers as a priori—at least there is no way that ‘cuts at the cognitive joints’. [Premiss]
- (6) Any surviving a priori/a posteriori distinction is of little theoretical interest. [Premiss]
- (7) There is no interesting sense in which intuitive judgements are a priori. [From 3–6]<sup>13</sup>

Unless specified, I will be working with a loose but standard notion of a priori justification— as justification that is suitably independent of the subject’s perceptual experience.<sup>14</sup> One might wonder what it takes

<sup>11</sup> A qualification is in place: Williamson does not affirm that intuitive judgements are a posteriori either. The exact claim is that there is no principled (and theoretically interesting) way to draw an a priori/a posteriori *distinction* among our counterfactual judgements— and so perhaps he is better represented as saying that intuitive judgements are *neither* a priori nor a posteriori. (See, in particular, his 2007a, Sect. 3, and 2007b, Ch. 5. See also the discussion in Sect. 3.5 below.) But for our purposes this does not matter: I take it that, if intuitive judgements are neither a priori nor a posteriori, then they are not a priori.

<sup>12</sup> Williamson 2005, p. 1. The other quoted phrases here are from the same paper.

<sup>13</sup> This is my reconstruction of the argument— as it appears in Williamson 2005 and 2007a. The most recent version looks slightly different (see his 2007b, Chs 5–6). But Williamson has agreed, in conversation, that my reconstruction captures the original version of the argument. And the bulk of my criticism applies to the present version too.

<sup>14</sup> Nothing of importance to my arguments here hinges on how that notion is best precisified; e.g. on whether ‘perceptual experience’ should be taken to include introspection, conceptually rich conscious experience, or only sensory experience.



for a notion of a priority to be of ‘theoretical interest’ (see step 6), but it would take us too far afield to discuss that here. And there is no pressing need to do so, since our main focus will be on the first three steps of the argument. More precisely, my agenda is this: in what remains of the first section, I clarify step 1, and explain the related problematic; then I argue that 1 is false — that intuitive judgements are not counterfactual judgements — and I go on to propose an alternative analysis. In section 2, I defend that analysis against a number of objections and competing views (e.g. that they are metaphysical necessity judgements of a certain kind, and that they are fictional judgements). In section 3, I return to Williamson’s argument and argue, on independent grounds, that there is no plausible reading of step 2 on which it supports 3: that we do not have a ‘general capacity to handle counterfactuals’ of the sort that is needed for the argument to go through. This objection also blocks a potential fallback manoeuvre — a structurally similar argument against rationalism that does not rely on the contentious step 1 (indeed, one that is compatible with the analysis that I propose). At the very end of the paper, I briefly discuss some of the considerations that Williamson gives in support of step 5. One upshot of that discussion is that the argument may not, after all, be an argument from on high.

### *1.3 The content problem*

As mentioned above, the viability of the method of cases seems to depend on the status of our intuitive judgements: the method only provides me with a justified output belief to the extent that my intuitive judgement is justified in turn; likewise for reliability. A natural explanation of this is that the method involves an inference — roughly, an inference from my intuitive verdict on the given problem case to the truth of the theory under evaluation (call this a ‘positive’ thought experiment) or to its falsity (a ‘negative’ experiment). It is time to consider in some detail what this inference might look like. Or rather — since positive and negative thought experiments require separate treatment, and we cannot do everything at once — to consider what the inference that is involved in a negative experiment might look like.<sup>15</sup> Throughout I will work with a concrete example: a negative thought experiment of the sort originally designed by Gettier.

<sup>15</sup> A positive experiment presumably involves an inference to the best explanation, but exactly what kind is a difficult question.

In asking what this inference might look like, I am asking what form of argument the relevant piece of reasoning exemplifies (or perhaps, what form it best approximates). It may still be misleading to talk of one inference here—even once we bracket positive experiments. There may be no unique argument form common to all negative thought experiments, or even to all tokens of the same such experiment. More cautiously, then, the aim is to capture an argument form that is common to at least a core set of negative experiments. (I treat this as understood in what follows.) That core set, in turn, only includes successful thought experiments—experiments that, other things equal, do provide the thought experimenter with a justified and reliably true output belief.<sup>16</sup> The argument form of interest, then, is the argument form our reasoning exemplifies in so far as it results in a justified and reliably true belief that the theory under evaluation is false.

Let the ‘Gettier case’ be a specific problem case, a case we might describe as follows:

Suppose that Smith believes that Jones owns a Ford, on the basis of seeing Jones drive a Ford to work and remembering that Jones always drove a Ford in the past. From this, Smith infers that someone in his office owns a Ford. Suppose furthermore that someone in Smith’s office does own a Ford—but it is not Jones, it is Brown. (Jones’s Ford was stolen and Jones now drives a rented Ford.)

Let the ‘Gettier judgement’ be the intuitive judgement that I (and many with me) would make about this case, if asked the appropriate question—a judgement that we might express by saying: ‘Smith has a justified true belief, but does not know, that someone in his office owns a Ford.’ And let the ‘Gettier inference’ be the inference by which we get from this judgement to the belief that the target theory—the theory that knowledge is justified true belief—is false.

On standard ways of understanding that theory (‘the JTB theory’) it entails a metaphysical necessity claim: that, necessarily, a subject knows that *p* if and only if she has a justified true belief that *p*. As a result of Gettier’s thought experiment, we somehow become justified in believing that this claim is false. But how? A first step towards an answer is to say that, by reflecting on the Gettier case, we become justified in believing that some particular person—namely Smith, the person featured in the case—has a justified true belief but does not

<sup>16</sup> So as not to beg the question against a sceptic, we might qualify this by saying that our concern is with successful experiments, *if any*. (This too will be treated as understood.)

know, and from this we can rationally infer that it is possible that someone does (which transparently contradicts the necessity claim). But of course, this is only a first step: the question of justification has simply been pushed back—it instantly reappears as the question how reflection on the case could justify us in believing that Smith has those properties. Furthermore, the suggested inference is only sound if the given problem case is actual, and arguably only warrant-transferring if known to be actual,<sup>17</sup> but the problem cases used in philosophy are usually hypothetical, and known to be so—including (let us suppose) our sample case.

This is not to deny that an actual problem case would do just as well as—perhaps even better than—a merely possible case, when it comes to testing a philosophical theory: we could easily bring about an actual Gettier-style problem case, and our intuitive judgement about it would have at least as much weight as our judgement about the corresponding hypothetical case.<sup>18</sup> But the problem cases that are used in philosophy are *normally* hypothetical—or at least not normally known to be actual. (We are, after all, talking about *thought* experiments.) This much is part of the data about the method of cases as currently practised, hence it is something that we should aim to capture in our representation of it. (More on the significance of actual cases later on.)

It is highly implausible, then, that reflection on the case gives us justification to believe that some particular, actual person has a justified true belief but does not know; nor is it plausible that this is what we (unjustifiably) come to believe as a result of such reflection—we are not deluded about the hypothetical nature of the case; we know that the name ‘Smith’ has no actual referent. On the face of it, however, we are sincerely asserting *something* when we say ‘Smith has a justified true belief but does not know’—but *what?* What is the (‘real’) content of our intuitive judgement?<sup>19</sup>

One line of response is that it is some kind of modal conditional—roughly speaking: that it is a generalization over some range of

<sup>17</sup> Or at least: justifiably believed to be actual. (Nothing here hinges on the difference.)

<sup>18</sup> As noted by Williamson; see his 2005, p. 15, and 2007b, p. 192.

<sup>19</sup> Perhaps there is even more at stake than the *truth* of the judgement: if *that Smith has a justified true belief but does not know* is an ‘object-dependent’ content, we cannot even *think* that content unless there exists a particular object for ‘Smith’ to refer to. (See e.g. Evans 1982; McDowell 1977; Recanati 1993; Russell 1905.) Since there is a problem here in either case, I prefer to set it up in a way that does not rely on this account of the surface content. (But I leave open that it may be the correct account.)

possible realizations of the Gettier case. What we ‘really’ judge is that, in any such realization, someone has a justified true belief but does not know. The proposal that Williamson attributes to his opponent — whom we might call the ‘target rationalist’ — falls into this category of responses (as does his own). The target rationalist identifies the content with a certain strict conditional — a generalization over all metaphysically possible realizations of the case, under a given description. That is, what we really judge is that,

NECESSITY Necessarily, anyone who stands to a proposition  $p$  as in the Gettier case (as described) has a justified true belief that  $p$  but does not know that  $p$ .

From this alone we cannot rationally infer that the JTB theory is false. But given the additional premiss that the case thus described *could* be realized — that it is possible that someone stand to  $p$  in the specified way — there is a straightforward entailment to the intended conclusion.<sup>20</sup>

It is worth emphasizing that NECESSITY is a claim about all subjects who stand to a proposition as in the Gettier case *under a given description* — all possible subjects who satisfy a standard or canonical description of the case. What does such a case description look like?

For one thing, it contains a lot of concrete detail: that there is an office, with at least three office workers, that one of the office workers recently saw another drive a Ford to work, etc. But it leaves out many more — by no means does it specify a *complete* possible world or situation. Hence there may be many different (internally consistent) ways of ‘filling out’ the description. This will prove important soon. Second, it exhibits a certain *neutrality* — it does not specify that the subject does, or does not, know the relevant proposition; nor does it specify that she does, or does not, have a justified true belief in it.

<sup>20</sup> On this view, then, the overall inference can be represented like this:

$K(x, p)$   $x$  knows that  $p$   
 $JTB(x, p)$   $x$  has a justified true belief that  $p$   
 $GC(x, p)$   $x$  stands to  $p$  as in the Gettier case as described

- (a)  $\Box \forall x \forall p (GC(x, p) \rightarrow (JTB(x, p) \ \& \ \neg K(x, p)))$
- (b)  $\Diamond \exists x \exists p GC(x, p)$
- (c)  $\Diamond \exists x \exists p (JTB(x, p) \ \& \ \neg K(x, p))$
- (d)  $\neg \Box \forall x \forall p (K(x, p) \leftrightarrow JTB(x, p))$

(But note that it *does* specify that she truly believes it.) More generally, a canonical description of a problem case is neutral with respect to the distribution of the ‘test properties’ — the properties whose modal connection is at issue in the thought experiment: it does not explicitly stipulate that the test properties are or that they are not instantiated in the given case (nor does it stipulate anything that transparently entails that they are/are not instantiated). This is very rough, but it suffices for now.

#### 1.4 Deviant realizations I

Williamson argues that mere reflection on the Gettier case does not give us justification to believe anything as strong as NECESSITY, and, moreover, that NECESSITY is ‘quite probably false.’ As mentioned above, a standard case description is radically incomplete, and there may be ways of completing it on which the subject does *not* have a justified true belief without knowledge. In fact it is quite easy to complete it in some such ways — for example, to describe a possible realization of the case in which the subject’s true belief is not justified. Here is one such way: suppose that Smith has good reason to believe that he tends to hallucinate people driving Fords to work, and to believe that he tends to misremember what cars people drove in the past (and so on for any other piece of evidence that is specified in the given description). Other things equal, Smith’s (prima facie) justification is here defeated — hence his belief is not a counter-instance to the JTB theory.<sup>21</sup>

Williamson does not say so, but it is equally clear that the description can be completed in ways such that the subject *does* know the relevant proposition, namely by some *other* means or route — some route not specified in the description. For instance, suppose that Smith has independent testimonial justification (of a strength sufficient for knowledge) to believe that someone in his office owns a Ford. Suppose further that this justification is not in turn defeated and not itself ‘Gettierized’ — then Smith knows by testimony that someone in his office owns a Ford (and so, again, his belief is not a counter-instance).

These realizations are not ruled out by anything that is specified in a canonical description of the case. But then NECESSITY is false.

<sup>21</sup> My example differs from Williamson’s own, but the basic idea is the same. (See his 2005, p. 8, and 2007b, p. 185.)

For related reasons, it is hard to see how mere reflection on the case—under some such description—could give us justification to believe NECESSITY.

Someone might object as follows: ‘Why is it a problem for the target rationalist that the judgement she ascribes to us is false and unjustified? Surely it is still an open question at this stage of the inquiry whether any given intuitive judgement is true and/or justified? Careful analysis of its content may reveal that it fails on both scores.’

Let me reply on Williamson’s behalf: as I understand it, the problem is not simply that the candidate content is false and/or unjustified—the problem is that it is *too obviously* false and unjustified. Our intuitive judgements may indeed be in poor epistemic shape, but it would be highly surprising if it turned out to be this easy to reveal that they are. Indeed, it would be surprising enough if it were this easy to show that the Gettier judgement alone was in poor shape—the usual presumption being that, if *any* intuitive judgement is in good standing, it is the Gettier judgement. But matters are worse, since parallel arguments are available for many other intuitive judgements.<sup>22</sup>

Next, the proposal under consideration is a psychological hypothesis about the content of our actual judgement, and as such it makes certain predictions that are not borne out. We simply do not behave as we would behave if the proposal were true. In particular, we do not take the possibility of the above realizations of the case to falsify our original intuitive judgement. And on the face of it, this reaction is appropriate—it is not just a result of stubbornness or a reluctance to admit mistakes. Rather, the reason why we do not retract or revise our original judgement is that the envisaged realizations do not agree with our (semantic) intuitions about what the case designer *meant*. (By ‘case designer’ I just mean whoever put together the case description. The case designer and the thought experimenter may of course coincide.) These realizations of the case are *deviant*, and they are easily recognizable as such. That is, the corresponding interpretations of the case description are clearly *unintended*—the description was not

<sup>22</sup> Take my intuitive judgement about a certain Trolley case (Thomson 1973)—a judgement that I might express by saying: ‘George may not shove the fat man onto the track, even though doing so brings down the death toll.’ It is consistent with everything that is stated in the given case description that, by shoving the fat man on to the track, George would cause a riot in which fifty people die. Or take my intuitive judgement about a certain Twin-Earth case (Putnam 1975)—a judgement that I might express by saying: ‘Oscar<sub>1</sub> and his physical duplicate Oscar<sub>2</sub> do not refer to the same thing by “water”’. It is consistent with the given description that Oscar<sub>2</sub> has been travelling slowly back and forth between Earth and Twin-Earth... (See Burge 1988, Boghossian 1989, on ‘slow switching’.)

meant to be read in some such way—and we rightly ignore these interpretations when running the thought experiment. And when it is pointed out to us that it *can* be filled out in some such way, we do not retract our original intuitive judgement or start qualifying it (e.g. by adding ‘what I should have said is that Jones does not know—*provided* that there is no other route to knowledge available to him’).<sup>23</sup> This suggests that NECESSITY fails to capture the intended generality of that judgement.

Rather, Williamson argues, the intended generality is that of a certain counterfactual conditional—in terms of possible worlds, we are generalizing over all ‘nearby’ realizations of the case, under a given description. That is, what we really judge is that,

COUNTERFACTUAL If someone were to stand to  $p$  as in the Gettier case (as described), then she would have a justified true belief that  $p$  but not know that  $p$ .<sup>24</sup>

### 1.5 Deviant realizations II

To recapitulate, the current puzzle concerns the content of our intuitive judgements, as exemplified by the Gettier judgement. The discussion so far has been guided by the thought that the real content supports—or at least, that it can reasonably be *taken* to support—the negation of the necessary bi-conditional entailed by the JTB theory.<sup>25</sup> The ‘surface’ content of our judgement—*that Smith has a*

<sup>23</sup> Suppose that someone did withdraw her judgement in the face of a clearly deviant realization of the sort outlined above. Other things equal, the natural explanation of this would be that she had not in fact understood the example.

<sup>24</sup> Williamson formalizes this as follows:  $\exists x \exists p GC(x, p) \square \rightarrow \forall x \forall p (GC(x, p) \rightarrow (JTB(x, p) \& \neg K(x, p)))$ . As he himself notes, it is debatable whether this is the right way to represent COUNTERFACTUAL; on the face of it, the string says something quite different—namely, that, if someone stood to a proposition as in the case as described, then *anyone* who stood that way to it would have a justified true belief without knowledge. But it is not obvious how to do better. This is an instance of a certain general technical problem in semantics—the problem of so-called ‘donkey anaphora’. (See Geach 1962.) However, my objection to Williamson is independent of this problem: it concerns whether the counterfactual claim expressed by the relevant English sentence adequately represents the real content of our intuitive judgement in the first place. For my purposes, then, the informal statement of the proposal will do, and I will not discuss the problem of donkey anaphora any further. (Williamson discusses the problem in some detail in his 2007b. He also considers, but ultimately rejects, a couple of competing formalizations. See pp. 194–9 and Appendix 2.) For a content proposal similar to Williamson’s, see Häggkvist 1996.

<sup>25</sup> I am not assuming that the real content deductively entails that claim—alone or in conjunction with other available premisses—although the surface content does in fact entail it, and the other two candidates entail it given the additional premiss labelled (b) in n. 20:  $\diamond \exists x \exists p GC(x, p)$ . The support relation could be weaker, e.g. it could be an inductive

*justified true belief but does not know that someone in his office owns a Ford*—would do the job, but the surface content commits us to an existential claim that we know to be false, or at any rate know that we lack justification for believing. And the problem is not the falsity or lack of warrant per se; it is that the falsity and lack of warrant are too easily detectable. This brings out a second guiding constraint—roughly, that our intuitive judgement should not come out *obviously* false or *obviously* unjustified. A good candidate content should at least be one that we could reasonably take to be true, and that we could reasonably take ourselves to have justification for believing. Third, an adequate content proposal should conform to our semantic intuitions about deviance—it should not count clearly deviant realizations of the given problem case as non-deviant, or conversely. The suggestion that the content of our judgement is ‘just what it seems to be’ fails on this score too.

The proposal that Williamson attributes to his opponent is that the content is a certain strict conditional—NECESSITY. This proposal fails on similar grounds: it attributes a judgement to us that can too easily be seen to be false and unjustified, and it classifies some clearly deviant realizations of the case as non-deviant. His own view is that the content is the corresponding counterfactual—COUNTERFACTUAL. Just like NECESSITY, COUNTERFACTUAL alone does not provide sufficient grounds for rejecting the JTB theory, but together with the additional premiss that the case thus described *could* be realized, it does (other things equal<sup>26</sup>).

Thus COUNTERFACTUAL too satisfies the first constraint articulated above (what we might call the ‘epistemic’ constraint). But does it fare better than the competitors with respect to the second and third (the ‘psychological’) constraints? With respect to the second—yes; but, as we will see, not with respect to the third.

Unlike NECESSITY, COUNTERFACTUAL cannot easily be shown to be false—in fact, it may for all we know be true. Moreover, we may well have justification to believe it. But it is easy to envisage (or even bring about) situations in which COUNTERFACTUAL is false, and/or we lack justification to believe it, but that nevertheless do not seem to

---

or abductive implication relation. (But note that if this turns out to be the case then the thought experiment has considerably less force than we typically assume.)

<sup>26</sup> That is, provided all additional requirements on the acquisition of justification by inference are satisfied.



falsify our original intuitive judgement—situations in which we would retain that judgement, and rightly so.

Consider, for instance, how we would react to the discovery that the Gettier case is *actually* realized in the following way:<sup>27</sup> my uncle Smith stands to the proposition that someone in his office owns a Ford in the exact way stipulated in the case description, but uncle Smith has good reasons to believe that he is prone to hallucinate people driving Fords to work and prone to misremember what cars people drove in the past (and so on). Other things equal, uncle Smith does not have a justified true belief without knowledge. But then COUNTERFACTUAL is false. However, on learning about uncle Smith's predicament, we would not retract or modify our original intuitive judgement. And just as before, this seems to be the appropriate reaction, since the envisaged (actual) realization of the case—'the uncle Smith realization'—is clearly deviant. It requires that we read the case description in a way we know it was not meant to be read. The fact that the realization is 'nearer' (in fact, as near as can be) does not help.<sup>28</sup>

It is hard to see how Williamson could respond to this objection without losing an important part of his motivation for rejecting target rationalism. One option would be to deny that the uncle Smith realization is in fact deviant. But that does not seem viable. If the appeal to our semantic intuitions about deviance is problematic here, surely it is equally problematic when used against NECESSITY. That is not to say that it is *never* problematic. The point is just that there seems to be no relevant difference between the uncle Smith realization and the corresponding—more 'distant'—realizations of the case that only falsify NECESSITY (and not COUNTERFACTUAL). On the face of it, those realizations are on a par: they are equally deviant, and for the same reasons.

Nevertheless, this is what Williamson replied when I first presented him with this objection,<sup>29</sup> and it also seems to be his considered response. In his most recent writings on the topic, he briefly considers the possibility of an uncle-Smith-type scenario (i.e. a seemingly deviant but actual or nearby non-actual realization of the case), and he

<sup>27</sup> The same point can be made using a non-actual but nearby realization of this sort, or a nearby—actual or non-actual—realization in which the subject knows the proposition in some other way (some way not stipulated in the case description).

<sup>28</sup> It was recently brought to my attention (by an anonymous referee for this journal) that Ichikawa and Jarvis 2009 use similar cases to argue against Williamson. Their positive proposal is criticized in Sect. 2.5 below.

<sup>29</sup> In conversation.

recognizes that we may not take our original intuitive judgement to be falsified by it. But he then goes on to argue that we would be wrong to so react: that realization *would* falsify our original judgement, and our failure to acknowledge this is just a symptom of a common character flaw: a general reluctance to admit mistakes.<sup>30</sup>

However, this move strikes me as completely ad hoc — it is ad hoc to diagnose our (expected) reaction to the uncle-Smith-type scenario in this way, but not allow the corresponding move to an advocate of the target proposal. Is it not *just as* plausible (or implausible) that our reaction to the more distant realizations that Williamson exploits against that proposal is erroneous?

Note that I am not just arguing that, *for all we know*, Williamson's content proposal is mistaken since, for all we know, the actual or a nearby world contains a deviant realization of the Gettier case (and so, for all we know, COUNTERFACTUAL is false). The point is that, on Williamson's view, there *could be* no deviant but actual or nearby non-actual realizations. But that seems wrong: on the face of it, the uncle Smith realization is one such realization, and it is easy to come up with more. This, I submit, is already enough to refute the proposal. In his response, Williamson does not contest the possibility of uncle-Smith-type scenarios, only their deviance. But he does not explain away their *apparent* deviance. And it is very hard to see how to do that in a way that does not backfire. To reiterate: why think that we are poor judges of deviance when it comes to nearby realizations of the case, but reliable when it comes to more distant realizations? (If indeed there is an asymmetry here, one might expect it to be the other way round.)

The above consideration seems to suggest that COUNTERFACTUAL is too strong; that the range of possible realizations it permits is too wide. In light of this, one might replace COUNTERFACTUAL with its mere possibility — the claim that it is *possible* that, if someone were to stand to a proposition as in the case as described, then she would have a justified true belief without knowledge.<sup>31</sup> The embedded counterfactual would not be falsified by the uncle Smith realization; moreover, it is

<sup>30</sup> Thus he writes: 'Many philosophers have the common human characteristic of reluctance to admit to having been wrong. We should not distort our account of thought experiments in order to indulge that tendency' (Williamson 2007b, p. 201).

<sup>31</sup> In symbols:  $\diamond(\exists x\exists pGC(x,p) \square\rightarrow \forall x\forall p(GC(x,p) \rightarrow (JTB(x,p) \& \neg K(x,p))))$ . Williamson considers and rejects this candidate (2007b, Ch. 6) on the grounds that it only yields a good inference within the modal system S5, and that it is implausible to attribute a tacit commitment to S5 to the average thought experimenter; especially since there is another

something that we plausibly believe, that we have justification to believe, and that (arguably) would yield a good inference.

But this proposal seems like overkill, given that there is another, simpler possibility claim in the vicinity—one that it is also plausible that we believe, that we have justification to believe, that would not be falsified by the uncle Smith realization, and that yields a good inference, namely:

POSSIBILITY 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  $p$ .<sup>32</sup>

I suggest that this is the real content of our judgement. If that is right, it looks like the Gettier inference has a very simple structure: other things equal, we can rationally infer that the JTB theory is false directly from the Gettier judgement.<sup>33</sup>

I will now defend this suggestion against some challenges and rival views.

## 2.

### 2.1 *Implicit generality I*

First, it might be objected that POSSIBILITY is *too specific*:

As competent thought experimenters, we typically realize right off the bat that many of the details that are included in a given case description are *inessential*—that we can abstract away from many of the stated facts, and ‘import’ certain unstated facts, without altering the distribution of the test properties. For instance, it is obvious that Smith’s having a justified true belief without knowledge does not depend on Jones’s driving a rented rather than a stolen Ford, or on

---

candidate available—namely, COUNTERFACTUAL—that does not require us to attribute that commitment to her. (This objection does not apply to my proposal.)

<sup>32</sup> We may represent POSSIBILITY as follows:  $\diamond \exists x \exists p (GC(x, p) \ \& \ JTB(x, p) \ \& \ \neg K(x, p))$ . This transparently entails the intended conclusion:  $\neg \square \forall x \forall p (K(x, p) \leftrightarrow JTB(x, p))$ . I take it that not any content with the same truth-conditions as POSSIBILITY counts as the same content, but, at least for our purposes, any content with the same logical form does. (The specific grammatical form is certainly not essential.) For example, each of the following variants counts as the same content: that someone could be related to  $p$  as stated in the case description and fail to know that  $p$ , despite having a justified true belief that  $p$ ; that it is possible for someone to stand to  $p$  as in the case as described, justifiably truly believe that  $p$ , but lack knowledge that  $p$ ; and (perhaps less obviously) that it is possible that someone *who* stands to  $p$  as in the case as described has a justified true belief that  $p$  but does not know that  $p$ .

<sup>33</sup> A little more structure will be suggested in Sect. 2.3 below.

Smith being male rather than female. We may not be able to distinguish all the inessential details from the essential ones, at least not right away, but we *do* know right away that not all the details matter equally. This suggests that the content of the Gettier judgement is not the highly specific claim that someone could stand to a proposition as in the case *as described*—be placed *exactly* as in the given case description, all picturesque details included—and have a justified true belief without knowledge. Our ability to abstract away from some of those details shows that we are committed to something more general. And there is an obvious alternative available: the claim that *someone could have a justified true belief without knowledge*. Besides, that claim is really all we need to get a good inference off the ground.

Note that there are two distinct complaints here: first, that the (competent) subject, who makes the Gettier judgement, typically realizes right away that she is responding to an instance of a general *schema*—that the Gettier case, as described, could be altered or ‘filled out’ in a number of ways, while calling for the same intuitive verdict. She realizes that things do not have to be *exactly* the way they are stipulated to be, for someone to have a justified true belief without knowledge. But—the objection goes—my content proposal cannot account for this, since *POSSIBILITY* specifically concerns situations where everything is exactly as stipulated (in the original case description). We might put this point by saying that the Gettier judgement has a certain *implicit generality* that *POSSIBILITY* fails to capture.

The second complaint is that we do not *need* a premiss as specific as *POSSIBILITY* to secure a good inference—a good inference to the conclusion that the JTB theory is false. The first conjunct in the embedded clause is completely redundant. All we need is the claim that,

*POSSIBILITY\** It is possible that someone has a justified true belief that *p* but does not know that *p*.<sup>34</sup>

Like *POSSIBILITY*, *POSSIBILITY\** transparently contradicts the modal claim entailed by the JTB theory—thus it too secures a good inference; but, unlike *POSSIBILITY*, *POSSIBILITY\** does not introduce any redundant elements, and it captures the implicit generality of the Gettier judgement. (Or so it is argued.)

<sup>34</sup> In symbols:  $\diamond \exists x \exists p (\text{JTB}(x, p) \ \& \ \neg \text{K}(x, p))$ . I have encountered this proposal in conversation many times.

However, not only is POSSIBILITY\* an inadequate content proposal, but both of the complaints that supposedly motivate it are confused. My reply to the first complaint is, in outline, that the implicit generality of the Gettier judgement can be explained in a way that is consistent with my content proposal. (For reasons that will emerge, the details of this reply are deferred until Sect. 2.2.) My reply to the second complaint is that, granted, POSSIBILITY\* is all we need to secure a good inference—but a good inference is not good enough. The present aim, recall, is not just to articulate a possible (rational) route to the rejection of the JTB theory (i.e. a so-called ‘rational reconstruction’ of the Gettier inference); the aim is to articulate our *actual* route—more precisely, a rational route from our actual intuitive judgement, one that is plausibly available to us. This, of course, is why the content problem matters in the first place: the real content of the Gettier judgement is a crucial premiss in the argument that we are trying to spell out.<sup>35</sup>

Now, as we have seen, there are multiple pressures on candidate solutions to this problem: since the Gettier inference ostensibly embodies a good piece of reasoning, an adequate candidate should meet the epistemic constraint (in effect: it should secure a good inference) but, since it purports to paraphrase the content of our actual judgement, it should also satisfy the other two—the psychological—constraints that we identified in section 1.6. And there may well be further requirements. The crucial question, then, is whether POSSIBILITY\* fares as well as POSSIBILITY across the board. (If it does—but only then—perhaps the relative simplicity of POSSIBILITY\* tips the balance in its favour.)

At first sight it may look like it does: POSSIBILITY\* does not fail at the point where NECESSITY or COUNTERFACTUAL fails—it is not obviously false or unjustified, and it does not seem to count any clearly deviant realizations of the Gettier case as non-deviant, or conversely. However, the list of constraints that we compiled above was not exhaustive. For one thing, we must add that an adequate content proposal should generalize in natural ways to intuitive judgements other than the Gettier judgement, and that it should not ride roughshod over our pre-theoretical classifications of those judgements—for example, by failing to distinguish between intuitive judgements that we firmly take

<sup>35</sup> This, in turn, looks like the right way to proceed if the overall aim is to capture our ‘doxastic’ justification—our justification *for believing* that the JTB theory is false. (And that is certainly the appropriate aim if we take ourselves to *know* that the JTB theory is false, on the basis of Gettier’s thought experiment, since doxastic justification is required for knowledge.)

to be distinct. But POSSIBILITY\* does precisely that (on what looks like the only natural way to generalize the proposal). Simply put: it is hard to see how we could accept POSSIBILITY\* — as an analysis of the Gettier judgement — without committing to giving exactly the same analysis of, for example, the intuitive judgement that might be expressed by saying: ‘Jill has a justified true belief but does not know that the president has been assassinated’, and of the judgement that might be expressed by saying ‘Henry has a justified true belief but does not know that there is a barn in front of him.’<sup>36</sup> But it is absurd to suppose that the Gettier judgement is the *same judgement* as — that it has the same content as — either of those judgements.

Rather, I submit, the Gettier judgement *does* concern the highly specific situation stated in the given case description — it expresses the claim that someone could be situated *exactly like that* and have a justified true belief without knowledge. The corresponding judgement about a structurally similar problem case expresses a distinct possibility claim (of the same abstract form) — for instance, that someone named ‘Jill’ could read in the paper that the president has been assassinated, be unaware of all the misleading counter-evidence (etc.), and have a justified true belief but not know.

## 2.2 Implicit generality II

Second, it might be objected that POSSIBILITY — indeed *any* possibility claim — is *too weak*:

In effect, what POSSIBILITY says is that there is at least one possible way of satisfying the given case description such that it is a counter-instance to the JTB theory. But surely we are committed to something stronger than that? As has already been noted by the advocate of POSSIBILITY\*, we are typically aware that the case description can be satisfied in a multitude of ways — many of which constitute counter-instances. (Williamson has only shown that not *all* of them do.) We are also typically aware that the description can be altered in a number of ways and still yield counter-instances. And, importantly, all it takes to realize this is reflection on *one* specific problem case, under *one* description. Furthermore, a *failure* to realize this — or at least, a failure to ‘catch on’ — is a *rationality* failure: it betrays some kind of inconsistency or misunderstanding on our part. (This point is overlooked by POSSIBILITY’s advocate.) To clarify: by making the Gettier judgement, it seems that I incur a certain *rational*

<sup>36</sup> See Harman 1968; Ginet 1975; Goldman 1976.

*commitment*—a commitment to making the same (or the corresponding) judgement in response to a wide range of possible variations of the original case description. If I make a contrasting judgement—if, say, I judge that the protagonist *knows*—when presented with a filled-out variation according to which s/he has two sons, I am being confused or inconsistent: there is a problematic discord between my earlier and later judgement. Likewise if I make a contrasting judgement in response to an otherwise identical description according to which, say, Jones now drives a stolen rather than a rented Ford. Now, the natural way to account for this—the natural diagnosis of my apparent rationality failure—is to say that I am here *contradicting* my original intuitive judgement (or that I am contradicting something that obviously follows from it). But then the real content of that judgement cannot be a mere possibility claim, such as POSSIBILITY (or POSSIBILITY\*, for that matter): the possibility claim is perfectly compatible with it being the case that a realization in which the protagonist has two sons is one where s/he knows. Presumably, the content of the judgement is after all some kind of strong modal conditional—most plausibly, it is a suitably restricted necessity claim.

There are different ways to elaborate on this suggestion, but the basic thought is just to weaken NECESSITY by restricting the scope of the necessity operator to some specific subset of possible worlds that satisfy the given case description. (Or better: that satisfy the given description *or some slight variant of it*—a qualification along these lines is needed to get the right generality. This introduces some additional complications, but we can ignore those here.<sup>37</sup>) Let us use ‘the *intended* Gettier case’ as a placeholder name for that subset of worlds. We can then express the suggestion by saying that the content of the Gettier judgement is the claim that,

NECESSITY\* Necessarily, anyone who stands to a proposition  $p$  as in the intended Gettier case has a justified true belief that  $p$  but does not know that  $p$ .<sup>38</sup>

<sup>37</sup> For one thing, note that there is a tension between achieving the right generality, and avoiding the problem facing POSSIBILITY\*: of assimilating judgements that we pre-theoretically take to be distinct. To avoid that problem, the proposal had better not incorporate the qualification I suggest above—the relevant subset of worlds must only include worlds that satisfy the given description. (In what follows I will understand the proposal in this way.)

<sup>38</sup> We can introduce a symbol for the intended Gettier case—‘IC’—and represent NECESSITY\* thus:  $\Box \forall x \forall p (IC(x, p) \leftrightarrow (JT(x, p) \ \& \ \neg K(x, p)))$ . But it is best to think

Add the premiss that the intended case *could* be realized, and once again we have a good inference. Thus NECESSITY\* too meets the epistemic constraint. It also seems to avoid the problems that fault NECESSITY and COUNTERFACTUAL: since the modal operator in NECESSITY\* only ranges over the intended (as opposed to all possible, or all nearby) realizations of the case, NECESSITY\* is not obviously false or unjustified, and it respects our semantic intuitions about deviance. Last, it looks like NECESSITY\* — unlike POSSIBILITY\* — has the resources to individuate our intuitive judgements in a way that lines up with our pre-theoretical classifications of them.

On the face of it, then, NECESSITY\* is a good bet — perhaps our best bet, if we wish to maintain that the content of our judgement is a necessity claim, in light of Williamson's objection.<sup>39</sup>

But it is doubtful that the proposal can be made to work. First, NECESSITY\* needs to be spelled out in more detail — in particular, the placeholder ('the intended case') must be eliminated — and this turns out to be extremely difficult. Second, just like with POSSIBILITY\*, the complaint that is supposed to motivate the proposal does no such thing: the phenomenon that NECESSITY\* is invoked to explain can be explained in a way that is compatible with POSSIBILITY being the real content of our judgement — and there is independent reason to seek an alternative explanation, since that phenomenon also arises in situations where the kind of explanation that NECESSITY\* provides is not even remotely plausible (see Sect. 2.3). Let us take these points in turn.

As stated, NECESSITY\* is way too schematic; we cannot properly evaluate it — in particular, we cannot adjudicate between NECESSITY\* and POSSIBILITY — in the absence of more detail. We need to be provided with a more informative characterization of the set of worlds that makes up the intended Gettier case (i.e. the set that the modal operator in NECESSITY\* ranges over) — a characterization that captures

---

of NECESSITY\* as a *template* for a content proposal, and we can expect substantial variation (in the content and truth conditions) of specific proposals that fit the template. See more below.

<sup>39</sup> It may even be what the actual rationalists to whom Williamson is reacting had in mind all along — the actual rationalists who maintain that the content of our judgement (and/or 'intuition') is a necessity claim. The exegetical question is complicated by the fact that those rationalists rarely express their view on the matter very rigorously. (At least that goes for those rationalists who were writing prior to Williamson's work on the topic.) Here is Bealer: 'when we have a rational intuition — say that if *p* then not not *p* — it presents itself as necessary: it does not seem to us that things could be otherwise; it must be that if *p* then not not *p*.' (Bealer 1998, p. 207 and p. 3; cf. his 2000.) More recently, D. Sosa has defended a content proposal that clearly fits the NECESSITY\* template (D. Sosa 2006). See also Grundmann and Horvath MS.



all and only non-deviant realizations of the Gettier case, but that does not itself contain terms like ‘deviance’ and ‘intended’. Moreover, the characterization had better not make NECESSITY\* come out trivial, nor too rich — that is, as presupposing information that the thought experimenter could not yet possess. But it turns out to be very difficult to provide a characterization that meets these constraints. Below I offer my own best attempt.

A natural starting point is to look at what specific, deviant realizations of the Gettier case have in common — presumably there is some shared non-trivial feature, or small set of such features, in virtue of which they are all deviant. (If Williamson were right, they would all be relatively ‘distant’ from the actual world, but as we have seen that is not so.) For brevity, let us refer to a realization of the Gettier case as a ‘G-world’. The obvious thing to notice about the deviant G-worlds encountered above is that they are worlds where the featured subject knows the target proposition, and/or her (prima facie) justification to believe it is defeated. This may inspire the suggestion that deviance is a simple function of how things are with the relevant test properties: that the non-deviant G-worlds comprise (all and only) those worlds where someone stands to  $p$  as stipulated in the given case description, has no defeaters for her justification to believe that  $p$ , and does not know that  $p$ .

But although this may well be *true*, it is useless for present purposes. We do not have to run a thought experiment to realize that any such world is a world where the subject has a justified true belief that  $p$  and does not know that  $p$  (but that is what NECESSITY\* comes to, on the suggested gloss). Recall that it is *stipulated* in the given case description that the subject truly believes that  $p$ . Conversely, if this were the content of the Gettier judgement, then those who reject that judgement would be seriously confused indeed.<sup>40</sup> All the controversy ought to be over whether someone could stand to  $p$  as in the case as described, lack defeaters, and *still fail to know*.<sup>41</sup> Indeed, on this view, the Gettier judgement turns out to be epistemically idle — it

<sup>40</sup> For instance, Pailthorp 1969 and Weatherson 2003. I owe this point to Patrick Greenough.

<sup>41</sup> More precisely: over the claim that someone could stand to  $p$  as described, lack defeaters for her justification to believe that  $p$ , and not know that  $p$ . A separate premiss to this effect must replace (b) above (cf. n. 20), lest the overall inference be invalid. (Of course, on my proposal all the controversy *is* over this claim, but since that claim is also the content of the Gettier judgement, the dialectic is not being misrepresented.)

plays no role at all in explaining how we are justified in believing that the JTB theory is false.

A more careful look at the specific deviant G-worlds encountered so far reveals that they are not just worlds in which the subject knows or lacks defeaters: they are worlds in which she knows or lacks defeaters *in virtue of some 'extraneous' fact*—some fact not explicitly stipulated in the case description. Perhaps, then, the range of the modal operator (in NECESSITY\*) is restricted to G-worlds that contain no such extraneous facts? But now this, in turn, must be precisified. Exactly *what* class of extraneous facts makes for deviance? It is tempting to take the phrase 'in virtue of' to express a strong modal dependency relation—for example, a logical or metaphysical entailment relation. Thus, one way to precisify the above would be to say that the modal operator ranges over all and only G-worlds that contain no set *S* of extraneous facts, such that *S* is (logically/metaphysically) sufficient for the subject to know that *p*, or *S* is sufficient to defeat her justification to believe that *p*. But this cannot be right. Consider the first clause: that clause rules out *some* deviant G-worlds where the subject knows—namely those in which she has a back-up route to knowledge that is modally independent of the stipulated facts. But there are other, and on the face of it equally deviant, G-worlds where she knows because the stipulated facts have been 'enhanced' in certain ways—where she knows that *p* but would not have done so unless she stood to *p* as stated in the given case description.<sup>42</sup>

In light of this, one might modify the last suggestion by saying that the modal operator in NECESSITY\* ranges over (all and only) those G-worlds that contain no set *S* of extraneous facts, such that *S* and the stipulated facts are *jointly* sufficient for the subject to know that *p*, or *S* and the stipulated facts are jointly sufficient to defeat her justification to believe that *p*. This rules out all deviant G-worlds that we have encountered so far (and any other that I can think of). But it is problematic for the same reason as the first suggestion mentioned above—it trivializes the Gettier judgement.

<sup>42</sup> For example: suppose that Brown knows that Smith believes that someone in the office owns a Ford on the basis of the false belief that Jones does. Suppose further that Brown also knows that Jones is about to reveal to Smith that he does not own a Ford, but that (for whatever reason) it is important to Brown that Smith keep believing that someone in the office owns a Ford. To ensure this, Brown tells Smith that someone in the office owns a Ford, and Smith thereby comes to know this by testimony. But he would not have come to know it unless he also stood to the proposition as stipulated in the original case description.

Here is why: the stipulated facts alone are not sufficient to determine whether the featured subject knows, or whether she is justified—*some* extraneous facts are needed to settle the matter. (This, of course, is what Williamson exploits against the target rationalist.) The question is *which* such facts we may legitimately ‘import’, when we evaluate whether some particular possible world that satisfies the given case description is a counter-instance to the JTB theory. The current suggestion is (in effect) that we may import anything *except* whatever facts, together with the stipulated facts, suffice for the subject to know or for her justification to be defeated. But, once again, that makes the whole exercise idle—we do not need to run a thought experiment to find out that any such world is one in which the subject has a justified true belief but does not know.<sup>43</sup>

To avoid trivializing the judgement, one might attempt to specify—in more informative terms—the relevant class of facts (i.e. those extraneous facts that, together with the stipulated facts, suffice for the subject to know, or for her justification to be defeated). However, even were that attempt to succeed, it is unclear how it would help. The trouble is that it is implausible that the *thought experimenter* has access to the needed specification—that she is able to have thoughts (and beliefs) about that class of facts, under a more revealing mode of presentation. It is certainly implausible that she has explicit propositional *knowledge* of what it takes for the subject in the case to know the target proposition, or of what it takes for her justification to be defeated. In so far as an informative specification of the relevant facts is available, accessing it requires a lot of theoretical work—witness the so-called ‘post-Gettier industry’ (cf. Shope 1983)—work that the thought experimenter cannot be presumed to have carried out prior to making the Gettier judgement.

The present problem is reminiscent of a familiar problem pertaining to *ceteris paribus* laws: it is notoriously difficult to eliminate a *ceteris paribus* clause—to come up with a non-trivial specification of the intended ‘exceptions’ (i.e. the conditions under which the *ceteris paribus* law does not apply). Possible candidates either render the law trivial, or are too rich to plausibly be known by the subjects who subscribe to it—at least, too rich to be known in advance of further

<sup>43</sup> The same problem arises if we gloss ‘in virtue of’ as some kind of epistemic or conceptual entailment: here too, the suggested restriction trivializes the judgement.

(here: empirical) inquiry.<sup>44</sup> It is not obvious that this problem cannot be solved or, if it cannot, that that shows *ceteris paribus* laws to be meaningless or vacuous. But it is certainly a problem worth taking seriously; likewise for the analogous problem that pertains to the NECESSITY\* proposal. At the very least, then, the burden of proof lies on the advocate of that proposal — she must show that the problem *can* be solved, or explain away its apparent significance.

That burden is especially pressing, given that NECESSITY\* does not seem to be particularly well motivated to begin with — the phenomenon that it is invoked to explain (as I have set things up) calls for a different explanation, one that is compatible with the POSSIBILITY proposal.<sup>45</sup>

### 2.3 *The generality of reason-based judgements*

The phenomenon that needs to be explained is the ‘implicit generality’ of intuitive judgements. We have seen this to be a rather complex phenomenon. The competent thought experimenter typically realizes that there is a wide range of innocuous variations on the given case description: variations that merit the same (or the corresponding) intuitive judgement. And she can be expected to behave accordingly: she can be expected to respond in the same way — with the same intuitive judgement — to any specific variation in that range. But furthermore, there is a sense in which she *should* respond in the same way to any such variation. Her response to the original description (of, say, the Gettier case) seems to *commit* her to upholding a certain wider pattern in her intuitive judgements — it commits her to responding in the same way (whatever way that was) to a filled-out variation that stipulates that Smith has two sons, that he dislikes cabbage, or that he wears drag on Wednesdays. Likewise for a variation that features Brown in lieu of Smith — and so on. A failure to respond in the same way to any such variant constitutes a rationality failure.

The NECESSITY\* proposal diagnoses this rationality failure as the familiar type of rationality failure that we are guilty of when holding

<sup>44</sup> For discussion, see Earman, Roberts, and Smith 2002; Fodor 1991; Schiffer 1991; Strevens MS; Woodward 2002.

<sup>45</sup> The motivation that I offer here, on behalf of NECESSITY\*, has to my knowledge not been put forward in print. But it is hard to see what *else* might motivate the suggestion that the content of our judgement is a necessity claim (of this or any other sort). And the motivation that I offer was recently explicitly endorsed, by Thomas Grundmann and Joachim Horvath, in a paper presented at a conference on Thought Experiments and the A Priori, at the University of Fortaleza in August 2009.

contradictory beliefs—when believing (e.g. judging) two or more contents that are logically inconsistent with one another. On this view, then, it is no mystery how a single intuitive judgement can incur a rational commitment to uphold a certain wider pattern in one's judgements: it does so just by having the particular content that it has (given our standing obligation to not hold contradictory beliefs). It is clear that this kind of explanation is not available on my account: *POSSIBILITY* does not contradict, say, the claim that someone could stand to *p* as in our sample case description, dislike cabbage, have a justified true belief that *p*, and know that *p*. But that is a drawback only if the suggested explanation is indeed the natural (read: best) explanation of the commitment—and it is not.

To see this, suppose that I encounter a known-to-be *actual* (non-deviant) Gettier-style problem case, and that I make an intuitive judgement about it, a judgement that I might express by saying 'Smith has a justified true belief, but does not know, that someone in his office owns a Ford'. There is no puzzle about the content of *this* judgement—the puzzle only arises when the problem case is non-actual. When the case is actual, and known to be so, we can take the judgement's surface content at face value. However, it looks like this judgement too—my intuitive judgement about the actual case—incur a rational commitment (arguably the very *same* commitment that was described above). If I retract my original judgement on finding out, say, that Smith has two sons, or that he dislikes cabbage, then again I betray some kind of inconsistency or confusion. Similarly if I make a contrasting judgement in response to an otherwise identical case that features Brown rather than Smith (etc.). But there should be no temptation to construe the content of my intuitive judgement about an *actual* problem case as a suitably restricted necessity claim along the lines of *NECESSITY\** (or as a modal or conditional claim of any other sort). We need a different explanation here—a different explanation of the apparent rational commitment—and if we can provide a different explanation here, why not apply that explanation across the board?

I suggest that, in both cases, the commitment reveals something about the *grounds*, not the content, of the intuitive judgement. There is a certain generality to my grounds (or *reasons*) for judging that Smith has a justified true belief but does not know—whether, as it were, Smith is actual or hypothetical—and this generality of grounds rationally constrains my options when it comes to making

certain other judgements. (This will shortly be spelled out in more detail.)

There are many situations where an explanation of this sort is appropriate — even forced. Here is a straightforward example: suppose that a student in my seminar, Anna, puts her feet up the table and I judge that she should put them down. Other things equal, I can be expected to react in the same way to the next student who puts her feet up. (As before: not only is this a plausible *prediction* of my behaviour — it seems that I am somehow *bound*, by my original judgement, to react in the same way.) If I do not, then I betray some kind of inconsistency or confusion. But here too, there is no temptation to paraphrase the content of my original judgement as any kind of modal or conditional claim. The original judgement is just what it seems to be: it is a judgement about *Anna*. Correspondingly, there is no temptation to say that my subsequent judgement — my judgement that, say, Balder may keep his feet up — *contradicts* the first judgement.

Rather, the natural explanation goes something like this: normally, the reasons for which an (overall reasonable) person would judge that a given student should put her feet down have a certain generality — perhaps they apply to everyone in the room, or at least to everyone in the room whose feet are dirty. What my divergent judgements betray is that I did not in fact base my original judgement on reasons of the presumed generality (even though, perhaps, I *should* have done so) — or I did, but I failed to see that those reasons applied to the next student too. A third possibility is that my reasons were defeated in some non-obvious way in the latter situation.<sup>46</sup> Any which way, the tension between my two judgements reflects something about the reasons on which they are based — in no way does it suggest that my original judgement is not just about Anna (and my subsequent judgement not just about Balder).

To be a little more precise: at least *some* of the reasons on which each judgement is based are presumably specific to the case at hand. For one thing, I would not — and I should not — have judged that Anna should take her feet down unless I took her to have put her feet up in the first place. On a very simple syllogistic model of my

<sup>46</sup> In the first and third case, the apparent rationality failure may not be genuine — at any rate it is not the kind of rationality failure that we are guilty of when failing to apply our reasons consistently. (But there are at least two other ways in which I may have done something wrong: I may have based my original judgement(s) on *bad* reasons — perhaps I relied on a principle that unfairly exempts my favourite student — or I may have based it on no reasons at all.)

reasoning, it includes the minor premiss that Anna put her feet on the table. But (if the case is normal) it also includes a major premiss roughly to the effect that anyone who puts their feet up in my seminar—and, perhaps, meets some further specification—should take them down. Crude as this model may be, it arguably has the right basic structure: a judgement like this—at any rate a *rational* judgement like this—results from the application of a general principle or rule to a particular instance, an instance that falls under it (and/or is taken by the subject to fall under it).<sup>47</sup> And ‘if I do it once I should do it twice’—I should apply that principle or rule in any other circumstance that is *relevantly similar* to that in which I first applied it—unless, of course, the new circumstance provides me with strong grounds for rejecting the principle or rule itself. This point is sometimes expressed by saying that *reasons must be consistently applied*; and, as the present example serves to illustrate, that requirement is not equivalent to our requirement to not believe logically inconsistent contents.

Let me safeguard against another potential confusion: the judgement that Anna should do such-and-such is an overtly *normative* judgement, and this may invite the suggestion that the rational commitment that is exemplified here is just a commitment to respect a certain *supervenience* relation—namely, the supervenience of normative facts (or properties) on non-normative facts (or properties).<sup>48</sup> But the normativity of the judgement is a red herring. Just suppose, instead, that Anna falls asleep in class, and I judge that she is bored; that

<sup>47</sup> I want to leave open whether this is the right model for *all* judgements (or judgement types), i.e. whether all rational belief-formation involves the application of general principles/rules. I also want to leave open whether—in the present example, and elsewhere—these principles/rules are best understood as *prima facie* reasons, or as *ultima facie* reasons. (This affects how we should think of their contents.) More generally, I want to leave open how the ‘crude model’ is to be refined. Some reject the model wholesale—namely, proponents of so-called ‘particularism’; e.g. McDowell 1979; Dancy 2001, 2004. (Their focus tends to be on moral reasons, but many of their concerns generalize to practical reasons more generally, as well as to reasons for belief.) Note, however, that even the particularist owes us an account of our obligation to apply reasons consistently—one that does not simply reduce it to our obligation to avoid contradictory beliefs. To my mind, extant attempts at doing so (within the particularist framework) have not been successful, but I cannot argue this in detail here. (See e.g. Dancy 2001, Ch. 5, Sect. 4.) And the most important point is that, since the particularist too must (and typically does) acknowledge that there is a distinctive requirement here—a consistency constraint, of *some* kind, on the application of reasons—the broad strategy that I adopt in this section may in principle be acceptable even to her. (For more on particularism, and relevant criticism, see e.g. Kaebnick 1999; Jackson, Pettit, and Smith 2000; Väyrynen 2006. For other worries about the crude model, see Boghossian 2008a.)

<sup>48</sup> Thanks to the editor of this journal for pressing this point.

Roger runs a five-minute mile, and Anna judges that he is fast; or that Sherlock Holmes finds blood on the butler's knife, and he judges that the butler did it. None of these judgements—that Anna is bored, that Roger is fast, and that the butler did it—are normative, but they all incur a commitment of the type that is at issue: a commitment to making the same (or the corresponding) judgement in any relevantly similar circumstances.

Next, that commitment goes well beyond a commitment to respect supervenience—even in those cases where the judgement in question *is* normative. To say that the normative supervenes on the non-normative is, to a first approximation, to say that there could be no normative difference without a non-normative difference. All I have to do, then, to respect supervenience, is make the same normative judgement in any two circumstances that are alike in all non-normative respects. And *that* constraint need not be violated for there to be a problematic discord between my two judgements in the example where I judge that Anna should put her feet down but that Balder may keep his up. The two sets of circumstances that trigger the judgements may differ in any number of non-normative respects (e.g. Balder is male and Anna female, they occupy different seats, and so on). But there is a problematic discord between my two judgements as long as the triggering circumstances do not differ in any *relevant* respects—normative or non-normative, as the case may be. In other words, since not any difference is a relevant difference, a commitment to respect supervenience cannot help explain what goes wrong in a case like this.<sup>49</sup> (Granted: a failure to respect supervenience is also a rationality failure, but it is a *different* rationality failure.)

To sum up: my suggestion is that the rational commitment that an intuitive judgement seems to 'bring on' is fundamentally the same kind of commitment that is manifest in the above examples—a commitment that can be found in any reason-based activity (in the practical as well as in the epistemic domain). The nature of this commitment is by no means fully, or even particularly well, understood as yet. But it is a commitment that we are all familiar with. It is implicitly invoked at any time someone is called on to defend why she made a certain judgement (or performed a certain action) in a circumstance  $C_1$ —given that  $C_1$  seems to match another circumstance

<sup>49</sup> See Dancy 2004, pp. 85–9, for related discussion.



$C_2$  in all relevant respects, and that, in  $C_2$ , she made a judgement (or performed an action) of a contrasting type. Of course, sometimes the right response to this challenge is that  $C_1$  and  $C_2$  are not relevantly similar after all—but the point is that, without a consistency requirement on the application of reasons, the challenge would never be well motivated in the first place.

The suggested explanation is compatible with the POSSIBILITY proposal. And recall that it has the distinct advantage—over the explanation that NECESSITY\* provides—of applying in the situation where the problem case is actual, as well as in the situation where the case is hypothetical. It is highly implausible that our intuitive judgement about the actual case is a disguised modal judgement; a better explanation of the commitment it incurs is that the judgement is based on reasons of a certain generality. But if that is the right thing to say here, when the problem case is actual, surely it is also the right thing to say when the case is hypothetical: my intuitive judgement about the hypothetical case is also based on reasons, reasons of a certain generality—presumably, at least in part, the *very same* reasons on which my judgement about the actual case is based.

However, it is controversial to suggest that intuitive judgements are based on reasons—*any* reasons. One might even worry that this makes the explanation I have outlined a non-starter, since those judgements are supposed to play a *foundational* role in philosophical methodology.<sup>50</sup> Now, it is not clear to me that that supposition is correct—that said, it is worth noting that the claim that intuitive judgements are reason-based is, at least in principle, compatible with the claim that they play a foundational role. The key feature that makes a judgement or belief fit to play that role is that it is *non-inferentially justified* (or non-inferential, for short)—roughly: that its justification does not rest on the justification of any of the subject's other beliefs. And it is possible for a judgement to be *both* non-inferential *and* based on reasons, as long as not all reasons are themselves beliefs (or belief contents)—that is, as long as there are at least some 'non-doxastic' reasons.<sup>51</sup> This principled compatibility would still only be of marginal interest, in the current context, if the kind of explanation of the rational commitment that I outlined

<sup>50</sup> Thanks to Paul Boghossian, and to an anonymous referee for this journal, for pressing this point.

<sup>51</sup> For discussion, see e.g. Bonjour 1985; Pollock and Cruz 1999; Pryor 2005.

above required a doxastic construal of reasons. But, at least on the face of it, it does not require that.<sup>52</sup>

Second, it is in fact *independently* plausible—although not widely recognized—that intuitive judgements are based on reasons. The guiding thought here is just this: in making the Gettier judgement, we are (loosely put) attributing justified true belief without knowledge to a subject who stands to a proposition in a certain peculiar way—namely, the way that is specified in the given case description. Indeed, on reflection it seems clear that we are attributing those properties to her in part *because we take her* to stand to a proposition in that peculiar way. Unless we took her to be thus situated, we would not—and we should not—judge that she has a justified true belief but does not know. (Recall and compare the example above: my judgement about Anna.) Of course, since the problem case is hypothetical, there is no actually existing subject to whom we are attributing anything—all of the above must be paraphrased. But the present point can be appreciated independently of any particular solution to the content problem. (And the point applies equally to the situation where the case is actual, and no paraphrases are called for.)<sup>53</sup> In

<sup>52</sup> Consider the following case: I enter a brightly lit room, eyes wide open, and I judge that it looks bright in here. This judgement is a prima facie paradigm of a non-inferential judgement, but it too incurs a rational commitment (to making the same, or the corresponding, judgement in any relevantly similar circumstances) and once again, an explanation in terms of reasons, not content, seems to be in place. How, in outline, might the explanation go? Perhaps the best way to conceive of the general reason that explains the commitment here is as a rule that takes me from a certain type of visual experience to a certain belief about how things look. (Nothing here hinges on the exact formulation of the rule.) Of course, we also need a story of what makes that rule my ('operative') reason, and the story had better not require me to have a *justified belief* in the rule, which serves as a step in some suitable reasoning process. But a different story is arguably needed *anyway*—and not just for the case of perceptual belief. (For related discussion, see Boghossian 2008a; Boghossian and Wright MS; Pollock and Cruz 1999, Ch. 5; Pryor 2005, Sect. 7.)

<sup>53</sup> 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:  $\diamond \exists x \exists p \text{GC}(x, p)$ .) 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

brief: an intuitive judgement is always made in response to a specific case description, and it seems very plausible that the thought experimenter's exposure to that description has more than *causal* significance—specifically, that (some or all of) the information that is explicitly stipulated in that description constitutes a *reason* for her to make a certain intuitive judgement. (If what I argued above is correct, then the stipulated information at best constitutes *some* or *part* of her overall reason(s), but that is all one would expect.)

There is obviously much more to be said on this issue, but that is a task for another occasion.<sup>54</sup> For now, let it suffice that it is by no means *obvious* that intuitive judgements are not based on reasons. In fact, it strikes me as a plausible working hypothesis that they are.

#### 2.4 *Just another fiction?*

The next rival content proposal to be considered here assimilates intuitive judgements about cases to certain judgements about standard *fiction*. On this view, there is no *special* content problem about intuitive judgements: the Gettier case is just a piece of standard fiction, and the Gettier judgement is just a judgement about a fictional character—specifically, it is an ‘internal’ fictional judgement, such as the one I might express by saying ‘Hamlet is a border liner’, or ‘Holmes lived on Baker Street’. (In general, an internal fictional judgement is one that goes beyond what is explicitly stated in the text, but that does not yet treat the fiction *as* a fiction.<sup>55</sup>) There is still a lot of disagreement over the semantics of fictional discourse, but whatever the correct account turns out to be, it can be expected to apply to intuitive judgements—or so the suggestion goes.<sup>56</sup>

This would not constitute a genuine alternative to my proposal, if (internal) fictional judgements were plausibly analysed along the lines

---

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.)

<sup>54</sup> For further discussion, see my doctoral dissertation (Malmgren MS), Ch. 3.

<sup>55</sup> As in, say, ‘Hamlet is modelled on the legendary folk hero Amleth’. (Cf. Currie 1990; Eagle 2007; Thomasson 2003.) For brevity, I sometimes omit the qualification ‘internal’ below. I take it that the assimilation of intuitive judgements to any other type of fictional judgements is a non-starter.

<sup>56</sup> Lewis seems to endorse a view of this sort—see his 1983a, p. 278: ‘note that the philosophical example is just a concise bit of fiction’. See also Ichikawa and Jarvis 2009. (Their view is discussed separately in Sect. 2.5.)

of POSSIBILITY. But I take it that they are not, and not *just* because we can, it seems, make true judgements even about explicitly *impossible* fictions.<sup>57</sup> Set that problematic aside—perhaps as part of a larger ‘divide-and-conquer’ approach to fictional discourse—and POSSIBILITY is still not a plausible model. That is, it is not plausible even if applied exclusively to judgements about so-called ‘realistic’ fiction, such as *The Tragedy of Hamlet: Prince of Denmark*. (Some of the reasons why will become apparent below.)

By way of initial gloss, we might express the new proposal by saying that the content of our judgement—the Gettier judgement—is the claim that,

FICTION In the Gettier fiction, Smith has a justified true belief but does not know that someone in his office owns a Ford.

FICTION is intended to be compatible with a variety of different accounts of fictional judgements (corresponding to different interpretations of the ‘in-the-fiction operator’);<sup>58</sup> the basic idea is just that the Gettier judgement is a fictional judgement—*however* such judgements are in general to be analysed. Certain questions cannot be usefully addressed at this level of abstraction (e.g. whether and how FICTION secures a good inference). But others can. We do not need more detail to see that the assimilation of the Gettier judgement to a fictional judgement is a mistake: the intuitive truth-conditions of the respective judgements differ enough to suggest that they do not have the same semantics.

For one thing, the judgement that I express by saying ‘Hamlet is a border liner’ would not be verified (or falsified) by the existence of some actual person who fit all the explicit descriptions provided by Shakespeare, and who did (or did not) have a borderline personality. In contrast, an intuitive judgement—for example, the Gettier judgement—*does* seem to be verifiable/falsifiable by actual (non-deviant) realizations of the relevant problem case. We might put the point by

<sup>57</sup> Relatedly: some fictions are *implicitly* impossible—impossible (at least in part) in virtue of some of the extraneous facts that are ‘true in the fiction’. But that is precluded by the POSSIBILITY-model. (Cf. n. 71.)

<sup>58</sup> FICTION is not compatible with just *any* account of fictional judgements: it presupposes that some form of ‘operator fictionalism’ is correct—and, more generally, that the basic problem with fictional judgements is a *content* problem, rather than an *attitude* problem (cf. n. 7). But my objections against the assimilation of intuitive judgements to fictional judgements, below, do not substantially depend on this framework (although they have to be reformulated if it is incorrect).

saying that fictional characters are *essentially* fictional,<sup>59</sup> whereas characters in philosophical problem cases are not. Just consider, once more, the uncle Smith realization: that realization is deviant, but surely, what *makes* it deviant does not have anything to do with Smith's being actual. (If uncle Smith did not have the stipulated defeaters, would the realization *still* be deviant?)

For another, what counts as a permissible interpretation of a case description seems to depend, at least in part, on the specific *use* to which the case is put; more precisely, the tacit constraints that govern our interpretation of a case description are sensitive to the *target* of the given thought experiment—to what theory is being tested. Consequently, they may vary with a change in target. For instance, we may not suppose that the subject in the Gettier case has more than one route to knowledge available, but we may suppose that killing him would cause a riot in which fifty other people die. The former extraneous fact is *ruled out*, but the latter is *left open*: it is indeterminate whether it obtains in the case.<sup>60</sup> (At any rate this is so when the case is used for its original purpose—to test the JTB theory.) The converse holds for the potential victim in a 'Trolley case': we may not suppose that killing him would cause a riot in which fifty people die, but we may indeed suppose that he has more than one route to knowledge available (here: knowledge of any proposition of our choice).<sup>61</sup> Interestingly, this shift is not just a result of the (literal) differences

<sup>59</sup> Cf. Kripke 1980, p. 157; Lewis 1983a, p. 265.

<sup>60</sup> One might wonder: 'If it is indeterminate that some fact obtains in the case, then it is not *true* (that it does) and so how can we be permitted to import it?' Here a comparison with fiction is indeed useful: we are permitted to creatively embellish a fiction at the points where it is indeterminate (and, it seems, only there); e.g. we can imagine Hamlet with or without a bruise on his leg, as inhabiting a world where there are several species of hedgehog or only one ... other things equal, we are any which way imagining *Hamlet*. Likewise for Smith: we can imagine him as liking or disliking cabbage, as someone whose potential killing does or does not cause a riot ... any which way we are imagining *Smith*. This seems mysterious if we think of a fiction/problem case as a single possible world that is indeterminate with respect to certain facts. But, as Lewis points out, that is in any case implausible (Lewis 1983a, p. 270, esp. n. 11); better to think of a fiction/problem case as a *set* of possible worlds that differ from one another with respect to certain facts—and as long as a world belongs to that set it is a non-deviant realization of the fiction/problem case. (This may still not be the *best* way to think about it—at least not the best way to think about *fiction*. There are a number of problems with understanding fictions in terms of possible worlds at all, especially if we bar the divide and conquer strategy; see e.g. Lewis 1983a, p. 277; Proudfoot 2006; and, for a defence, Hanley 2004. But these problems do not seem to carry over to problem cases.)

<sup>61</sup> Again: at least when the case is used for its original purpose—to test Utilitarianism, or the Doctrine of Double Effect. See Foot 1978; Kamm 1989; Thomson 1976, 1985.

between the two case descriptions. Suppose that we explicitly ‘tag on’ a Trolley scenario to our sample description of the Gettier case. One might expect the resultant text to automatically inherit all the constraints—or rather, the intersection of the constraints—that govern the interpretation of its constituent case descriptions (in their normal contexts). But this is not automatic. What we may and may not import into the ‘Gettier-Trolley case’ once again depends on what *we are using it for*: whether we are using the case to test the JTB theory, Utilitarianism, both—or something else altogether.

Note that there is no analogous phenomenon in the case of standard fiction: the interpretation of a fictional text appears to be governed by certain *standing* constraints, *modulo* the author’s specific intentions—unless, of course, the text is being ‘hijacked’ for the purpose of a thought experiment. (We might, for instance, construct a Trolley case that features Hamlet. More on this below.)

These standing constraints—whatever exactly they are—also seem to impose a greater degree of overall similarity to the actual world (or better: the actual world as the author and her immediate audience *take it to be*<sup>62</sup>) on the fiction, than the corresponding constraints impose on a problem case. At any rate this goes for so-called ‘realistic’ fiction; perhaps outré fantasy, science fiction, and surreal fiction are associated with constraints that are more permissive in this respect. But let us set these genres aside.<sup>63</sup> In the case of realistic fiction, the audience is not only permitted but *required* to import a lot of extraneous ‘background’ facts that obtain in the actual world. These background facts, together with the stated facts, generate further extraneous facts that they are also required to import. Thus it is true—or ‘true in the fiction’—that Hamlet sometimes needs to go to the bathroom, and that water boils at 100°C, even though neither fact is explicitly stated (or is entailed by anything that is explicitly stated) in *The Tragedy of Hamlet*. And it is *false*, albeit consistent with the text, that Hamlet

<sup>62</sup> A qualification along these lines is needed to avoid anachronism, and too much specificity, in the fiction (e.g. to avoid the result that it is true in *The Tragedy of Hamlet* that there are exactly *n* species of hedgehog, where *n* is the actual number of species). Compare Lewis’s ‘Analysis 2’, in Lewis 1983a, p. 273. For further discussion of this aspect of fictional discourse, see Alward MS; Byrne 1993; Carlshamre 2004; Currie 1986, 1990; Eagle 2007; Hanley 2004; Proudfoot 2006; Sainsbury 2009; Walton 1990.

<sup>63</sup> It is safe to put them aside, since it would be very implausible to assimilate problem cases to, say, outré science fiction stories. (Although *some* extant case descriptions contain enough bizarre details to inspire such assimilation, most do not—e.g. standard descriptions of the Gettier case, the Trolley case, or even the Twin-Earth case.)

wears drag on Wednesdays, and that Ophelia has a local psychic power — say, a telekinetic ability to move cups across tables.

Contrast problem cases: on the face of it, Smith in the Gettier case might indeed wear drag on Wednesdays; he (or Brown or Jones) might even have a telekinetic ability to move cups across tables. It is not *true* — or ‘true in the problem case’ — that he does, but it is not false either: it is simply indeterminate. I take it that the corresponding claims about Hamlet and Ophelia are determinately false. Likewise for the claim that Hamlet never needs to go to the bathroom, and that, in his and Ophelia’s world, water boils at 67°C. In the Gettier case, however, water may or may not boil at that temperature, and Smith may or may not have the relevant physical need. All of this is *left open* by the constraints that govern the interpretation of case descriptions.<sup>64</sup>

It is worth noting that standard fictions too contain a lot of indeterminacy — even with respect to actual (or taken-for-actual) background facts. This suggests that we, *qua* audience of fiction, do not seek to *maximize* overall similarity between fiction and actuality. But the contrast between fictions and problem cases remains: absent stipulations to the contrary, much *more* overall departure from actuality appears to be permitted in the interpretation of case descriptions.

It is also worth noting that a philosophical problem case can, of course, be *presented as* a piece of standard fiction (just like a piece of fiction can be presented as, and serve the function of, a problem case in a thought experiment). Our sample case description might, for instance, be included in a collection of short stories — and, if it were, then most or all of the disparities that I have drawn attention to would presumably disappear.<sup>65</sup> But that is irrelevant. (Compare: both case descriptions and fictional texts can be presented as *news*

<sup>64</sup> To bring home the point, suppose that a thought experimenter informed us that, the way she imagines the Gettier case, Smith wears drag on Wednesdays and water boils at 67°C. In response, we might well complain that she has filled out the given description in *idiosyncratic* or *irrelevant* ways — irrelevant, since her colourful embellishments are not going to help her settle how the test properties are distributed in the case. But I do not think we would complain that her embellishments are *illegitimate*, or that the envisaged realization of the case is *deviant* — as indeed we would if she had filled it out in such a way that, say, Smith’s justification is defeated. (But is this not what we would do, if the FICTION proposal were correct: would we not complain that the realization at hand *is* deviant — and would we not be *right*?)

<sup>65</sup> Provided that neither author nor audience has any previous exposure to Gettier’s thought experiment, and that the question whether Smith has a justified true belief without knowledge is not salient — e.g. is stated in the immediately surrounding text. (Creative elaboration on the original case description — as in Lodge 2001, Ch. 16 — may further facilitate a ‘fictionalist’ reading.)

reports— with rather drastic effects on their proper interpretation — but I take it that this has little, if any, significance for the respective content problems.) The crucial question, for us, concerns the proper interpretation of a case description in its ‘home’ context: the context of a (specific) thought experiment. I have argued, in effect, that this context comes with its own set of interpretative principles — a set that differs in several ways from that which is normally operative in our understanding of fiction. And these differences combine to suggest that intuitive judgements and fictional judgements should not receive the same treatment.

In defence of FICTION, one might propose that the problematic disparities can be explained away with appeal to the relative *brevity* of a canonical case description (compared to, say, the text of *The Tragedy of Hamlet*, or — for that matter — the text of an average short story). But I doubt that the disparities can be explained away that easily. First, it seems clear that short — even *very* short — realistic fiction typically requires that we import more actual, or taken-for-actual, background facts than does a problem case.<sup>66</sup> Second, even characters in (very) short fiction are essentially fictional. And, third, the relative brevity of a standard case description cannot begin to explain the fact that what is true — and false and indeterminate — ‘in the case’ is partly determined by what theory the case is being used to test, and so may shift with a change in target theory. There is nothing about short fiction, in general, that predicts or explains this.

As the attentive reader may have realized, the third point raises a further important question: does FICTION even have the resources to classify all clearly deviant realizations of a problem case as deviant? Roughly put, a realization is deviant if it contains some extraneous facts that in obvious (and unintended) ways affect the distribution of

<sup>66</sup> For a stark example, consider the following piece of ‘flash fiction’: ‘For sale: baby shoes, never worn.’ (Allegedly Hemingway, undated, word length: 6.) The brevity of the text notwithstanding, this fiction too contains a lot of detail that is naturally explained (in part) by our operating with a principle of overall similarity to the taken-for-actual world. For example, presumably (it is determinately true in the fiction that) the baby died — before, during, or shortly after birth — but that it was not first abducted by aliens, that both of its parents breathe oxygen, that giant squids are illiterate, and that there is more than one continent on Earth, and so on. For two other examples, consider *The Joy* (Chekhov 1883, word length: 697), and *The Boy Who Cried Wolf* (Aesop, sixth century BC, word length in tr. by Gibbs 1994: 91) — although none of this is explicitly stated, I take it that Mitya in *The Joy* cannot fly, that he is a citizen of a tsardom, and that (he inhabits a world in which) water freezes at 0°C; and I take it that the shepherd-boy in Aesop’s fable has exactly one head, that he sometimes sleeps at night, and that (he inhabits a world in which) wolves are mortal.



the relevant test properties.<sup>67</sup> FICTION had better imply that everything that—for a given problem case and thought experiment—falls into this category is *false* ‘in the case’: that it is false in the Gettier case that Smith has another route to knowledge available, false in the Trolley case that killing the victim would cause a riot in which fifty people die, and so on. However, it is not at all clear that the constraints that govern our interpretation of standard fiction (long and/or short) *rule out* these facts—as opposed to just leaving them open. Needless to say, there is no general presumption in force to the effect that, unless stated, a fictional character never has a backup route to knowledge of a proposition they believe. (Or that killing her would never cause a riot, and so on.) That suggestion fails to account for the fact that what is true/false/indeterminate in a problem case depends in part on—and can shift with a change in—the target theory; furthermore, it has wildly implausible consequences for the case of standard fiction.

I return to this worry below—and elaborate on it in more detail—since it arises in almost identical form for the next (and final) competing proposal.

### 2.5 A hybrid view

Let me wrap up the discussion of rival content proposals by addressing an interesting hybrid—a hybrid of NECESSITY\* and FICTION—that was recently put forward by Jonathan Ichikawa and Benjamin Jarvis (Ichikawa and Jarvis 2009). On this view, a problem case is indeed a piece of standard fiction, but an intuitive judgement is not *simply* a fictional judgement (of the usual kind). Rather, it is a certain necessity judgement about a given fiction—a judgement roughly to the effect that, in any metaphysically possible non-deviant realization of a fiction, the test properties are distributed in such-and-such a way.<sup>68</sup> Importantly, the non-deviant realizations are supposed to be *demonstratively* identified (in thought): an intuitive verdict is a

<sup>67</sup> This is at best a sufficient condition for deviance, and a circular one at that, but this does not matter—we do not need to solve the problem in Sect. 2.3 to formulate the current objection to FICTION.

<sup>68</sup> I am going along with the authors’ assumption that (internal) fictional judgements ‘of the usual kind’ cannot be analysed as necessity judgements—at least not as necessity judgements along the lines of FICTION\*. (Here are three *prima facie* reasons why not: a FICTION\*-type analysis of fictional judgements fails to accommodate both impossible fictions and the essential non-actuality of fictional characters; it is also highly uninformative—hardly more informative than FICTION.)

generalization over all possible worlds in which *things are like that*—where ‘things are like that’ just in case everything that is true in the fiction is true *simpliciter*.

In application to the Gettier judgement, then, the new candidate content is the claim that,

FICTION\* Necessarily, if things are like *that*, then someone has a justified true belief that *p* but does not know that *p*.<sup>69</sup>

Given the additional premiss that things *could* be like that—in effect, that everything that is true in the fiction could be true *simpliciter*—FICTION\* secures a good inference. And the particular fiction in question here is, of course, the ‘Gettier fiction’: the fiction that results if we read a canonical description of the Gettier case as we would read the text of a standard fiction.

The first thing to notice is that FICTION\* does not avoid any of the objections that I raised above against FICTION. If what I argued there is correct, we do not (not normally—in the context of specific thought experiments) read case descriptions in the way that we read standard fictional texts. What is true/false/indeterminate in a problem case does not line up (across the board) with what is true/false/indeterminate in a fiction; in consequence, FICTION threatens to classify some non-deviant realizations of the Gettier case as deviant, and conversely. But it should be clear that the same goes for FICTION\*—the necessity operator and the demonstrative component make no difference at all with respect to this issue. For instance, FICTION\* counts any realization of the Gettier case where Smith has a local psychic power as deviant, but that does not seem right. (*Some* such realizations are presumably deviant, but not for that—or just for that—reason.) Moreover, a

<sup>69</sup> Ichikawa and Jarvis represent their proposal as follows:  $\Box[r \rightarrow \exists x \exists p (JT(x, p) \& \neg K(x, p))]$ , where *r* is the proposition that things are like that—or, I should add, some other proposition that is true if and only if all the propositions that are true in the fiction are true *simpliciter* (Ichikawa and Jarvis 2009, pp. 229–32; Ichikawa 2009, p. 442). But it is unclear what else *r* could be. The most obvious alternative—the proposition that *all the propositions that are true in the fiction are true simpliciter*—is explicitly ruled out by the authors’ insistence that a good candidate content must not include the concept of truth in fiction, or even the concept of a fiction. (By the same token, I assume, it must not include concepts like that of a story, or of what happens in a story, etc.) It seems that at least some of their reasons for insisting on this are confused—e.g. the observation that intuitive judgements about *actual* cases are not plausibly analysed in terms of truth in fiction (Ichikawa and Jarvis 2009, p. 228). But—given the restriction that they set themselves—that *things are like that* may well be the best way to gloss *r*. This is why, in the text, I present the proposal in the way that I do. (The authors acknowledge that the only other gloss they provide is psychologically implausible; Ichikawa and Jarvis 2009, p. 231.)

deviant realization that is overall sufficiently similar to the actual, or the taken-for-actual, world may come out non-deviant<sup>70</sup> — unless, of course, that realization is ruled out by some *other* general interpretative constraint(s) on standard fiction. But *is it*? The burden is on the proponent of FICTION\* to show this. And remember that it is not enough that the problematic facts — say, Smith’s justification being defeated — come out *indeterminate* ‘in the case’; they should come out *false*. (To make the point vivid: Smith’s justification being defeated is just not on a par with his disliking cabbage, or with Hamlet having a small bruise on his shoulder.)

It might be suggested that any such (i.e. any deviant but sufficiently nearby) realization is ruled out, not by some general interpretative constraint associated with standard fiction, but by a specific overriding intention on the part of the author/case designer — perhaps an intention that only appears in the context of a philosophical thought experiment. We might think of the proposed intention as an exercise of authorial authority: an author’s ability to stipulate what is true and false in her fiction. The idea would be that, by having a certain intention, the author/case designer *tacitly* stipulates that none of the problematic — the ‘deviance-making’ — facts obtain in the Gettier fiction. But there are big problems with this move (even if sense can be made of the notion of a tacit stipulation). Note first that, on this elaboration of FICTION\*, all the real work is being done by the case designer’s intention, rather than by the assimilation of problem cases to standard fiction. Furthermore, the ‘specification problem’ that faced NECESSITY\* crops up in almost identical form — as the problem of spelling out the content of the case designer’s intention (in a way that yields the right result, but that is neither question-begging nor too rich).

A related problem concerns the minor premiss in the present version of the Gettier inference: the claim that *things could be like that* — in effect, the claim that the Gettier fiction, as a whole, could be realized. Suppose it turns out that, although most of the propositions that are true in that fiction are compossible — including all those that are literally expressed by the case description — there are a few odd men out; perhaps because some of the contingent extraneous facts that

<sup>70</sup> Supposing that the Gettier fiction is a piece of realistic fiction. Williamson (2009, p. 466) has essentially the same worry, although he assumes that enough similarity to the actual (rather than the taken-for-actual) world is enough for a realization to be classified as non-deviant by FICTION\*. It is clear, however, that that assumption is dispensable. (Note also that FICTION\* may not classify any *actual* deviant realizations as non-deviant, since fictional characters are essentially fictional.)

we are required to import (to achieve enough overall similarity to the actual, or taken-for-actual, world) are incompatible with others. Ichikawa and Jarvis point out that, in many cases, an apparent impossibility in a fiction can be explained away as an indeterminacy (Ichikawa and Jarvis 2009, p. 234). That is correct, but it cannot *always* be explained away that way: some fictions really are impossible, in overt or subtle ways.<sup>71</sup> The worry here is not the sceptical worry that, for all we know, the Gettier fiction—or some other paradigm ‘problem case fiction’—is actually impossible.<sup>72</sup> The worry is that, in a hypothetical situation where it turns out to be impossible—specifically, where it turns out to be impossible in some subtle way that, on the face of it, has no bearing at all on the distribution of the test properties—we do not behave as we would behave if the FICTION\* proposal were correct. We would not take that discovery to show that the thought experiment *fails*. But that, it seems, is the behaviour that the proposal predicts (and recommends).

Finally, FICTION\* fails to capture the fact that, in making the Gettier judgement, we are attributing justified true belief without knowledge to the subject (or to any of a range of subjects), who is related to a proposition in a certain peculiar way. What FICTION\* says, in effect, is that any possible world where the case description is satisfied—and everything else that is true in the fiction is true—is a world where *some subject(s)* or other has a justified true belief without knowledge (in some proposition(s) or other): it need not be the same subject as the subject who plays the ‘Smith role’ in that world. Indeed, FICTION\* is true even if the *only* possible worlds in which someone plays the Smith role are worlds in which that subject does not have a justified true belief without knowledge (but someone else does). On the face of it, however, that is not enough to make the Gettier judgement true.<sup>73</sup>

<sup>71</sup> Arguably, a realistic fiction is never overtly (or explicitly) impossible—impossible in virtue of an overt inconsistency in the text—but it can still be subtly (or implicitly) impossible: impossible once all the inexplicit details are filled out. For more on this, See e.g. Proudfoot 2006; Sainsbury 2009, Ch. 4.

<sup>72</sup> Compare the argument against COUNTERFACTUAL in Sect. 1.5.

<sup>73</sup> The ‘first-stab’ proposal that Ichikawa and Jarvis consider (and reject) fares better in this respect; see Ichikawa and Jarvis 2009, p. 227. This problem is reminiscent of the ‘donkey anaphora’ problem facing Williamson’s preferred formalization of COUNTERFACTUAL—cf. n. 24. But, in the present case, we have more than just a technical problem—the issue arises even on the natural language formulation of FICTION\*.

## 3.

3.1 *Rationalism revisited*

I have argued that the real content of the Gettier judgement is a certain metaphysical possibility claim—and, in particular, that it is not a counterfactual conditional, a strict conditional, or a claim about fiction. It is time to return to the argument that sparked my discussion of the content problem: Williamson's argument against rationalism. In the rest of this paper, I consider the relation between the content problem and the question of rationalism in some detail. I argue that that relation is much less direct than Williamson takes it to be, and I close with a dilemma.

Williamson, recall, tries to get a lot of leverage out of his content proposal. In rough outline, he argues as follows:<sup>74</sup>

Intuitive judgements are judgements of certain counterfactual conditionals; we have a general capacity to handle counterfactuals; hence there is no need to invoke a special-purpose capacity or mechanism—such as a faculty of rational intuition—to explain the formation and the epistemic properties of intuitive judgements. The general capacity to handle counterfactuals is not 'exclusively a priori'; nor is there a principled way to single out only some of the judgements that this capacity delivers as a priori. So intuitive judgements are not a priori.

Here I would like to draw attention to the key assumption that we have a *general* capacity to handle counterfactuals. Is it legitimate to suppose that we do? I will argue that it is not; more precisely, that it is not legitimate to suppose that we have a general capacity *at the appropriate level of implementation*. (This will be explained shortly.) The immediate upshot is that the argument fails to go through, even if we *grant* that intuitive judgements are judgements of counterfactuals.

Importantly, the consideration that I am about to give also blocks a certain fallback manoeuvre otherwise available to Williamson: an argument against rationalism that, unlike the above, does *not* depend on his specific content proposal—indeed, that is compatible with mine. Williamson has recently argued that our judgements of *metaphysical modality* are not a priori (either).<sup>75</sup> But then it looks like he could in principle dispense with the claim that intuitive judgements are

<sup>74</sup> For more detail, see Sect. 1.2 above.

<sup>75</sup> See Williamson 2007a and 2007b, Ch. 5. The argument is foreshadowed in his 2005, Sect. 3. See also Hill 2006. (My objection applies *mutatis mutandis* to Hill's argument.)

judgements of counterfactuals, accept my alternative suggestion instead, and still reach the desired conclusion.<sup>76</sup> As we shall see, however, my next objection applies to the ‘fallback argument’ too.

We may summarize the fallback argument as follows:

Claims of metaphysical necessity and possibility are logically equivalent to certain counterfactual claims.<sup>77</sup>

This suggests that no special-purpose capacity or mechanism is needed to explain our judgements of metaphysical modality either: rather, our general capacity to handle counterfactuals is responsible for those judgements too. (Anything else would indicate a ‘bizarre lack of cognitive economy’ (Williamson 2007b, p. 162).) But, again, this general capacity is not exclusively a priori; nor is there a principled way to single out only some of the judgements it delivers a priori. Add to this that intuitive judgements are metaphysical possibility judgements — and it follows that rationalism is false.

The fallback argument raises some new concerns, but rather than pursuing those, let us take note of the features it shares with the original argument against rationalism: both arguments require that we have a general capacity to handle counterfactuals — a general-purpose cognitive capacity that is causally responsible for all, or at any rate most, of our counterfactual judgements.<sup>78</sup> Moreover, the explanation of our intuitive judgements that this capacity provides is supposed to *make redundant* an explanation in terms of a special-purpose capacity or mechanism. Last, the general capacity is supposed to provide, not just a causal explanation of our intuitive judgements, but an *epistemology* for them. Do we have a general capacity of the requisite sort?

Nothing of relevance here hinges on whether capacities are distinct from dispositions (or from abilities); for our purposes, then, a

<sup>76</sup> Thanks to an anonymous referee for pressing this point.

<sup>77</sup> Informally: ‘necessarily *p*’ is equivalent to ‘if it were not the case that *p*, a contradiction would be true’ and ‘possibly *p*’ is equivalent to ‘it is not the case that, if it were the case that *p*, a contradiction would be true’. (See Williamson 2007a, 2007b; Lewis 1973; Stalnaker 1968.)

<sup>78</sup> In his 2005 and 2007a, Williamson writes as if the capacity in question is responsible for *all* counterfactual judgements. In 2007b (pp. 151–2) he writes as if there are exceptions: the general capacity is only responsible for *most* or *typical* such judgements. On the face of it, this weakens the argument, absent compelling reason to think that intuitive judgements are not in fact atypical. But I will not press this point here. And the objection I give below is not substantially affected by this change of view since, as we will see, I am willing to grant that our intuitive judgements are explained by the general capacity Williamson has in mind — whatever the exact scope is of that capacity.

capacity is just a disposition, and a ‘cognitive’ capacity is a disposition to engage in some cognitive activity — such as reasoning, seeing depth, or writing a dissertation. But what it is to ‘handle’ a counterfactual? Judging from Williamson’s usage, it is (at least) to rationally assess it, and to determine its truth-value in light of that assessment.<sup>79</sup>

### 3.2 Capacities and counterfactuals

There is a sense in which it is uncontroversial, indeed *trivial*, that we have a general capacity to handle counterfactuals: for any activity  $\varphi$ , we can introduce the general capacity to  $\varphi$  — the unique disposition that is manifest when and only when someone  $\varphi$ s. In this sense, I obviously have the general capacity to  $\varphi$ , for any activity  $\varphi$  in which I am able to engage. The attribution to me of the general capacity to  $\varphi$  can still be both informative and explanatory — it can even provide a minimal explanation of particular  $\varphi$ -ings of mine. Thus my general capacity to learn languages explains my learning French, and my general capacity to handle counterfactuals explains why I believe that if I made the supper it would be inedible. Whether such explanations are causal explanations is debatable, but even if they are, it should be clear that they are not *exhaustive* — that in each case there is room for further explanations, in particular, room for ‘lower-level’ explanations: explanations that make reference to the specific way or ways that the general capacity to  $\varphi$  is realized or implemented in me.<sup>80</sup> (Presumably there is a hierarchy of implementation levels, but for brevity I will sometimes write as if there is only one level.) For many purposes, a lower-level explanation is what we need: the existence of a minimal explanation of the first mentioned sort — a so-called ‘virtus dormitiva’ explanation — certainly does not make this kind of explanation redundant.

Next, a virtus dormitiva explanation puts almost no constraints on the available lower-level explanations; importantly, it is compatible with lower-level explanations that trade in a variety of realizing mechanisms. It does not rule out the possibility that the target capacity is realized in many different ways, even in the same subject — that it is realized by a multitude of different mechanisms, processes, and/or further capacities (perhaps at each level of implementation).

<sup>79</sup> That is: form the belief that it is true/false (or neither), in light of that assessment.

<sup>80</sup> Neurophysiological and biological explanations would be clear examples. But importantly, there is also room for further *psychological* explanations. (See Cummins 2000; Davies 2005; Marr 1982.)

Furthermore, it tells us nothing about the detailed workings of the realizing mechanism(s).

The upshot of this should be clear: on the current reading, the claim that we have a general capacity to handle counterfactuals is *obviously* true but, on that reading, it is compatible with there being many different cognitive processes, mechanisms, and/or further capacities involved in the production of counterfactual judgements. In particular, it leaves open that there is a special-purpose capacity or mechanism involved in the evaluation of all ‘philosophically interesting’ counterfactuals (such as COUNTERFACTUAL<sup>81</sup>) and that this special-purpose capacity is exclusively a priori. It even leaves open that the evaluation of philosophically interesting counterfactuals recruits a faculty of rational intuition—whatever exactly that is supposed to be. Those who postulate such a faculty are not plausibly interpreted as saying that their account precludes a *virtus dormitiva* explanation of the relevant class of judgements.

Of course, there are strong *independent* reasons to reject the view that some counterfactual judgements, or for that matter any other judgements, are arrived at by a faculty of rational intuition—at least as that faculty is traditionally conceived: as a quasi-perceptual organ that somehow gives us direct ‘insight’ into the truth of the contents in question.<sup>82</sup> Crucially, none of its advocates has been able to explain how this faculty would work, or to provide independent evidence for its existence.<sup>83</sup> But Williamson does not reject the view on these, familiar grounds; he rejects it on the grounds that it is unmotivated. On the current reading, however, the claim that we have a general capacity to handle counterfactuals does not in fact undercut the motivation for the view. (Nor does it undercut the motivation for postulating some less extravagant special-purpose capacity.) Presumably, then, this is not the intended reading.

What Williamson needs is *generality at the appropriate level of implementation*: that our general capacity to handle counterfactuals, identified as such, be realized in us by another highly general capacity or mechanism; indeed, it must be thus realized at the *very same* level of implementation at which the faculty of rational intuition, or other

<sup>81</sup> Or counterfactuals with a contradiction as the consequent (see n. 77).

<sup>82</sup> Or that gives us direct insight into some necessary truths from which the relevant counterfactuals follow (cf. Sect. 3.4).

<sup>83</sup> See Boghossian 2003; Wright 2004.



special-purpose mechanism, is supposed to be located. Only then do we have two genuinely competing explanations on the table.

Whether there is generality at the appropriate level is at least in part an empirical question. But the empirical data is notoriously hard to interpret — especially, of course, for a non-expert — and the relevant literature is only of limited help: the research on counterfactual processing is still at an early stage, and many of the going theories are seriously underdeveloped. That said, there is one issue on which there is already near-consensus: most cognitive scientists who work on the topic seem to agree that the evaluation of counterfactuals is far from a unified affair — that it involves many different capacities and/or mechanisms. Which one gets recruited in a specific case appears to depend, among other things, on the content and complexity of the given counterfactual claim, and the pragmatic concerns and background beliefs of the subject.<sup>84</sup> This near-consensus may seem to suggest that all the available evidence points towards a negative answer. But that would be too quick.

### 3.3 *Mental simulation*

According to Williamson, the way the general capacity to handle counterfactuals is realized in us, in the vast majority of cases, is as a capacity to run certain *other* cognitive capacities ‘offline’, or as a capacity for offline *mental simulation*. Here is a representative passage:

We have our ordinary capacities for making judgements about what we encounter, and a further capacity to evaluate counterfactuals by running those capacities ‘offline’; that is already enough for philosophy to get going, without any need of a kick-start from a special faculty of intuition. (Williamson 2005, p. 18)<sup>85</sup>

In general, and to a first approximation, to say that a given capacity is (or is realized by) a capacity for offline mental simulation is to say that it involves the ‘redeployment’ of some other cognitive capacity or mechanism — one that normally performs a different function. And the primary difference between the normal (online) and the offline function does not lie in the type of *information*, if any, that the capacity or mechanism draws on, but rather in the type of *representational state* that it takes as input — for example, perhaps it manipulates

<sup>84</sup> See e.g. Byrne 2005; Evans and Over 2004; Kahneman 2003; Roese 1997; Sloman 2005; Stanovich and West 2003.

<sup>85</sup> See also his 2007a, Sect. 3, and 2007b, pp. 147–55.

desires and beliefs when it is run online, but manipulates functionally similar pretence correlates (so-called ‘pretend beliefs’ and ‘pretend desires’) when run offline. Other than that, the capacity behaves in much the same way in both cases.<sup>86</sup>

This is very rough, but it should be enough to appreciate two crucial points: in contrast to the uncontroversial (but trivial) claim discussed above, the claim that our general capacity to handle counterfactuals is for the most part realized as a capacity for mental simulation is far from trivial — it is a substantial empirical hypothesis.<sup>87</sup> However, it is a *highly abstract* empirical hypothesis. The details of the underlying cognitive machinery are left wide open: what they are like depends in the first instance on what other capacity or mechanism is being run offline in any given simulation exercise (and on how *that* capacity or mechanism in turn is implemented). The hypothesis also leaves open — as it surely should — that what is being run offline in the evaluation of a given counterfactual claim may depend, among other things, on the specific *content* of that counterfactual.

In one way, this is good news for Williamson — mental simulation is a suitably high-level realizer, and so the simulation hypothesis is compatible with the near-consensus noted earlier: that the evaluation of counterfactuals involves a rich variety of low-level cognitive capacities or mechanisms.<sup>88</sup> But the abstractness of the hypothesis also makes it uncongenial to his overall aims. On the face of it, nothing prevents his rationalist opponent from endorsing it: the only type of explanation that the simulation hypothesis precludes is a ‘theory-based’ explanation — one that postulates a body of tacitly known, domain-specific information to account for the evaluation of (some or all) counterfactuals. But there is no principled reason why a ‘special

<sup>86</sup> See Evans and Over 2004, and Goldman 1992, for simulation-based accounts of our capacity for counterfactual reasoning. Other capacities that have been explained in terms of mental simulation include our capacity for mental imagery (Currie 1994), our capacity to predict behaviour (Goldman 1989; Gordon 1985), and our capacity for empathy (Goldman 1993). For useful discussions, see Davies and Stone 1998; Nichols, Stich, Leslie, and Klein 1996; Weinberg and Meskin 2006.

<sup>87</sup> At least this is the most plausible way to understand it in this context. (For a contrasting approach, see Heal 1998.)

<sup>88</sup> This compatibility is noted by Evans and Over, whose simulation-based account of counterfactual processing is among the best developed in the literature. They present their account as ‘a high level description of many different processes that contribute to hypothetical thought’ (Evans and Over 2004, p. 158). Apparently Williamson too understands the simulation hypothesis this way (see his 2007b, p. 152), but he does not appreciate the problem this creates for his argument.

capacity rationalist' — say, the advocate of a faculty of rational intuition — would have to subscribe to that. For all that has been said, the relevant special purpose capacity may have an online as well as an offline use, or it may be part of some more complex capacity that does.

Take the case of interest here. Perhaps my intuitive judgement about an actual problem case, and my judgement about the corresponding hypothetical case are — at some relatively high level of implementation — generated by the same capacity (run offline and online respectively). On one model, a certain inferential mechanism is involved: when run online, it takes me from certain beliefs about the actual case — for example, the belief that Smith based his belief on a false lemma — to the judgement that Smith has a justified true belief without knowledge; when run offline, it takes me from certain suppositions and pretend-beliefs to the Gettier judgement (on present assumptions: a counterfactual judgement). This inference mechanism draws on information that in turn is provided by means of some special purpose capacity — for example, it draws on the 'rational insight' that, necessarily, anyone who relies on a false lemma fails to know — and it does so in *both* cases.

Importantly, that does not imply that both of the resulting judgements are a priori: since some of the beliefs that are implicated in the online use of the mechanism are (at best) empirically justified, the resulting judgement — that Smith has a justified true belief without knowledge — is empirical too. And I take it that this is the right thing to say about that judgement. But the *offline* use of the mechanism may still yield a judgement that is a priori. The suppositional reasoning process that generates my intuitive judgement about the hypothetical case may or *may not* draw on any empirical evidence; if it does not, and if my judgement is justified, it is a priori justified. Of course more must be said, but the present point is just that the simulation hypothesis is strictly compatible with the view that intuitive judgements about hypothetical problem cases are a priori;<sup>89</sup> indeed, it is compatible with the view that they are arrived at by means of a faculty of rational intuition.

To sum up: an explanation of our counterfactual judgements in terms of mental simulation does not preclude or make redundant further explanations of some (or all) such judgements and, although it is less permissive than a *virtus dormativa* explanation, it still does

<sup>89</sup> Note that this is overlooked by Yablo 2002.

not constrain these explanations enough — in fact, it too is compatible with the most extreme form of special-capacity rationalism. The upshot is that Williamson's argument against rationalism (or his arguments, counting the fallback manoeuvre) does not go through even on the second — more substantial — reading of the claim that we have a general capacity to handle counterfactuals.

At this point, a natural move would be to explore whether there is some further reading on which the argument *does* go through. But we can spare ourselves the effort: there are principled reasons to think that there is no such reading.

### 3.4 *The epistemology of counterfactuals?*

The assimilation of intuitive judgements to counterfactual judgements is supposed to undermine the motivation for special-capacity rationalism. But recall that that view is motivated by *two distinct* explanatory needs — the special-purpose capacity is invoked to explain both the formation of intuitive judgements *and what makes them justified and reliable* (cf. Sect. 1.2 above). Accordingly, the general capacity to handle counterfactuals must be able to provide a competing *epistemology* for intuitive judgements (in addition to a competing causal-psychological explanation). Is it fit for the purpose?

It should be clear by now that it is not, if conceived as a capacity for mental simulation: the simulation hypothesis allows for a variety of different mechanisms at lower levels of realization and, crucially, leaves it open that a difference in realizing mechanism at some such level may make an epistemic difference.<sup>90</sup> But in fact that was to be *expected*, given the scope of the hypothesis. By all appearances, counterfactual judgements can be justified in a variety of ways: deductively, inductively, by memory or testimony, etc. What kind of justification one has to believe a given counterfactual depends in the first instance on *which* counterfactual it is — what content it has — and on one's overall cognitive and epistemic situation. This should be totally uncontroversial: if there is variation of this sort at all — if there are interestingly different kinds of justification and knowledge — then there is variation of this sort among our counterfactual judgements. But then *any* hypothesis — a fortiori any causal-psychological hypothesis — concerning counterfactual judgements as such is bound to be very limited in its epistemological implications. We should certainly not expect to get an (acceptable) epistemology for some specific

<sup>90</sup> This is noted by Peacocke, in his 1998, p. 132.

subclass of counterfactual judgements out of it — say, for our judgements of philosophically interesting counterfactuals.

Notably, counterfactual judgements are heterogeneous also in the following respect: by all appearances, some such judgements are capable of being a priori justified, whereas others are not. The only kind of justification that I could have for believing that if I had made the supper it would have been inedible, or that if I fell out the window I would break my neck, is a posteriori justification. Likewise for many other actual and possible counterfactual beliefs — but not *all*. Thus, for instance, my belief that if twelve people had been killed more than eleven people would have been killed,<sup>91</sup> my belief that if God had existed then God would have existed, and my belief that Balder would be a bachelor if he were an unmarried man, are all, on the face of it, capable of being a priori justified. (Indeed, they are as good candidates as *any*.)

This casts further doubt on the idea that we can give an epistemology for intuitive judgements — explain what makes them justified and reliable — in terms of the general capacity to handle counterfactuals, *regardless* of how that capacity is conceived: as a capacity in the minimal sense, as a capacity for mental simulation, or as something else altogether. Any account that is pitched at this level of generality fails to sustain a number of central epistemic distinctions — including that between a priori and a posteriori justification. (Furthermore, for all we have been told so far, our judgements of philosophically interesting counterfactuals fall on the a priori side of the distinction more often than not.) Of course, if there were *independent* reason to think that the a priori/a posteriori distinction does not survive scrutiny — that, ultimately, there is no such distinction — then the failure of a theory to sustain that distinction would not be a major flaw. But if we had independent reason to think that, we would also have a much quicker way with rationalism: we could simply argue that, since nothing is a priori, intuitive judgements are not a priori. The appeal to the general capacity to handle counterfactuals now drops out as irrelevant.<sup>92</sup>

<sup>91</sup> Cf. Williamson 2007a, p. 12, and 2007b, p. 143.

<sup>92</sup> Another conceivable tactic, on which that appeal remains essential, would be to argue that there is no a priori/a posteriori distinction among *counterfactual judgements in particular*. But how would one argue that? What reason could there possibly be for doubting the a priori of my belief that if God were to exist then God would exist that is not also, at the very least, a reason to doubt the a priori of my belief in the corresponding metaphysical necessity claim or the corresponding indicative conditional?

We seem to have reached a dilemma: either there is independent reason to doubt that the a priori/a posteriori distinction survives scrutiny, or there is not. If there is, then the assimilation of intuitive judgements to judgements of counterfactuals does no work in the argument against rationalism. If there is not, then a plausible epistemological theory should be able to accommodate that distinction. A theory that (just) appeals to our general capacity to handle counterfactuals violates this constraint, since it fails to discriminate between different counterfactual judgements.

Clearly, then, there is explanatory work (of the second sort too) left to do: the general capacity to handle counterfactuals is simply *too* general to be of much use in explaining what makes intuitive judgements reliable and justified—even if those judgements are judgements of counterfactuals. Whether the best explanation involves appeal to a special-purpose capacity is a further question. But it is a question that can only be settled by assessing individual candidate explanations on their merits.

### 3.5 Final remarks

Williamson acknowledges that there are prima facie paradigms of the a priori and the a posteriori to be found among our counterfactual judgements, but he goes on to argue that appearances deceive: that the distinction does not survive scrutiny (at least not beyond a narrow range of ‘easy cases’).<sup>93</sup> For our purposes, the important point is that the main reasons he provides for thinking so are *perfectly general*—they do not require mention of the capacity to handle counterfactuals. Thus it looks like he opts for the first horn of the dilemma.

To clarify: Williamson argues that there are cases where the subject’s perceptual experience plays a role, in the formation of her (justified or knowledgeable) judgement, that is neither strictly *evidential* nor purely *enabling*. If the judgement were a posteriori, experience would play an evidential role; if it were a priori, experience would (at most) play an enabling role—thus, it seems, the judgement is neither a priori nor a posteriori. The central example he uses to argue this point involves a certain counterfactual judgement, but the argument is supposed to generalize to many other counterfactual judgements—and beyond. That it generalizes is crucial, since the existence of a few hard (even borderline) cases does not undermine the a priori/a posteriori distinction—at most it shows that the distinction is not perfectly sharp.

<sup>93</sup> See his 2007a, pp. 32–3, and 2007b, Ch. 5, Sect. 5.

Two considerations are provided: first, he argues that our perceptual experience can (and sometimes does) influence the *reliability* of our cognitive capacities — including the capacities that are run offline when we engage in mental simulation. Derivatively, it can affect the reliability with which our simulation-based judgements are produced. This is not yet enough for experience to play a strictly evidential role (in the formation of these judgements). But it makes for *more* than a purely enabling role, since reliability is epistemically relevant — according to Williamson, it is essential to both knowledge and justification.<sup>94</sup> The possibility of this ‘intermediate’ role for experience is illustrated with a case in which a subject arrives at the judgement that *if two marks had been nine inches apart, they would have been at least nineteen centimeters apart*, by running her capacity to make ‘naked eye judgements’ of distances in inches and centimetres offline. (She does not know the conversion ratio.) And that capacity is supposed to depend for its reliability — both online and offline — on her past perceptual experience (see Williamson 2007b, pp. 165–6).

The judgement in the example is a counterfactual judgement, but it is clear that a judgement with a similar causal history and a different content (e.g. the corresponding indicative, material, or strict conditional) would have served equally well: it is the fact that the judgement is simulation-based — not that it has a counterfactual content — that does the work in the example.<sup>95</sup> This does not yet show that *the capacity to handle* counterfactuals is idle in the argument since, for Williamson, that capacity *just is* a capacity for mental simulation. However, the example, on its own, does not take us very far: the hard work lies in showing that experience plays the specified intermediate role in a critical mass of cases. And that is by no means obvious. In particular, it is not obvious that it plays that role relative to any judgements that are *prima facie* candidates for being a priori justified — not even relative to any (in the present context) *controversial* candidates, such as the Gettier judgement.<sup>96</sup>

<sup>94</sup> See, in particular, his 2007b, p. 166. (Nothing here hinges on the difference between ‘safety’ and other notions of reliability; cf. Williamson 2000a.)

<sup>95</sup> And it is very implausible that *only* judgements of subjunctive conditionals are simulation-based (as Williamson would be the last to dispute) — one would at least expect the simulation hypothesis to extend to judgements of other conditionals (cf. Byrne 2005; Evans and Over 2004).

<sup>96</sup> The problem is this: it is indeed plausible that many of our cognitive capacities depend for their reliability on repeated *practice* — practice in applying the target capacity. Such practice may or may not involve specific perceptual experiences on the part of the subject. In some

It is at this point that Williamson resorts to perfectly general considerations. A few judgements with a *prima facie* claim to a priority — including the Gettier judgement, the judgement that it is necessary that whoever knows something believes it, and the judgement that whoever knew something believed it — are explicitly assimilated to the comparative distance judgement in the central example (Williamson 2007b, pp. 168, 188). But he does not explain why we should think that the *prima facie a priori* judgements, like the comparative distance judgement, are based on the offline use of a capacity that crucially depends on perceptual experience for its reliability. Instead he outlines, and rejects, a number of specific *positive accounts* of the epistemology of these judgements — ‘traditional’ accounts, on which experience either plays a strictly evidential, or a purely enabling, role. The upshot is supposed to be that there is no plausible account of either sort available. (To say that, of course, is just to say that there is no plausible account available on which the judgements come out a posteriori, and a priori, respectively.)

Williamson does not examine very many candidates — notable omissions include special-capacity rationalism<sup>97</sup> — but let us suppose that he examines the best. The crucial point is that his criticism of the selected candidates does not mention, or presuppose, that any of the judgements whose status is in question are simulation-based, or that

---

cases, it seems to *require* it; the present example is a case in point — it is impossible to practice the capacity to judge distances in inches and centimetres *by sight* in the absence of certain visual experiences. But that is because the capacity at issue is a (partly) *perceptual* capacity: it operates on specifically visual input (when performing its normal function, i.e. when run online). In other cases, the practice of a capacity typically — or even invariably, in the actual world — involves experience, but it does not require it; e.g. the practice of various mathematical capacities typically involves experiences of marks on paper; however, it is possible — say, for someone with abnormally strong short-term memory — to practice the same capacities, equally well, in their head). Now, I take it that experience plays a purely enabling role in the latter type of case — e.g. it plays a purely enabling role in the formation of our mathematical judgements, even if our experience-involving practice significantly affects the reliability with which those judgements are produced (and even granted that reliability is epistemically relevant). The former type of case is the only one that is at all problematic — the type that is exploited in Williamson’s central example. Is that enough? As I just pointed out, the judgement in that example is based on the offline use of a certain *perceptual* capacity — without the aid of any perceptual beliefs, or any other perceptual reasons. Why should we think that a wide range of judgements fall in this category? For one thing, it seems very implausible that any judgements with a *prima facie* claim to a priority fall in it (including those of most interest here: intuitive judgements about cases).

<sup>97</sup> Of course he takes himself to have refuted that view already; but, if what I argued above is correct, he has not.



they have counterfactual—or other modal—contents. The individual account that receives the most attention is the (rationalist) view that the judgements are ‘epistemically analytic’—very roughly, that they are explained and justified by the mere fact that we *understand* their contents. Against this, he argues that it is possible to understand each of the contents at issue—for example, COUNTERFACTUAL—without (justifiably) believing that content. And, as anyone familiar with Williamson’s overall work on that topic will know, this is part of an extended effort to show that there are *no* epistemically analytic judgements.<sup>98</sup>

Fortunately, there is no need for us to engage that effort here—the present (and final) point is dialectical: it looks like the weight of the argument against the a priori/a posteriori distinction is carried by considerations that are completely unrelated to the simulation hypothesis, and to the assimilation of intuitive judgements to judgements of counterfactuals; namely the alleged absence of a good positive account of the a priori (and/or of the a posteriori)—one with wide enough scope to take us beyond the narrow range of ‘easy cases’. Williamson’s criticism of extant accounts may, for all I have said, be compelling—but nowhere does it appeal to our general capacity to handle counterfactuals. Thus that capacity does in fact drop out as irrelevant.<sup>99</sup>

## References

- Aesop: ‘The Boy Who Cried Wolf’, in *Aesop’s Fables, A New Translation by Laura Gibbs*, ed. Laura Gibbs, Oxford World Classics. Oxford: Oxford University Press, 2002.
- Alexander, Joshua and Jonathan Weinberg 2007: ‘Analytic Epistemology and Experimental Philosophy’. *Philosophy Compass*, 2, pp. 56–80.
- Alward, Peter MS: ‘Truth in Fiction: The Story Concluded’.

<sup>98</sup> See e.g. his 2003b and 2007b, Ch 4. See also Boghossian 2003, and forthcoming.

<sup>99</sup> Thanks to Paul Boghossian, Albert Casullo, Helge Malmgren, Aidan McGlynn, Michael Raven, Stephen Schiffer, David Sosa, Michael Strevens, Timothy Williamson, and Crispin Wright for helpful discussion and feedback. Thanks also to the participants of the Arché Epistemology Seminar at the University of St Andrews, the General Colloquium at the University of Texas at Austin, and the Dissertation Seminar at New York University, where earlier versions of this paper were presented in the fall of 2006. Finally, thanks to the editor and the anonymous referees of this journal for their detailed and stimulating criticism.

- Bealer, George 1996: 'A Priori Knowledge and the Scope of Philosophy'. *Philosophical Studies*, 81, pp. 121–42.
- 2000: 'A Theory of the A Priori'. *Pacific Philosophical Quarterly*, 81, pp. 1–30.
- 2002: 'Modal Epistemology and the Rationalist Renaissance'. In Gendler and Hawthorne 2002, pp. 71–125.
- Benacerraf, Paul and Hilary Putnam (eds) 1983: *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press.
- Beyer, Christian and Alex Burri (eds) 2007: *Philosophical Knowledge: Its Possibility and Scope*. Amsterdam: Rodopi.
- Bishop, Michael and Dominic Murphy (eds) 2009: *Stich and His Critics*. Oxford: Wiley-Blackwell.
- Boghossian, Paul 1989: 'Content and Self-Knowledge'. *Philosophical Topics*, 17, pp. 5–26.
- 2003: 'Epistemic Analyticity: A Defence'. *Grazer Philosophische Studien*, 66, pp. 15–35.
- 2008a: 'Epistemic Rules', in his 2008b, pp. 109–34.
- 2008b: *Content and Justification: Philosophical Papers*. Oxford: Oxford University Press.
- forthcoming: 'Williamson on the A Priori and the Analytic'. Forthcoming in *Philosophy and Phenomenological Research*.
- Boghossian, Paul and Crispin Wright MS: 'Reflections on Rules, Regresses and Rational Belief'.
- BonJour, Laurence 1985: *The Structure of Empirical Knowledge*. Cambridge MA: Harvard University Press.
- Brown, J. Robert 1991: *Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.
- Burge, Tyler 1988: 'Individualism and Self-Knowledge'. *Journal of Philosophy*, 85, pp. 649–63.
- Byrne, Alex 1993: 'Truth in Fiction: The Story Continued'. *Australasian Journal of Philosophy*, 71, pp. 24–35.
- Byrne, Ruth 2005: *The Rational Imagination: How People Create Alternatives to Reality*. Cambridge, MA: MIT Press.
- Carlshamre, Staffan 2004: 'Truth in (of) Fiction'. *Ursus Philosophicus: Essays Dedicated to Björn Haglund on his Sixtieth Birthday*, Philosophical Communications, Web Series, 32.
- Carruthers, Peter 2006: *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.

- forthcoming: ‘How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition’. Forthcoming in *Behavioural and Brain Sciences*.
- Carruthers, Peter and Peter Smith (eds) 1996: *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Chalmers, David 1996: *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
- Chekhov, Anton 1883: ‘The Joy’, trans. Patrick Miles. In *Anton Chekov, Early Stories*, Oxford World Classics, trans. and ed. Patrick Miles and Harvey Pitcher. New York: Oxford University Press, 1999.
- Cummins, Robert 1998: ‘Reflection on Reflective Equilibrium’. In DePaul and Ramsey 1998, pp. 113–28.
- 2000: “‘How Does it Work’ vs. ‘What Are the Laws?’ Two Conceptions of Psychological Explanation’. In Keil and Wilson 2000, pp. 117–44.
- Currie, Gregory 1986: ‘Fictional Truth’. *Philosophical Studies*, 50, pp. 195–212.
- 1990: *The Nature of Fiction*. New York: Cambridge University Press.
- Dancy, Jonathan 2001: *Practical Reality*. Oxford: Clarendon Press.
- 2004: *Ethics Without Principles*. Oxford: Clarendon Press.
- Davies, Martin 1994: ‘The Mental Simulation Debate’. *Philosophical Issues*, 5, pp. 189–218.
- 2001: ‘Mental Simulation, Tacit Theory and the Threat of Collapse’. *Philosophical Topics*, 29, pp. 127–73.
- 2005: ‘Cognitive Science’. In Jackson and Smith 2005, pp. 358–96.
- Davies, Martin and Tony Stone 1998: ‘Folk Psychology and Mental Simulation’. In O’Hear 1998, pp. 53–82.
- DePaul, Michael and William Ramsey (eds) 1998: *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*. Lanham, MD: Rowman & Littlefield.
- Devitt, Michael 1994: ‘The Methodology of Naturalistic Semantics’. *Journal of Philosophy*, 91, pp. 545–72.
- Eagle, A. 2007: ‘Telling Tales’. *Proceedings of the Aristotelian Society*, 107, pp. 125–47.
- Earman, John, John Roberts, and Sheldon Smith 2002: ‘Ceteris Paribus Lost’. *Erkenntnis*, 57, pp. 281–301.
- Evans, Gareth 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

- Evans, Jonathan and David Over 2004: *If*. Oxford: Oxford University Press.
- Fodor, Jerry 1989: *Psychosemantics*. Explorations of Cognitive Science Series, 2, Bradford Books. Cambridge, MA: MIT Press.
- Foot, Philippa 1978: *Virtues and Vices*. Oxford: Basil Blackwell.
- Keil, Frank and Robert Wilson (eds) 2000: *Explanation and Cognition*. Cambridge MA: MIT Press.
- Geach, Peter 1962: *Reference and Generality: An Examination of Some Medieval and Modern Theories*. Ithaca, NY: Cornell University Press.
- Gendler, Szabo Tamar and John Hawthorne (eds) 2002: *Conceivability and Possibility*. Oxford: Oxford University Press.
- Gettier, Edmund 1963: 'Is Justified True Belief Knowledge?' *Analysis*, 23, pp. 121–3.
- Gilovich, Thomas, Dale Griffin, and Daniel Kahneman (eds) 2002: *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge: Cambridge University Press.
- Ginet Carl 1975: *Knowledge, Perception, and Memory*. Springer Science and Business.
- Goldman, Alvin 1976: 'Discrimination and Perceptual Knowledge'. *Journal of Philosophy*, 73, pp. 771–91.
- 1989: 'Interpretation Psychologized'. *Mind and Language*, 4, pp. 161–85.
- 1992: 'In Defense of the Simulation Theory'. *Mind and Language*, 7, pp. 104–19.
- 2006a: 'Imagination and Simulation in Audience Responses to Fiction'. In Nichols 2006b, pp. 41–56.
- 2007: 'Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status'. *Grazer Philosophische Studien*, 74, pp. 1–26. Also published in Beyer and Burri 2007, pp. 1–26.
- Gödel, Kurt 1964: 'What is Cantor's Continuum Problem?' In Benacerraf and Putnam 1983, pp. 470–85.
- Gordon, Robert 1986: 'Folk Psychology as Simulation'. *Mind and Language*, 1, pp. 158–71.
- Graham, George and Terrence Horgan 1998: 'Southern Fundamentalism and the End of Philosophy'. In DePaul and Ramsey 1998, pp. 189–202.
- Grundmann, Thomas and Joachim Horvath MS: 'On the Logic and Structure of Thought Experiments'.
- Gunderson, Keith (ed.) 1975: *Language, Mind, and Knowledge*. Minneapolis: University of Minnesota Press.

- Häggkvist, Sören 1996: *Thought Experiments in Philosophy*. Stockholm: Almqvist & Wiksell.
- Hanley, Richard 2004: 'As Good as it Gets: Lewis on Truth in Fiction'. *Australasian Journal of Philosophy*, 82, pp. 112–28.
- Harman, Gilbert 1968: 'Knowledge, Inference, and Explanation'. *American Philosophical Quarterly*, 5, pp. 164–73.
- Heatherington, Stephen (ed.) 2006: *Epistemology Futures*. Oxford: Oxford University Press.
- Hemingway, Ernest. undated, untitled: <<http://www.litkicks.com/FlashFiction/>>.
- Hill, Christopher 2006: 'Modality, Epistemology, and the Metaphysics of Consciousness'. In Nichols 2006b, pp. 205–36.
- Hintikka, Jaakko 1999: 'The Emperor's New Intuitions'. *Journal of Philosophy*, 96, pp. 127–47.
- Hooker, Brad and Margaret Olivia Little (eds) 2000: *Moral Particularism*. Oxford: Clarendon Press.
- Ichikawa, Jonathan 2009: 'Knowing the Intuition and Knowing the Counterfactual'. *Philosophical Studies*, 145, pp. 435–43.
- Ichikawa, Jonathan and Benjamin Jarvis 2009: 'Thought-Experiment Intuitions and Truth in Fiction'. *Philosophical Studies*, 142, pp. 221–46.
- Jackson, Frank 1998: *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
- Jackson, Frank, Philip Pettit, and Michael Smith 2000: 'Ethical Particularism and Patterns'. In Hooker and Little 2000, pp. 79–99.
- Jackson, Frank and Michael Smith (eds) 2005: *The Oxford Handbook of Contemporary Analytic Philosophy*. Oxford: Oxford University Press.
- Kaebnick, Gregory 1999: 'Particularist Moral Reasoning and Consistency in Moral Judgments'. *The Journal of Value Inquiry*, 33, pp. 43–56.
- Kagan, Shelley 2001: 'Thinking About Cases'. *Social Philosophy and Policy*, 18, pp. 44–63.
- Kamm, Francis Myrna 1989: 'Harming Some to Save Others'. *Philosophical Studies*, 57, pp. 227–60.
- Katz, Jerrold 1998: *Realistic Rationalism*. Cambridge, MA: MIT Press.
- Kornblith, Hilary 2002: *Knowledge and its Place in Nature*. Oxford: Oxford University Press.
- 2005: 'Replies to Alvin Goldman, Martin Kusch, and William Talbott'. *Philosophy and Phenomenological Research*, 71, pp. 427–41.

- 2006: 'Appeals to Intuition and the Ambitions of Epistemology'. In Heatherington 2006, pp. 10–25.
- Kripke, Saul 1980: *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, David 1973: *Counterfactuals*. Oxford: Blackwell.
- 1983a: 'Truth in Fiction', in his 1983b, pp. 261–80.
- 1983b: *Philosophical Papers*, volume 1. New York: Oxford University Press.
- Lodge, David 2001: *Thinks*. New York: Viking.
- Luper, Steven (ed.) 2003: *The Sceptics: Contemporary Essays*. Burlington, VT: Ashgate.
- Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen Stich 2004: 'Semantics, Cross-Cultural Style'. *Cognition*, 92, pp. B1–B12.
- Malmgren, Anna-Sara MS 2009: 'Essays in the Theory of Justification'. PhD dissertation, New York University, ProQuest/UMI.
- Margolis, Eric and Stephen Laurence 2003: 'Should We Trust Our Intuitions? Deflationary Accounts of the Analytic Data'. *Proceedings of the Aristotelian Society*, 103, pp. 299–323.
- Marr, David 1982: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman: New York.
- McDowell, John 1977: 'On the Sense and Reference of a Proper Name'. *Mind*, 86, pp. 156–85.
- 1979: 'Virtue and Reason'. *The Monist*, 62, pp. 331–50.
- Nichols, Shaun 2004: 'Imagining and Believing: The Promise of a Single Code'. *Journal of Aesthetics and Art Criticism* (Special Issue on Art, Mind and Cognitive Science), 62, pp. 129–39.
- Nichols, Shaun 2006a: 'Imaginative Blocks and Impossibility', in his 2006b, pp. 237–56.
- 2006b: *The Architecture of the Imagination: New Essays on Pretence, Possibility, and Fiction*. Oxford: Oxford University Press.
- Nichols, Shaun and Stephen Stich 2003: *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.
- Nichols, Shaun, Stephen Stich, Alan Leslie, and David Klein 1996: 'Varieties of Off-Line Simulation'. In Carruthers and Smith 1996, pp. 39–74.

- Nichols, Shaun, Stephen Stich, and Jonathan Weinberg 2003: 'Metaskepticism: Meditations in Ethno-Epistemology'. In Luper 2003, pp. 227–47.
- O'Hear, Anthony (ed.) 1998: *Current Issues in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Peacocke, Christopher 1998: 'Implicit Conceptions, the A Priori, and the Identity of Concepts'. *Philosophical Issues*, 9, pp. 121–48.
- Pollock, John and Joseph Cruz 1999: *Contemporary Theories of Knowledge*. Lanham MD: Rowman and Littlefield.
- Proudfoot, Diane 2006: 'Possible Worlds, Semantics, and Fiction'. *Journal of Philosophical Logic*, 35, pp. 9–40.
- Pryor, James 2005: 'There is Immediate Justification'. In Steup and Sosa 2005, pp. 181–202.
- Pust, Joel 2000: *Intuitions as Evidence*. New York: Garland Publishing.
- Putnam, Hilary 1975: 'The Meaning of Meaning'. In Gunderson 1975, pp. 131–93.
- Recanati, François 1993: *Direct Reference*. Oxford: Blackwell.
- Rescher, Nicholas (ed.) 1968: *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, 2. Oxford: Blackwell.
- Roese, Neal 1997: 'Counterfactual Thinking'. *Psychological Bulletin*, 121, pp. 133–48.
- Russell, Bertrand 1905: 'On Denoting'. *Mind*, 14, pp. 479–93.
- Sainsbury, Mark 2009: *Fiction and Fictionalism*. New Problems of Philosophy Series. London: Routledge.
- Schiffer, Stephen 1991: 'Ceteris Paribus Laws'. *Mind*, 100, pp. 1–17.
- Shope, Robert 1983: *The Analysis of Knowing: A Decade of Research*. Princeton: Princeton University Press.
- Sloman, Steven 2005: *Causal Models. How People Think About the World and its Alternatives*. Oxford: Oxford University Press.
- Sosa, David 2006: 'Skepticism about Intuitions'. *Philosophy*, 81, pp. 633–48.
- Sosa, Ernest 2007a: 'Intuitions: Their Nature and Epistemic Efficacy'. *Grazer Philosophische Studien*, 74, pp. 51–67.
- 2007b: *A Virtue Epistemology: Apt Belief and Reflective Knowledge*, volume 1. Oxford: Oxford University Press.
- 2009: 'A Defense of Intuitions of the Use of Intuitions in Philosophy'. In Bishop and Murphy 2009, pp. 101–12.
- Stalnaker, Robert 1968: 'A Theory of Conditionals'. In Rescher 1968, pp. 98–112.

- Stanley, Jason 2001: 'Hermeneutic Fictionalism'. *Midwest Studies in Philosophy*, 25, pp. 36–71.
- Stanovich, Keith and Richard West 2000: 'Individual Differences in Reasoning: Implications for the Rationality Debate?' *Behavioural and Brain Sciences*, 23, pp. 645–65.
- Steup, Matthias and Ernest Sosa (eds) 2005: *Contemporary Debates In Epistemology*. Oxford: Blackwell.
- Stich, Stephen 1998: 'Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity'. *Synthese*, 74, pp. 391–413.
- Strevens, Michael MS: 'Ceteris Paribus Hedges and the Role of Causal Hypotheses in Science'.
- Swain, Stacey, Joshua Alexander, and Jonathan Weinberg 2008: 'The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp'. *Philosophy and Phenomenological Research*, 76, pp. 138–55.
- Thomasson, Amie 2003: 'Speaking of Fictional Characters'. *Dialectica*, 57, pp. 205–23.
- Thomson, Judith Jarvis 1973: 'Killing, Letting Die and the Trolley Problem'. *The Monist*, 59, pp. 204–17.
- Unger, Peter 1983: 'The Causal Theory of Reference'. *Philosophical Studies*, 43, pp. 1–45.
- 2002: *Philosophical Relativity*. Oxford: Oxford University Press.
- Väyrynen, Pekka 2006: 'Moral Generalism: Enjoy in Moderation'. *Ethics*, 116, pp. 707–41.
- Walton, Kendall 1990: *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge, MA: Harvard University Press.
- Weatherson, Brian 2003: 'What Good Are Counterexamples?' *Philosophical Studies*, 115, pp. 1–31.
- Weinberg, Jonathan and Aaron Meskin 2006: 'Puzzling Over the Imagination: Philosophical Problems, Architectural Solutions'. In Nichols 2006, pp. 175–204.
- Weinberg, Jonathan, Shaun Nichols, and Stephen Stich 2001: 'Normativity and Epistemic Intuitions'. *Philosophical Topics*, 29, pp. 429–60.
- Williamson, Timothy 2003a: *Knowledge and Its Limits*. Oxford: Oxford University Press.
- 2003b: 'Blind Reasoning'. *Aristotelian Society Supplementary Volume*, 77, pp. 249–93.



- 2005: 'Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking'. *Proceedings of the Aristotelian Society*, 105, pp. 1–23.
- 2007a: 'Philosophical Knowledge and Knowledge of Counterfactuals'. *Grazer Philosophische Studien*, 74, pp. 89–123, Also published in Beyer and Burri 2007.
- 2007b: *The Philosophy of Philosophy*. Oxford: Blackwell.
- 2009: 'Replies to Ichikawa, Martin and Weinberg'. *Philosophical Studies*, 145, pp. 431–34.
- Woodward, James 2002: 'There is No Such Thing as a Ceteris Paribus Law'. *Erkenntnis*, 57, pp. 303–28.
- Wright, Crispin 2004: 'Intuition, Entitlement and the Epistemology of Logical Laws'. *Dialectica*, 58, pp. 155–75.
- Yablo, Stephen 2001: 'Go Figure: A Path Through Fictionalism'. *Midwest Studies in Philosophy*, 25, pp. 72–102.
- 2002: 'Coulda, Woulda, Shoulda'. In Gendler and Hawthorne 2002, pp. 441–92.