

Oberlin

## Digital Commons at Oberlin

---

Honors Papers

Student Work


---

2013

### The Content of Thought Experiments and Philosophical Context

Kevin G. Gilfether  
*Oberlin College*

Follow this and additional works at: <https://digitalcommons.oberlin.edu/honors>

 Part of the [Philosophy Commons](#)

---

#### Repository Citation

Gilfether, Kevin G., "The Content of Thought Experiments and Philosophical Context" (2013). *Honors Papers*. 323.

<https://digitalcommons.oberlin.edu/honors/323>

This Thesis is brought to you for free and open access by the Student Work at Digital Commons at Oberlin. It has been accepted for inclusion in Honors Papers by an authorized administrator of Digital Commons at Oberlin. For more information, please contact [megan.mitchell@oberlin.edu](mailto:megan.mitchell@oberlin.edu).

## **The Content of Thought Experiments and Philosophical Context**

Kevin G. Gilfether

Philosophy Honors Thesis

Oberlin College

Advisor: Dorit Ganson

April, 2013

Philosophers deal with many issues where there is empirical debate. For instance, many hypotheses in philosophy of mind, if true, would have consequences relevant to empirical psychology. Similarly, hypotheses in empirical psychology, if true, would likely impact work in philosophy of mind. But, philosophy and the so-called empirical sciences tend to have different methodologies. Philosophy is often viewed as a more *a priori* exercise than empirical sciences. The name of the most generally practiced academic branch of philosophy in the English-speaking world, "analytic philosophy," gives some hints to its supposed primary method of conceptual analysis. This "armchair method" is often articulated as holding that philosophy is somehow more reflective or conceptual than other disciplines. Empirical sciences, on the other hand, are typically construed as *a posteriori* ventures, where the evidential factors are experimentation and analysis of data gathered through examinations of items or situations in the observable world. Negotiating the relationship between philosophy and empirical sciences has been increasingly scrutinized of late because hypotheses in areas of philosophy such as philosophy of mind, epistemology, and areas of value theory deal with cognition. Philosophy seems to have consequences for theorizing in empirical sciences, and vice versa. Increasingly, intuition-driven philosophical treatments are polarized; whether they are a good or bad method

for philosophy seems to depend on your standpoint.

My goal here is to examine two takes on the thought experiment as it is used in philosophy. Timothy Williamson has the view that thought experiments are valid arguments about counterfactual possibilities, and that the intuitive judgements about them are intuitions about counterfactual conditionals.<sup>1</sup> Williamson makes many claims about foundational principles of traditional analytic philosophy. I will briefly examine these in the course of providing exposition of his views about thought experiments and intuitive judgements. Anna-Sara Malmgren has the view that thought experiments, generally understood, are not about, designed to show, or really based upon counterfactual conditionals at all.<sup>2</sup> The results of thought experiments are rather judgements of a sort of metaphysical possibility claim. I do not intend to spend a large time fleshing out Malmgren's positive view, but instead wish to focus on her critique of Williamson's stance. I seek to show that Malmgren can accept or deflect many or all of Williamson's revisionary claims about traditional analytic philosophy and still have a cogent criticism of his view. I will exhibit a few potential technical solutions which Williamson might adopt, the most successful I think coming from some suggestions that Tamar Szabó Gendler makes about intuitions and thought experiments. I will then close with a brief examination of ways this research can continue.

### **Caveats and Delimiting Remarks**

A few caveats are in order. First, Malmgren's criticisms in "Rationalism and the Content of Intuitive Judgements" are not generally focused upon Williamson's work in *Philosophy of Philosophy*. Her complaints are mostly focused on some earlier issues raised in a 2005 work of

---

1 Williamson, Timothy., pgs. 180, 194-195, 204-207.

2 Malmgren, Anna-Sara. pg. 307

Williamson's called "Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking."<sup>3</sup> However, Malmgren thinks, and I agree, that the body of her complaints here could be used against much of what Williamson says in *The Philosophy of Philosophy*, with relevant changes. A small part of my task here is to attempt to diagnose what specific changes will need to be made to the structure of Malmgren's argument. My larger goal is to show how Malmgren's critiques of Williamson's view on thought experiments, even if successful, leave a large portion of his revisionary work relatively unscathed.

Second, Malmgren and Williamson both have purposes and commitments beyond considerations here. Their views on issues further afield than just thought experiments inform their thinking on thought experiments. Deep and formalized inquiry of philosophical methodology of the sort that Malmgren and Williamson are doing is a somewhat new field, even if philosophers thinking about the nature of philosophy is not. The relationships between holding certain views in this discourse of philosophical methodology and having associated co-commitments are not very well defined yet. I have the view that as a literature builds up, and more commentary is done, better maps of the philosophical territory will emerge. Malmgren and Williamson's arguments deal with issues (much) larger than just what is at stake in the narrow aspects of their argumentation I consider here; both Malmgren and Williamson seem to have goals beyond just talking about the nature of thought experiments. Williamson has at least the basis for some positive proposal of revising current thinking about philosophical method. Malmgren has the views that some forms of currently orthodox philosophical practice are able to survive Williamson's criticisms with certain more modest revisions.

Third, there are certain critiques of intuitions and thought experiments that I will not be

---

3 Malmgren, Anna-Sara., pg. 270.

considering here. Generally speaking, I will not do a serious examination of skeptical critiques of philosophy because I do not have enough time to do them justice. For instance, experimental epistemology provides many critiques of philosophical methodology derived from the view that the norms governing certain areas of philosophical research are not just pure facts of the matter. Instead, they are, for some reason or another, very probably contingent upon cultural situation (like race, ethnic background, gender, sexuality, socioeconomic status, or other factors), or something else suitably outside of orthodox (philosophical) theorizing. Weinberg, Nichols, and Stich provide some statistical evidence for this hypothesis in "Normativity and Epistemic Intuitions." There, they analyze the results of numerous surveys where respondents of two or more different social groups (groups entirely made up of people from different races, groups with members of different socioeconomic status, and so on) are presented with specific example cases from epistemology which are thought to show justified true belief without an incidence of knowledge. It turned out that statistically significant numbers of non-white, non-high socioeconomic status respondents responded in some of these cases differently than white, high socioeconomic status respondents (a population to which the majority of academic philosophers probably belong).<sup>4</sup> Should philosophers' methods be viewed as exceptionally good at doing what they set out to do? It seems as though certain realizations deviant from mainstream philosophical orthodoxy might be somehow acculturated through exposure to certain academic experiences. Further, one might assert that those orthodox views are contingent; had the orthodoxies been different, the intuitive judgements might have been as well. I have only provided a very light gloss on a lot of theorizing that Weinberg, Nichols, and Stich do about the appropriateness, objectivity, or bias of epistemic norms, but I think the gist remains. While this critique is

---

4 Weinberg, Jonathan M., Shaun Nichols, and Stephen Stich.

extremely interesting, and I find it deeply troublesome for a few of the views I consider here, it is not the focus of this work. (For his part, Williamson responds to these sorts of criticisms by splitting the difference. He tries to make space for the norms philosophical training while saying that perspectives outside of academic philosophical orthodoxy might have interesting and cogent input. Williamson thinks that in careful examination of the subtle differences between different concepts and cases, a trained philosopher can come to good judgements and avoid the deep quicksand of contingency or relativism. He does say that philosophers without formal training might be able to come up with valuable iconoclastic points. If anything, though, Williamson views good philosophical training as a virtue, and not a vice.)<sup>5</sup> Again, I am focusing on what Williamson and Malmgren have to say about thought experiments.

Fourth and finally, studying philosophy by way of philosophy might present some problems. For instance, much of this current inquiry is about intuitions and intuitive judgements. But, I am using intuitive judgements as evidence, at least some of the time. In philosophy, many subfields of inquiry also use intuitive judgements or "seems that" statements as evidence, so there is at least tradition on my side here. (Malmgren thinks it better to say "intuitive judgements" than "intuitions," to make room for non-quasiperceptual judgements, and generally avoid controversy at the outset.<sup>6</sup> I will use "intuitions" and "intuitive judgements" roughly interchangeably, though I will pay special care to follow Malmgren when I am looking at her work.) However, there is a *prima facie* reason that intuitive inquiry might not help us in considering philosophical methodology: one of the things in our consideration here is what role intuition and intuitive judgements should play in philosophical methodology. Thus, we risk

---

5 Williamson, Timothy., pg. 191.

6 Malmgren, Anna-Sara., pgs. 267-268.

circularity in that our tools for diagnosing the situation are themselves the things we are assessing. (Williamson deals with an interesting parallel problem in the consideration of whether or not philosophy is a primarily linguistic or conceptual venture.<sup>7</sup> I will touch on this more in describing Williamson's anti-mentalistic picture of philosophical methodology.)

Do we risk circularity? One reason why we should not worry too much that we do (or should maybe worry a lot): a way of characterizing the cognitive sciences of psychology and neuroscience is that they study human activity and the actions of the nervous system, respectively. Also, there is a very common hypothesis (endorsed by many in the disciplines of psychology and neuroscience), that our thoughts are wholly the result of physical processes in the nervous system. Indeed, so a certain commonly accepted line goes, change the nervous system's physical state in a certain way, and you will change the thoughts of those creatures. Our thoughts about our brains, nervous systems, and human activity would presumably be included if the right changes were made. Thus, in some sense, our brains and nervous systems are implicated in what we think about them, even if it is not totally transparent to us how to characterize this. We are, to follow the theory, relying upon specific brainstates to tell us about brainstates generally. I hold that, similarly, we are worried here about the ability of specific intuitive judgements to tell us about intuitive judgements.

So, 1) if the problem of using brainstates to interrogate brainstates is not worrisome to you, and 2) the case of brainstates in cognitive sciences and the case of intuitions in philosophical methodology are appropriately analogous, then probably the problem of using intuitions to interrogate philosophical methodology should not terribly worry you. Conversely, if the case of brainstates is worrisome for you, or if you think that mediation concerns from

---

<sup>7</sup> Williamson, Timothy., pgs. 46-49.

brainstates and intuitions present different problems, then this story about the seemingly parallel problem between brainstates and intuitions has probably not allayed your fears. This example speaks to me, but it might not convince all.

At any rate, Williamson and Malmgren both proceed carefully, but without deep skepticism. The goal of their projects is not to convince deep skeptics about philosophy that skepticism is wrong, but rather to sharpen parts of philosophical methodology which have gone unquestioned. I think that it will be best to sidestep the navel gazing on these issues for now. That is not to say that the mediation concern is an unimportant one. It is just not one that I think is best handled here. I will proceed under the view that a few careful "seems that" claims or intuitive judgements could be employed here without the whole project unraveling. I will try as much as possible to make claims I make from my intuitions clear and unambiguous.

### **Williamson's Account of Philosophical Methodology**

I wish to briefly sketch Williamson's comments about the nature of analyticity, intuitions, the *a priori/a posteriori* distinction, and philosophical anti-mentalism, with the goal of getting his commitments on the table. Then, I will assess Malmgren's negative arguments against Williamson.

Timothy Williamson thinks that the foundations of analytic philosophy are under-considered. Much of what he is raging against is the idea that the intuitions that serve as the "inputs" for traditional philosophical inquiry are well explicated. One of Williamson's foils is a particular variety of intuition-driven rationalism: crude rationalists have the view that intuitions, even just of a specific sort, serve as evidence for the analytic, conceptual or linguistic truths which philosophy allegedly uses as inputs.<sup>8</sup> The rationalist, on Williamson's formulation,

---

<sup>8</sup> Williamson, Timothy. pgs. 2-3.



typically relies on the *a priori* character of these intuitions. At the same time, Williamson also has an opponent in the skeptical empiricist, who sees the whole exercise of *a priori* philosophical research as untenable. Williamson ultimately tries to split the difference by attempting to show that *a priori* and *a posteriori* in philosophy are not so clearly divided as typically considered. He does this in part by showing that the method of analyzing transparent concepts which underpins the allegedly false clean break between *a priori* and *a posteriori* is wrong. Williamson lays out several arguments against strong conceptions of analyticity in *The Philosophy of Philosophy* to show that the examination of allegedly transparent concepts as being the core of academic philosophy does not hold.<sup>9</sup> A further consequence of most of these arguments, if successful, is that the division between *a priori* and *a posteriori* truths is at least partially a specious one. The *a priori* nature of philosophical inquiry and the *a posteriori* method of the natural sciences is not so deep as often conceived.<sup>10</sup>

I wish to introduce the term "anti-mentalism" to characterize some of Williamson's leanings. Here, by "anti-mentalism," I just mean that important philosophical truths are not just truths about language, thoughts, or concepts (though they might be those things, they are not merely those things). Rather, they are truths about things in the world. Part of the idea here is that the intuitions that serve as "inputs" for philosophical inquiry have to be saying things about items in the world (at least some of the time), if philosophical inquiry does what we think it does. Williamson's arguments against conceptual transparency are also arguments for an anti-mentalist picture of philosophy, and Williamson's ultimate suggestions for fixing philosophical underpinnings are anti-mentalized suggestions.

---

<sup>9</sup> Williamson, Timothy. pgs. 7, 21.

<sup>10</sup> Williamson, Timothy. pg. 3.

As I have said, Williamson also opposes hardcore articulations of rationalism and empiricism. In complicating the boundary between *a priori* and *a posteriori*, Williamson is also trying to make sure that his new picture of philosophical methodology does not play into the errors of hardcore rationalism or empiricism. Williamson thinks the project of trying to complicate conceptual analysis and the *a priori/a posteriori* distinction can make room for some methods which hit the sweet spot between being too rationalist, or too much empiricist. I will make clear how he tries to make this transition in more detail as I go on.

To start, let us deal with Williamson's anti-mentalism: philosophical truths, though they sometimes might be conceptual (or analytic) in nature, are not always so. Williamson says, "Some philosophers of time argue that not only the present exists by appeal to special relativity. Philosophers of mind and language dispute whether there is a language of thought; whatever the answer, it is no conceptual truth."<sup>11</sup> Similar sorts of examples from moral, political, or aesthetic philosophy could also be constructed: just take a dispute about a philosophical issue where there are appeals to empirically discovered facts about humans, and you have an example. For instance, much of philosophy of mind deals not just with conceptual truths about the mind, but about hypotheses relating to actual minds in the world. Many philosophers of mind, such as Peter Carruthers, do large tracts of their research on the consequences of psychological or neurological hypotheses for philosophy, and considering the overall consistency of the explanatory entities posited by those entities. Carruthers is very clearly in the tradition of philosophy of mind, and thus, philosophy, but is also very clearly using non-philosophical (and seemingly not merely conceptual) facts. Thus, at the very least in a *prima facie* sense, there is a way in which non-conceptual facts or truths are important in philosophical discourse.

---

<sup>11</sup> Williamson, Timothy. pg. 49.

One might object by saying that our philosophical faculties are only really only tracking concepts, and that these non-conceptual facts are outside of proper philosophical inquiry. "Our intuitions in classic analytic philosophical inquiry are just what philosophers make judgements about: they are the thing that philosophers 'go on.' We cannot throw these sorts of non-conceptual facts in, because then we would not be able to properly use our philosophical faculties," this person could say. Then philosophy would have to be an armchair discipline, because it deals with conceptual cases and conceptual issues.<sup>12</sup> The person would just be making a claim to the effect that all of our important philosophical theses, even those which deal with empirically informed hypotheses are couched in language, concepts, or thought. Thus, philosophy would really be about (only) those bits of language, concepts, or thoughts. (There are more linguistically, conceptually, or thought-centered sorts of mentalism. I am not going to discriminate between different sorts here.) Williamson's response is that philosophical inquiry does not just deal with conceptual or linguistic issues, but with the content of those linguistic expressions, concepts, or thought. We are sometimes interested in asking, "what do we mean when we say 'x,' 'y,' or 'z' (or think  $x$ ,  $y$ , or  $z$ )?", but we are interested in not merely the consequences of what our language shows us or our concepts entail.

Just as practitioners of other disciplines must consider their methods and devices on the way to finding the answer to the questions they originally started with, philosophers sometimes do consider language and concepts.<sup>13</sup> But issues of linguistic or conceptual content are not all that philosophy is about, says Williamson. Similarly, scientists often have to consider the types of tools they use to make observations. When hypothesizing, the scientists are not just talking

---

12 Williamson, Timothy. pg. 48.

13 Williamson, Timothy. pg. 46.

about the tools which mediate their awareness of the things in the world. They are talking about things in the world!

This metaphor might obscure just what we are talking about, though. A linguistically or conceptually informed mentalist might respond to Williamson here by saying that there are some core philosophical commitments that are conceptual. Maybe all philosophy is supported by some conceptual knowledge? Williamson supposes that this alleged "purely conceptual core" of philosophy is close to the method advocated by traditional analytic philosophy of consideration and argumentation on the basis of analytic truths. Williamson in chapters three and four of *The Philosophy of Philosophy* moves to show that just because analytic truths, however conceived, are the core of philosophical methodology, it does not follow that philosophy is itself intrinsically a linguistic or conceptual discipline.<sup>14</sup> Williamson defends that such an anti-mentalist picture secures philosophy as being a partially empirical discipline. Williamson attempts to lay down some arguments against common conceptions of analytic truths, and by extension, against some of the methods of contemporary academic philosophy dependent upon those conceptions. Throughout, transparency about analyticity, intuitions, hardcore rationalism, and hardcore empiricism are all on the chopping block. In this section, I want to examine some of Williamson's arguments against analytic transparency, and what consequences these are supposed to have for Williamson's rationalist opponent. Williamson thinks that the best way to expose problems with a posited conceptual core of philosophy is by trying to show how strong articulations of analyticity are complex and problematic. In doing so, Williamson thinks that he can show that a conceptual core of philosophy is untenable, and too vulnerable to technical problems. I will try to provide a quick gloss on most of the problems Williamson raises to

---

14 Williamson, Timothy. pg. 21.

generally characterize the issues that he thinks confront a traditional mentalist account of philosophy, and thus point to an anti-mentalist picture of philosophy (and philosophical thought experiment) as most appropriate.

Williamson starts in chapter three by trying to show that metaphysical conceptions of analyticity are at a loss to show that philosophy is an essentially linguistic or conceptual venture. Metaphysical conceptions of analyticity are those that hold analytic truths are analytically true just in virtue of their meaning, as opposed to synthetic truths, which are true in virtue of their meaning and some facts about the world.<sup>15</sup> Williamson separates these metaphysical conceptions of analyticity from epistemological conceptions of analyticity. Epistemological conceptions of analytic truths treat analyticity as being a special category of knowledge or justification, and are typically articulated in a way that properly understanding the constituent words or concepts is enough for one to be justified in believing a certain allegedly analytic sentence is true.<sup>16</sup>

Why does Williamson think metaphysical conceptions of analyticity fail to justify philosophy as a solely conceptual or linguistic venture? Williamson considers a broad cross-section of strategies which are used to anchor different metaphysical conceptions of analyticity. I will go through each briefly in turn. The upshot of this section is just that Williamson thinks most strategies to establish a conceptual core of philosophy fail. I will not be trying to adjudicate whether or not Williamson succeeds or fails, but rather am just trying to provide a quick gloss of Williamson's process.

1. Analytic truths might be disambiguated from synthetic truths by disquotational rules about when claims are or are not true which are sensitive to a distinction between analytic

---

<sup>15</sup> Williamson, Timothy. pg. 52.

<sup>16</sup> Ibid.

or synthetic sorts of truths. Williamson believes that analytic and synthetic truths are not distinguishing different senses of "true."<sup>17</sup> Williamson considers disquotational principles for truth ("*P*' is true if and only if *P*"), finding that because the different senses of truth are not disambiguated by just this phrase, new forms must be considered. He then looks at consequences in truth table and conditional situations for statements like, "*P*' is [analytically or synthetically] true if and only if *P*." Williamson asserts that truth-conditional semantic approaches to disambiguating synthetic and analytic meanings of truth are likely to fail because they do not track analyticity or syntheticity of truth in the right way; analytic or synthetic conditional truths each might have either analytic or synthetic truths as antecedents. The compositional semantics of assessing the truth of different analytic and synthetic claims in a way which is sensitive to the analyticity or syntheticity of truths is technically daunting, and does not seem to follow clearly from how compositional truth-conditional semantics are done. Drawing rules general enough to be correct and specific enough to be useful does not seem likely, or so Williamson says.<sup>18</sup>

2. Analytic truths might be disambiguated from synthetic truths by focusing on the fact that analytic truths are so (solely) "in virtue of meaning," while synthetic truths are so "in virtue of meaning and some fact or facts about the world." One might then rely upon modal-analyticity to disambiguate whether truths are analytic or synthetic. A truth is modal analytic when it is a truth whose meaning is sufficient for truth. Consider for example this case: "Barristers are lawyers" is modal-analytic, while "Barbara is a

---

17 Williamson, Timothy. pg. 54.

18 Williamson, Timothy. pgs. 54-58.

barrister" is not. (In fact, any non-indexical sentence expressing a necessary truth is modal-analytic.) So maybe it is that the conceptual core of philosophy is made up of wholly modal-analytic claims. Williamson worries that the epistemology and context of such modal-analytic truths are still obscure. For instance, in the "Barristers are lawyers" example, we need an account of how we can know or justifiably believe that claim. Williamson thinks this is true even after an explanation relying upon us grasping the language and meanings of words.<sup>19</sup>

3. Frege-analyticity, where analytic truths are true if and only if they are synonymous with a logical truth, might be employed as a method of disambiguating analytic truths from synthetic. Two problems: first, the obscure epistemology from the modal analyticity account could still hold. Second, it seems *prima facie* apparent that not all philosophical truths are logical truths and that very important philosophical truths will be likely be excluded from the alleged purely conceptual core of philosophy.<sup>20</sup>

In short, Williamson thinks that if someone is to give an analytic conceptual core account of philosophy with the sorts of strategies which are typically used, then they must grapple with the epistemology of analytic truths.

Williamson is equally pessimistic about the ability of epistemological conceptions of analyticity to provide a compelling and substantive picture of a conceptual core of philosophy. As above, I do not mean to go through exhaustively the epistemic strategies that proponents of epistemological conceptions of analyticity have access to. I intend to give a brief sketch of the debate and put a gloss on how Williamson comes down. Williamson views these strategies at

---

<sup>19</sup> Williamson, Timothy, pgs. 58-63.

<sup>20</sup> Williamson, Timothy., pgs., 63-71.

some abstraction, generally characterizing them in terms "links" between different epistemic or psychological states (assent, belief, truth, knowledge, justification), and the sort of turn being taken in that link, linguistic or conceptual.<sup>21</sup>

Williamson goes through a very long series of refined versions of these linked conceptions to try to get to an epistemological conception of analyticity which provides room for a conceptual core of philosophy. A very large portion of this section is spent trying to see if there is a way to link the psychological stances of assent or belief to epistemic ones like knowledge or justification (after all, it is epistemic boundaries which seems to interest us in determining whether or not there is an analytic conceptual core in philosophy).<sup>22</sup> Williamson says that any attempts to make these linked conceptions the main strategy for pointing to a purely conceptual core of philosophy is not likely to be successful.

An understanding-assent link proponent takes the view that the epistemic state of understanding a particular claim leads to assenting to it. Further, they would say that understanding language is the appropriate level of analysis for epistemologically analytic claims if they take the linguistic turn, and concepts are the most important if they took the conceptual turn. Williamson's example for a linguistically based understanding-assent claim is, "Necessarily, whoever understands the sentence 'Every vixen is a female fox' assents to it."<sup>23</sup>

Williamson then says if anything is going to help the advocate for epistemological conceptions of analyticity pointing to a conceptual core of philosophy, it will be understanding-assent or understanding-disposition to assent links. These links fail, on his account. The gritty details are a bit too much for my purpose here, and Williamson employs

---

21 Williamson, Timothy., pg. 74.

22 Williamson, Timothy., pgs. 77-80.

23 Williamson, Timothy., pg. 73.



different strategies to defuse different articulations of understanding-assent links.<sup>24</sup> Williamson is clear that he thinks that most any strategy that depends upon understanding-assent links can be quickly defused with "a little ingenuity."<sup>25</sup> The key is just to understand the structure of each specific theory and find counterexamples as fit. Let us return to the "every vixen is a female fox" example. In the understanding-assent link, "Necessarily, whoever understands the sentence 'Every vixen is a female fox' assents to it," the word "every" might be a point for making a cogent objection to this articulation. For instance, one might object that a claim like "every F is G" implies that there is at least one F (such that F is G). So, in cases where there is no F, say, that foxes are all intricate hallucinations, there is no reason to accept that "Every vixen is a female fox," need be assented to.<sup>26</sup> (Though all foxes being intricate hallucinations might stretch this counterexample's plausibility, one could imagine a more plausible case from history. For instance, phlogiston was widely supposed to be something like what we mean when we say "elementary particle." So, one might think the sentence "All phlogiston motes are elementary particles," implying "There is at least one phlogiston particle," could be a good parallel example.) Other counterexamples might be constructed from a great many concerns about intermediate cases or potentially faulty logical presuppositions of the examples. In the interest of space, I will not rehearse any these, as they are very technical. The upshot is Williamson is skeptical of a project that constructs links which move from other epistemological states to psychological ones as a method of demonstrating an analytic conceptual core of philosophy.<sup>27</sup>

Does it follow from the fact that there is no conceptual core of philosophy that philosophy must "get up out of the armchair," and start to employ non *a prioristic* methods?

24 Williamson, Timothy., pgs. 85-120.

25 Williamson, Timothy., pg. 120.

26 Williamson, Timothy., pgs. 86-87.

27 Williamson, Timothy., pgs. 132-133.

Williamson makes the claim that there is a division between armchair methodology and *a priori* theorizing. He makes this on the grounds that the *a priori/a posteriori* distinction is more fraught than typically conceived in philosophical theorizing. Rather than a clear binary distinction, *a priori* or *a posteriori*, Williamson thinks we need more theoretical nuance. This nuanced view Williamson adopts follows from the view he holds that we have a generalized capacity to consider counterfactuals, which in turn gives us a capacity to consider metaphysical modality.<sup>28</sup>

"Metaphysical modality" here refers to our capacity to consider the contingency, necessity and impossibility of a great many metaphysical claims. (It might be true that our capacity to consider metaphysical modality does not wholly depend upon consideration of counterfactuals; Williamson's point is that the capacity to consider counterfactuals is responsible for human capacity to consider a wide range of claims of metaphysical modality. He further thinks that many modal claims we make are made from a general capacity to consider counterfactuals.) Our capacity to consider counterfactuals can in turn be seen from our ability to plan, use abductive reasoning, and make a wide variety of inferences about the way the world is from observed evidence. In further justifying that we have such a capacity to consider counterfactuals, Williamson claims that such a capacity would account for many cognitive abilities of typical humans, and is evolutionarily supported.<sup>29</sup> For instance, it seems that counterfactuals play an important role in causal thinking; one can think about situations by way of thinking what would be the case if the world had been different, such as how it would look if the bicycle were painted red instead of blue, or if you had the french fries instead of the onion rings.<sup>30</sup> It seems like the evolutionary pressures which produced humans probably involved

---

28 Williamson, Timothy., pgs. 136-137.

29 Williamson, Timothy., pgs. 135-136.

30 Tamar Szabó Gendler puts a nice gloss on the way that certain thought experiments about the natural world are quasi-perceptual in "Thought Experiments Rethought – and Repercieved." Roughly, her point there is that many

something like the ability to make plans about specific situations which might or might not have been the case.<sup>31</sup> Williamson further believes that this general capacity to consider counterfactuals allows humans general, fallible knowledge of metaphysical modality in many situations.<sup>32</sup>

So, supposing that we do have a general capacity to consider counterfactuals, and even further that this capacity gives us some fallible knowledge of metaphysical modality, why should we abandon an *a priori/a posteriori* distinction? An advocate for the traditional split might sensitive to the ways in which there are troubled cases might say, "There is already a pretty useful way of accounting for experience in *a priori* cases. In those cases which seem to be both *prima facie a priori*, and contingent upon experience, experience might be merely enabling, but not evidential." Williamson claims that how we come to grasp concepts and how we deploy them in some paradigm *a priori* cases is more complicated than just the traditional split between evidential and enabling roles that experience might play, though. Williamson attempts to capitalize on the seeming similarities between two cases where traditional orthodoxy might label the first *a priori* and the second *a posteriori*.<sup>33</sup> Here are the two claims:

(D1) If two marks had been nine inches apart, they would have been at least nineteen centimeters apart.

(D2) If two marks had been nine inches apart, they would have been further apart than the front and back legs of an ant.

Very generally, his point is that experience might be more implicated in the judgement of truth

---

thought experiments in the natural sciences involve imagistic reasoning. Such reasoning would almost certainly need to deeply involve our cognitive abilities to imagistically reason, and that reasoning would be epistemically crucial to the assessment of the thought experiment in question. A nice example case is that some sort of imagistic reasoning seems necessary for me to judge whether three elephants could fit in a certain room or not. Gendler, Tamar Szabó. "Thought Experiments Rethought – and Repercieved."

31 Williamson, Timothy., pgs. 136-137.

32 Williamson, Timothy., pgs. 155-165.

33 Williamson, Timothy., pgs. 165-168.

(for both claims) in a way that is more than "merely" enabling, but not yet being wholly evidential. Williamson holds that he could believe both (D1) and (D2) without doing any sorts of mental conversions. He even holds that one need not remember any time where one has seen an ant to hold (D2), one need just know what an ant looks like, and its size. If one follows Williamson in this thought, then one will probably accept his conclusion that while the *a priori/a posteriori* distinction is somewhat fraught. The distinction might not be wholly useless, but there are likely cases that strain its ability to fully explain them. If that much is true, then there is likely an area of armchair knowledge which is not what would traditionally be called *a priori* because experience might be more than merely enabling (but less than evidential).<sup>34</sup>

In short, Williamson's view on a new methodology for philosophy might be sketched as having these commitments:

1. A commitment to anti-mentalism about the subject matter of philosophical inquiry. Philosophy is not an especially linguistic or conceptual venture in his view, even if it often deals with these issues.
2. There is no conceptual core to philosophy which can be found from either metaphysical or epistemological conceptions of analyticity.
3. Humans have a general capacity to consider counterfactuals, and this capacity is responsible for many of our judgements of metaphysical modality.
4. The distinction between *a priori* and *a posteriori* knowledge is not comprehensive. There is likely a body armchair knowledge which is not strictly *a priori*.

I think Malmgren would probably view many tenets of Williamson's proposed account

---

<sup>34</sup> Williamson, Timothy., pg. 169.

for philosophical theorizing as still being open issues. Malmgren chastises Williamson for omitting certain sorts of rationalism in his earlier work, and I think it reasonable that those same sorts of criticisms might apply too for Williamson in *The Philosophy of Philosophy*.<sup>35</sup> Williamson might also be very wrong in his leap to saying that any conceptual core of philosophy will deal with only in analytic truths. Then his lengthy remarks on different conceptions of analyticity might be only an attractive nuisance. These projects might be Williamson's way of trying to ship certain precommitments he has (such as the unanalyzability of knowledge, or worries about philosophical problems like vagueness) into a picture of how philosophy should work.

Williamson's view that humans have a general capacity to consider counterfactuals is a point to which we will return later. For now, it will do to say that Malmgren has the view that we do not have the sort of general capacity to handle counterfactuals that Williamson supposes we do.<sup>36</sup>

Let us move to Williamson's picture of thought experiments. Thought experiments in philosophical orthodoxy are generally thought to be different from thought experiments in the empirical sciences.<sup>37</sup> Are they though? If we have one general capacity to consider counterfactuals, then it seems that scientific and philosophical thought experiments would be significantly similar. Williamson thinks that the consideration of the Gettier case as a paradigm thought experiment is worth examining to see if thought experiments more generally could be subsumed into thinking about counterfactuals. Part of this goal is to show that the thought experiment in philosophy is not an instrument that is "peculiarly philosophical," but an extension of or analogue to normal means of cognition. Here is the paradigm of the paradigm, one of Gettier's Gettier cases:

---

35 Malmgren., pgs. 318-319.

36 Malmgren, Anna-Sara., pgs. 307, 309-311.

37 Williamson, Timothy., pg. 179.

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

(d) Jones is the man who will get the job and Jones has ten coins in his pocket.

Smith's evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones's pocket ten minutes ago. Proposition (d) entails:

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e) and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket. Proposition (e) is then true, though proposition (d), from which Smith inferred (e), is false. In our example, then, all of the following are true: (i) (e) is true, (ii) Smith believes that (e) is true, and (iii) Smith is justified in believing that (e) is true. But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.<sup>38</sup>

The upshot of this example is supposed to be that Smith has justified true belief, without knowledge. How we should characterize this thought experiment is part of what is on the table in the debate between Williamson and Malmgren. Williamson claims that the Gettier argument is a "valid modal argument for a modal conclusion."<sup>39</sup> His interpretation of the "central premise" of the Gettier argument is:

(WFG) if a thinker were Gettier-related to a proposition, he/she would have justified true belief in it without knowledge.<sup>40</sup>

(WFG) is a counterfactual claim, which follows from Williamson's view that modal reasoning is subsumed by counterfactual reasoning. Williamson further thinks that thought experiments like the Gettier cases have irreducibly imaginatively warranted premises: that is, good thought experiments' premises are such that our imaginations are implicated in their acceptance.<sup>41</sup> I'll return to this point about imagination in my discussion of fallback maneuvers by considering

---

38 Gettier, Edmund., pgs. 192-193.

39 Williamson, Timothy., pg. 187.

40 Williamson, Timothy., pg. 195.

41 Williamson, Timothy., pg. 187.

some of what Gendler has to say about framing.

One might worry that (WFG) is not a good formulation of the realization we are supposed to have from a Gettier case. For instance, why suppose that Gettier is engaged in counterfactual reasoning? Gettier introduces his example with "Suppose that...", which could certainly be read as introducing a counterfactual situation. That said, it is a long way of from the paradigm philosophical practice of expressing counterfactual claims by way of the subjunctive mood ("If so-and-so were to thus-and-such, blah would result.").<sup>42</sup> Or one might worry that WFG is too strong: maybe it is only *possible* that a subject being Gettier-related would have an instance of justified true belief, but not knowledge. (This is where Malmgren will stake some ground.) Or maybe it is fantasy to suppose that an imaginary case can point to anything useful at all?

Williamson's response to these worries is first to draw attention to conversational context.<sup>43</sup> In short, thought experiments might be such that large features about them are left unsketched. That thought experiments take place at a level of generality to the arguments they are presenting (or are indicative of) is not necessarily a demerit. Consider the case of a childhood favorite: Opposite Day. On Opposite Day, all factual claims are negated, all imperatives issued must be for a doing opposite what you want done, and all questions must ask for the opposite of what one wants to know. Putting aside some worries about whether or not Opposite Day is self-defeating, childish, or has some boundaries that might be unclear, it can help illuminate just what Williamson means here. The Gettier case could be construed very plainly to not function to show what it purports on Opposite Day. But, in our articulation of the Gettier case as evidence, we are not supposing that it is Opposite Day, inside or outside of the example. Now strictly

---

42 Williamson, Timothy., pgs. 204-205

43 Williamson, Timothy., pg. 200.

speaking, this might serve as a point against the Gettier example, because the example does not rule out that it is or is not Opposite Day. But it seems as though the conversational context of the example does a fairly good job of delimiting that the consideration of Opposite Day should not be entertained.

Williamson's next response is to say that in the face of some alleged counterexample that one might tighten stipulations on the situation. That is, if one raises a case that is a counterexample, one might further sketch out the case to defuse it. For instance, if someone says that the Gettier case does not show what it seeks to because it might be Opposite Day, I can just further stipulate that it is not. If this further stipulation does not make the thought experiment untenable, then it might prove or support whatever claim the first articulation was meant to motivate. Williamson is very careful to mention that further stipulations should be considered as new thought experiments, though "not far from the old one."<sup>44</sup> He worries that fairly good thought experiments might quickly prove "too heavy" if they are loaded down in advance with stipulations meant to prevent any sort of refutation. In this way, thought experiments must walk the thin line between stipulating too much and not enough. Similarly, proposed counterexamples to the thought experiment must not be "too heavy," lest we depart the realms of sufficient generality and closeness to our general capacity to consider counterfactuals.

The upshot is that thought experiments, even if they must occasionally be revised, are not wholesale bad, illegitimate, or even peculiarly philosophical. Good thought experiments, like the Gettier case, are valid arguments about counterfactual possibilities. They are processed by our general capacity to consider counterfactuals, are most typically modal, and have limits placed on their effectiveness by humans' abilities to adequately do the counterfactual processing necessary

---

44 Williamson, Timothy., pg. 201



to assess them.<sup>45</sup>

### **Malmgren's Response:**

What does Malmgren have to say about all this? First, she is extremely skeptical about Williamson's proposed general capacity to consider counterfactuals. Malmgren identifies that Williamson's hypothesis is an empirical one.<sup>46</sup> Though Williamson is somewhat abstract about the level of implementation for this capacity, it either exists or does not. (Williamson is actually *less* explicit in *The Philosophy of Philosophy* about the details of the capacity to consider counterfactuals than the works to which Malmgren is responding. In *The Philosophy of Philosophy*, Williamson's main specific claim about the level of implementation of the general capacity to consider counterfactuals is that it evolved due to pressure to make plans about situations that might or might not be actual.<sup>47</sup> So, Williamson's view is still pretty vague.) Malmgren asserts that cognitive science is the best way to discover whether or not it exists. This consideration from cognitive science weighs against Williamson in two ways. First, whether we have a general capacity to consider counterfactuals is empirically unsettled. We clearly consider counterfactuals from time to time. But, whether or not one cognitive faculty assesses them or not is an open question, as is the question of thought experiments involving this capacity. Malmgren thinks that Williamson's abstractness might count against his fight with rationalists, as proponents of rationalism can find adaptations which avoid some theoretical trouble with his capacity.<sup>48</sup>

Malmgren thinks specific sorts of rationalism might make room for a special general capacity to consider counterfactuals which operates in some way close to Williamson's finessing of armchair methodology. She calls the proponent of this theory a "special capacity rationalist."

45 Williamson, Timothy., pgs. 206-207.

46 Malmgren, Anna-Sara., pg. 312.

47 Williamson, Timothy., pgs. 136-137.

48 Malmgren, Anna-Sara., pgs. 312-313.

What might this sort of rationalist look like? Malmgren is vague on the details, but I think that the basic picture would need to include a few points. Such a rationalist could have a special capacity to consider counterfactuals where the evidence involved is not wholly *a priori*. The origins of this capacity could be unlearned, developed, or contingent on experience, the details could be left open. It could also be that the special capacity of the rationalist is the ability to make modal judgements about necessity or possibility, or to imagine worlds in the way that they would need to be for the thought experiment's conclusion to follow, or something else like that. One example of a judgement that this special capacity rationalist could make might be a judgement about the nature of spatial experience, like "Necessarily, physical objects take up space." Such a claim might well require more than just *a priori* accessible information to epistemically ground it. (If one cared to, one could put a counterfactual spin on this claim, "Necessarily, if the world were to follow familiar spatial laws, then physical objects take up space." The antecedent of this conditional could make reference to the laws of physics, facts about the constituent components of physical objects, facts about space, etc. I just picked an antecedent I thought would be sufficiently general.) A modality claim like this one seems such that many rationalists would assert it, and a less abstract "special capacity" might account for why. Malmgren's posited "special capacity" might be a way of filling in how an advocate for philosophical rationalism could cope with Williamson's worry that rationalists about philosophy ignore empirical facts which have consequences for philosophical theorizing.

What is most important in that strategy is that there be some way that the special capacity rationalist could be a philosophical rationalist – they have some significant knowledge on the basis of intuition – while having a way for experience to count for something. Specifically,

experience is going to be necessary for seeing the ways in which the world does or does not comport with the intuitions of the rationalist. I think that the special capacity rationalist that Malmgren has in mind in this section is supposed to be one with some room to consider counterfactuals. This does not have to be Williamson's hypothesized general capacity: I read Malmgren as being more supportive of specific bounded counterfactual capacities with different cognitive processes involved in each, rather than one global capacity to generally consider counterfactuals. That said, Malmgren seems to want to stay out of the fray of positive arguments made by cognitive sciences, so this is very tentative.

If Malmgren is right in thinking that there might be rationalists who make room for a capacity to consider counterfactuals that is similar to Williamson's, then I think that Williamson has a few options. He might say that it does not matter if a rationalist could believe the same thing he does, because rationalism is less his foe than skeptics about thought experiments or philosophical methodology and hardcore advocates of a core conceptual realm of philosophy. Or he could say that such rationalists are not paradigmatic, even if they are possible. Or he could attempt to mount an argument about how such a capacity undermines what it is to be a rationalist. I am neutral on the specifics of Williamson's response. The idea of a special capacity rationalist is certainly conceivable to me. At any rate, I think the plausibility of Malmgren's posited "special capacity rationalist" makes it worthy of consideration. Similarly, I am not sure which path Williamson might take. Given Malmgren's complaint that Williamson tries to ship certain precommitments he has into his picture of philosophical methodology, it could be that he will try to dismiss it somehow from the outset.<sup>49</sup> Malmgren also thinks that Williamson is committed enough to excluding even a special capacity rationalism that he will be forced to

---

49 Malmgren, Anna-Sara., pgs. 317-318.

revise his view if he is to be charitable as an interlocutor.<sup>50</sup> Or, it might be that Malmgren's reading is uncharitable: Williamson does not just fail to include the special capacity rationalist as an option to prematurely rule out certain views opposed to his own.

Malmgren does not wholly rely on an argument that some variety of rationalists might be able to deal with the points of Williamson's critique. Malmgren also presents Williamson with a dilemma about his view on the blurriness of the *a priori/a posteriori* distinction. First, she introduces the point that a general capacity to handle counterfactuals must have the epistemic component refining why judgements made from such a capacity are justified and reliable.<sup>51</sup> Can we give an epistemology of these intuitive judgements from an alleged general capacity to handle counterfactuals? Can we give a description of what does or does not make them justified? Malmgren makes the point that counterfactuals all seem (*prima facie*) justified *a priori* or *a posteriori*. In a sense, Malmgren is just digging in against Williamson's claim that the *a priori/a posteriori* division is not so clean. Then the issue might well be intractable until some new evidence comes along. I do not think we need to worry about the intractability, though. Malmgren's claim about epistemology of counterfactuals might be read another way which is, I think, somewhat more charitable to Williamson. Malmgren might just be making the point that most philosophically interesting hypotheses which are gleaned by reflection on our intuitions are more *prima facie a priori* than they are *prima facie a posteriori*. The relevant point then becomes that there would need to be some way of sorting through the counterfactuals that seem more *a priori* and the ones that seem more *a posteriori*. So, when we make a judgement on the basis of our alleged general capacity to consider counterfactuals, we are picking the broad category

---

50 Malmgren, Anna-Sara., pg. 314.

51 Malmgren, Anna-Sara., pg. 314.

which best fits the sort of judgement we are making. Then, even if we are just picking the broad category which fits the case we are considering best, we would still need an epistemological account of how our judgement were justified.

Malmgren also presents a dilemma for Williamson's conception of the *a priori/a posteriori* split. Here is the dilemma:

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 [of the more or less *a prioristic* sort].<sup>52</sup>

How does this work, or not, for Williamson's case in *The Philosophy of Philosophy*? My view is that it does work, at least to an extent worthy of close consideration. Williamson is still interested in combating the rationalist. Recall Malmgren's point that specific sorts of rationalism could (at the very least plausibly) account for a general capacity to consider counterfactuals. (Moreover, it could be that there is a rationalist reason to doubt the *a priori/a posteriori* distinction.) Thus, there is a way in which Williamson might be sandwiched between his dispute with rationalism and his commitment towards non *a priori* armchair knowledge. Williamson might want to give up on the arguments against the rationalist. I do not think that he is interested in trying to let the rationalist about thought experiments get away with "business as usual" after so much argument about philosophy as not being an especially linguistic or conceptual venture. (Rationalism without mentalism is certainly possible, but the conceptual or linguistic analysis school of philosophy is both significantly rationalist and mentalist.)

The second horn, I think, is particularly harmful to Williamson's vision of philosophical

---

52 Malmgren, Anna-Sara., pg. 316.

methodology. To be clear, Williamson does say that even though there is a realm between paradigm *a priori* and *a posteriori* cases, an account of how we know which is which is not an easy thing to provide. Williamson's view is that we should accept the *a priori/a posteriori* split as too simplistic because our posited general capacity to consider counterfactuals, which in turn allegedly gives us knowledge of metaphysical modality, seems to suggest that the enabling/evidential split is too simplistic to account for all the sorts of counterfactual discriminations that we are capable of making.<sup>53</sup> This claim is not obvious though; it takes philosophical work to make it.

Williamson's view on the *a priori/a posteriori* split likely needs to be supported by significant of research in philosophy of mind, not to mention the empirical sciences of mind. Malmgren is really worried about the epistemic status of knowledge of metaphysical modality gathered from a general capacity to consider counterfactuals, and not just the cognitive basis. Williamson's account is vague on the cognitive details of the alleged general capacity to consider counterfactuals, but it is even vaguer on the epistemic details. I think Williamson's best bet is to tie the epistemic details to the cognitive details, where cognitive processes or states could be said to imply or signify epistemic processes or states. But, tying cognitive and epistemic to cognitive would require a lot of theorizing, much of which Williamson is opposed to in *Philosophy of Philosophy*.<sup>54</sup> Williamson is going beyond just what is empirically suggested, and though his theorizing is plausible, it is not yet confirmed. So, either we will wait for confirmation, much of it empirical, for Williamson's theorizing, or we will have to have a better epistemological account for the general capacity to consider counterfactuals which does not necessarily rely upon the

---

53 Williamson, Timothy., pg. 165-169.

54 Williamson, Timothy., pgs. 132-133.

cognitive processes. Either way, there is much of that Williamson abstractly theorizes must be filled in.

I will proceed by considering some thought experiments more generally, and examining some general theoretical features of thought experiments common to philosophy. I will pay special attention to ways in which I think that Williamson and Malmgren might analyze these thought experiments, but I am also interested in the more discursive features of thought experiment which might be revealed. Seeing how Williamson and Malmgren's models account for thought experiments other than just the Gettier paradigm is useful for this project; we are interested in how good their accounts really are for thought experiments in philosophical theorizing. But this analysis serves another purpose: rather than theorize about thought experiments in a contextual void, I think that solutions to some of Malmgren's objections can be seen by examining the different philosophical contexts of thought experiments. After this section, I will move on to a more formal explanation of how I think that Williamson can avoid some of the problems that Malmgren thinks affects his view.

### **Other Thought Experiments**

How well does Williamson's general and abstract picture of thought experiments work with some other paradigm examples of thought experiments? How might Malmgren's picture differ? I do not have space to analyze every example I would like to, but I think that some careful consideration of a few cases from ethics and metaphysics might augment the picture of thought experiments that is based upon consideration of the paradigm Gettier case. I will analyze here, John Locke's "locked in a party" thought experiment for assented to non-free will, Judith Jarvis Thomson's 1971 "violinist" case, and a joke by Michael F. Patton.

In doing my analysis, I will pay special attention to disentangling what might be called evidential cases made by the thought experiments from rhetorical ones. Evidential cases made by thought experiments are, roughly, the ways in which a thought experiment is supposed to give reason to believe or make an argument for certain views or propositions. The rhetorical cases made by thought experiments are those ways in which a thought experiment functions in the context of a work's larger argumentative structure, and roughly, what sorts of overtures that thought experiment makes towards positively affecting the overall plausibility, believability, or acceptability of the argument writ large, *for reasons other than evidential ones*. That said though, something should be said for the evidential role of a specific thought experiment playing a rhetorical role in the argument. You would have to be quite seriously committed to some form of radical error theory about our belief-forming process to say that evidence allegedly given by a specific argument or set of claims never serves as good reason to believe. Evidential goodness might have some very specific role in an example's rhetorical power. To draw a rough analogy, a doctor might have several tools available in trying to convince a patient to accept a certain surgery with particularly debilitating short-term outcomes, but long-term benefits. One tool of convincing the patient might be presenting medical evidence for why such a treatment is warranted or might improve a patient's condition. Let us call this practice "evidence giving." Still, the doctor has other tools than just evidence giving. The doctor might try to relate to the patient, to use good bedside manner, to talk to a patient in a way that makes the patient feel as though the doctor really thinks that the long-term benefits outweigh the short-term detriments. These reasons are the doctor's rhetorical tools. I think that philosophers and doctors might be somewhat alike insofar as they use evidence giving and rhetorical tools in concert.



Examining the relationship between evidence, rhetoric, and the theoretical fit of thought experiment to philosophy is why I have chosen to separate evidential and rhetorical reasons in this way of analyzing thought experiments. I think it is important to recognize that philosophers making arguments from thought experiments probably use rhetorical tools other than pure evidence giving, much the same way that doctors might. Thinking that philosophers might consider things other than pure evidential considerations in assessing theories cuts against the paradigm of philosophical inquiry as being particularly reflective and assessing only those things which count as "evidential." But I am committed to it for much the same reason as I am committed to the view that philosophical context matters: philosophers are humans, and humans do not exactly fit the paradigm of only considering those things which are evidential (and not considering the form or mode of those facts' presentation, whether they like the way they are being presented to, or any of the other multifarious factors which mediate which sorts of views they come to accept), even at their most reflective.

This division between evidential and rhetorical reasons anticipates a certain argument from Bernard Molyneux and Joshua Earlenbaugh which I will examine later. Their argument has it that intuitions or intuitive judgements are non-evidential, but are merely socio-rhetorically important. I do not mean to try to ship in Molyneux and Earlenbaugh's way of assessing thought experiments. I just think that a division between evidential and rhetorical reason giving recognizes a common distinction between the philosophical evidence given by an argument and the form of that argument, or what role it is supposed to serve in a larger work. The idea is that different thought experiments might operate differently given either their philosophical contexts, or their roles in argumentation.

I should probably say something to critics from the die-hard "no thought experiments in philosophy" camp. There is certainly room to criticize the view that thought experiments might give evidence; one might think that thought experiments make no claims useful in determining anything about the way the world is, for instance. To this, I have two things to say. First, I am not attempting to deal with skeptical critiques of philosophical methodology, and many arguments against any evidential value of thought experiments are skeptical arguments. For those arguments against evidential roles of thought experiments which are not skeptical about philosophical methodology, I think that the act of disambiguating the alleged evidential content of a thought experiment from the rhetorical content of that thought experiment (instead of just trying to brute force assert that the only rhetorical content of a thought experiment *is just* the evidence it gives or argument it makes) represents a change of pace.

John Locke takes up the difference between will and volition in Book II of *An Essay Concerning Human Understanding*. His hope is to make clear how we could intelligibly believe that a person could voluntarily do something when they could not have done otherwise. Put another way, Locke is concerned with showing that what is often referred to as "free will" is not necessary for an account of voluntary or assented to action. The example:

Again: suppose a man be carried, whilst fast asleep, into a room where is a person he longs to see and speak with; and be there locked fast in, beyond his power to get out: he awakes, and is glad to find himself in so desirable company, which he stays willingly in, i.e. prefers his stay to going away. I ask, is not this stay voluntary? I think nobody will doubt it: and yet, being locked fast in, it is evident he is not at liberty not to stay, he has not freedom to be gone. So that liberty is not an idea belonging to volition, or preferring; but to the person having the power of doing, or forbearing to do, according as the mind shall choose or direct. Our idea of liberty reaches as far as that power, and no farther. For wherever restraint comes to check that power, or compulsion takes away that indifferency of ability to act, or to forbear acting, there liberty, and our notion of it, presently ceases.<sup>55</sup>

What should we make of this? A good first question might be if this example is even a thought

---

55 Locke, John. Book II, Chapter 21, Section 10.

experiment in the relevant sense. Most thought experiments that Williamson and Malmgren are considering are really arguments of a certain sort (at the very least, they could be interpreted as arguments). The role that they play in the broader structure of a work sometimes is obscure, but giving a version of the thought experiment that is similar to the more abstracted and formalized argumentative structure of premises and conclusion is typically possible. In fact, many authors do this with their own thought experiments: a thought experiment serves as a way that a reader could "lock onto" the application of the abstract details of an argument, and formalized versions are presented either before or after a thought experiment. (That some people present formalized arguments first and others after a thought experiment might point to a way in which the evidential and rhetorical roles of thought experiments might be disambiguated and assessed.)

In a sense, I think Locke's example fits the thought experiment as generally conceived by Williamson and Malmgren. In sharing a vivid example, Locke is attempting to provide a positive argument for the plausibility and acceptability of the views that volition and will are 1) separable concepts and 2) practically distinct in that one could happen without another. This is also a counterexample to commonsense thinking that volition and will somehow determine each other, at least typically. Thus, Locke is trying to give an argument, from our intuitive views about a particular situation, to combat an interlocutor with an incorrect sense as to what our general intuitive views are. Of course, the Locke is a very old thought experiment, and philosophy has changed a lot since 1689 (or so it is commonly thought). It would be an interesting historical exercise to try to trace how thought experiments or examples have changed since Locke's time, but I will have to take a pass on it now.

Still other thought experiments are (it seems) intended to track our intuitions in a general

way. There, the thought experiment is used to attempt to show a more general intuitive view that we hold: our specific reaction to the thought experiment is supposed to convince us that we have a general view which we would apply in a great many similar situations. Let us consider another one of these thought experiments: Judith Jarvis Thomson's violinist example. Thomson's goal in the violinist case is to show us something about our intuitions about the moral permissibility of abortions. The example:

You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available medical records and found that you alone have the right blood type to help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you, "Look, we're sorry the Society of Music Lovers did this to you--we would never have permitted it if we had known. But still, they did it, and the violinist is now plugged into you. To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you." Is it morally incumbent on you to accede to this situation? No doubt it would be very nice of you if you did, a great kindness. But do you have to accede to it? What if it were not nine months, but nine years? Or longer still? What if the director of the hospital says. "Tough luck. I agree. But now you've got to stay in bed, with the violinist plugged into you, for the rest of your life. Because remember this. All persons have a right to life, and violinists are persons. Granted you have a right to decide what happens in and to your body, but a person's right to life outweighs your right to decide what happens in and to your body. So you cannot ever be unplugged from him." I imagine you would regard this as outrageous, which suggests that something really is wrong with that plausible [slippery slope from fetus to person] argument I mentioned a moment ago.<sup>56</sup>

Thomson thinks the example shows us something about our intuitions around the "right to life." She further thinks the violinist case points to a general intuition that even having a right to life does not entail that you have a right to receive what is required to sustain your life from someone else (especially when what is required to sustain your life represents substantial inconvenience to the person that would be asked to provide it). In the more narrow arc of her work, the point is to try to show us that our views about abortion either 1) could be aligned with that general intuition, or 2) that a specific way of framing abortion analogous to the violinist case

---

56 Thomson, Judith Jarvis. "A Defense of Abortion.," pgs. 48-49.

could lead a reader to make the judgement that abortion is a morally permissible, even given a fetus having a right to life.

This invites a somewhat technical question about just how much we want to align our norms with cases that map onto our intuitions. Following Williamson, it might be that our intuitions about a specific hypothetical case are evidential just insofar as they express or point to a counterfactual possibility claim. But whether or not a specific thought experiment, like the violinist case can even point to such an intuitive judgment about counterfactual possibility depends very significantly upon whether our general ability to assess counterfactuals can adequately process the specific thought experiment. Whether our general ability to consider counterfactuals can process such a case depends on many things, but closeness to the actual world is one really important criterion: we are almost certainly better at processing cases which are closer to the actual world than they are far away. Thus, for Williamson, an important feature of thinking about the violinist case would be whether or not the violinist case is "close enough" to reality for our general ability to consider counterfactuals to be properly effective.

I hold that the answer to whether or not our general ability to consider counterfactuals can adequately process a case significantly depends on the philosophical subject matter of the case. From roughly broader to less broad categorizations, I would say that the violinist example could plausibly be grouped into value theory, ethics, bioethics, personal ethics, or possibly just non-ethical decision making about one's personal life. All of these different domains of philosophical theorizing could very well have different sorts of evaluative or evidential standards associated with them. Additionally, if we follow Williamson in the view that our alleged general capacity to consider counterfactuals evolved not so that we could do philosophy, but rather so

that we would be able to make plans about all sorts of practical situations that might or might not be actual, then we might very well be better at making intuitive judgements about certain cases than others. So, some sorts of philosophical domains might have standards or problems which are more in line with our general ability to consider counterfactuals, and others less. In this way, the situation of the violinist case is little far-fetched compared to normal sorts of life situations. If the violinist case ever did occur in the real world, it would almost certainly involve some very perplexing thoughts. And we would likely think just that *for the reason that* our commonsense thinking is so "far" from these situations, and is probably informed by more relatively normal situations. A similar way of putting this nearness or farness condition is that as one goes further out from what is typically considered the bounds of normal everyday thinking and normal everyday experience that one's intuitions might be more likely to mislead or obscure what can justifiably be said about a situation. Whether or not the violinist is close enough to the actual world for our general ability to consider counterfactuals to adequately process it could be a matter of some debate.

Nearness of farness of the thought experiment to the actual world is not likely to be the only consideration which affects our ability to counterfactually assess thought experiments. (Williamson talks about how Gettier cases he's constructed in the actual world are not typically harder to assess than imagined ones with mundane forms.<sup>57</sup>) Also prominently involved will be the ability of the assessor to adequately think about those things about which the thought experiment is theorizing. I am going to say that exercising this ability is "tracking" the things about which we are supposed to have an intuitive judgements because of what the thought experiment reveals. (I have purposely chosen the vague label of "things" as what is tracked just

---

57 Williamson., Timothy., pgs. 192-193.

because I think that saying that we track concepts or something like that might be unacceptable to Williamson because of his anti-mentalism.) Tracking will matter in my later discussion of philosophical context, in that we might have different abilities to track in different contexts. I hold that for a thought experiment to be useful in philosophy, we must be able to track the things we are philosophizing about through the turns they take in considering a particular thought experiment. Put another way, unless we are able to track the philosophical entities the thought experiment is saying something about, we cannot call the thought experiment philosophically useful in the paradigm sense. For instance, in the violinist case, the entities posited to make the point about our moral intuitions are somewhat fantastic. Who know if our moral judgements track to such a fantastic case? They certainly might, but they also might not. So long as the world is not so fantastic that our ability to assess moral judgements is somehow different (and then ceases to track the things about which we are philosophically theorizing), then the thought experiment might work.

Let us imagine a pretty far out example: Aliens kill every person on earth tomorrow, save one old woman in Peoria, IL. One thing that we can judge on the basis of this thought experiment is that in such a case, the human race is probably going to go extinct. (If you like, I could bracket any debates about cloning or some such by stipulating that the woman has no special scientific knowledge or skills.) The reason is that in the "alien invasion" case, nothing about our normal understanding of biology is stretched by example: aliens invading is biologically possible, the human race ending if there can be no more procreation is biologically necessary. Nothing in the case suggests that our understanding of biology should be altered from commonsense understandings of biology, so we probably would not adopt an altered picture of biological

theory while assessing the example. Even though the "alien invasion" case is strange, we can probably come to a pretty safe defense for the claim that there would be no more human race soon after the imagined invasion. So, in the relevant sense, our understanding of biological theory tracks to the example.

I think that Williamson's general ability to consider counterfactuals account of thought experiments could only work on cases that are close enough to our world in the relevant ways that philosophically considered entities would be tracked in those thought experiments. Thomson argues that our normal body of everyday moral intuitions track to the violinist case (and then back again to the less far-fetched moral issue of abortion). Thomson further thinks that the violinist example is not so far-fetched that we should doubt any insight it gives us. These tracking and nearness or farness formulations also have a palatable theoretical conclusion for Williamson; thought experiments about hard to grasp metaphysical concepts might be such that our intuitions do not track what we are theorizing about well. When dealing with such a thought experiment, a hypothesized general capacity to consider counterfactuals might be more fallible than when processing thought experiments about simple spatial concerns like whether or not a dog could fit in this room.<sup>58</sup>

In Williamson's account of thought experiments, all of these factors will constrain or contribute to counterfactual judgement which we make. However thought experiment might be able to be processed without involving any sort of general ability to consider counterfactuals. Malmgren, who is skeptical of humans possessing a nontrivial ability to generally consider counterfactuals, would try to introduce a different way that we might consider the modality of this case. Thought experiments are not about providing an appropriately closed and delimited

---

<sup>58</sup> Williamson, Timothy., pg. 164.



valid counterfactual argument, they are making a metaphysical possibility claim by means other than the counterfactual.<sup>59</sup> (Recall that Malmgren rejects Williamson's counterfactual reading of thought experiments because of her doubts about our general ability to assess counterfactuals. She also worries that one might be confronted with the odd case where a counterfactual reading is false, but that we still want to hold on to our intuitive judgement. One case she points out is a Gettier case realized where the agent with supposed justified true belief but not knowledge has good reason to believe that their evidence is hallucinatory. Thus, in at least some cases, Malmgren thinks a counterfactual reading of a thought experiment might not be specific to our ability to make counterfactual judgements.)<sup>60</sup>

Malmgren thinks thought experiments generally point to just a metaphysical possibility claim, rather than considering ways that the world would be, given some specific counterfactual situation. I think Malmgren's weak general commitment to defending certain forms of rationalism might mean that she would be willing to defend that some sort of rationalist capacity to assess thought experiments. She has no comments in "Rationalism and the Content of Intuitive Judgements" which commit her one way or another on this though. Because she gives sparse details about the special capacity rationalist who might be able to avoid Williamson's arguments, it is hard to say for certain whether she thinks that it is just plausible that such a rationalist might exist, or if it is likely. As such, I do not have a strong sense of how to interpret Malmgren there.

The rhetorical role of the violinist case is a good example of how thought experiments can have somewhat vague operations. Whether or not a person accepts its conclusion or not depends very much on sort of valence that a person might have towards abortion already. A

---

59 Malmgren, Anna-Sara. pgs. 307.

60 Malmgren, Anna-Sara. pgs. 277-280.

reader who is introduced to the experiment by way of what it has to say about abortion, or in the abstract, or in a myriad set of different circumstances might have very different reactions. Because thought experiments require detail (and context, I argue later) beyond general or abstract points, an individual considering the thought experiment could have very different reactions to a very similar thought experiments depending on their phrasing, which imagined devices are employed, and how the reader is lead or directed to think about assessing the thought experiment. Then, it is not wholly in the realm of rational and pure logic that we are considering thought experiments, even if logic and a sort of rational consciousness are involved in their assessment somewhere along the line. This all follows very much the theorizing of Gendler's in "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium," more of which I will turn to later.

Our own inclinations and rational leanings are not all that determine the way we might read a thought experiment. Following Gendler's considerations of Hume, we are not just posing questions in a way neutral of consideration when presenting thought experiments, we are presenting valences which the person considering the thought experiment might follow or attempt to understand.<sup>61</sup> Different modes of presentation might well bring up different concerns for the assessor. To make this vivid, suppose that you were presented the violinist thought experiment not as though you were the innocent person nabbed and hooked up to the violinist, but as though you were instead the violinist. In fact, non-abstracted arguments for or against the permissibility of abortion often do just this. Advocates for the permissibility of abortions will typically stake out the issue as being an important consideration for the woman or women's health (health is often construed broadly to include mental or financial health or bodily

---

61 Gendler, Tamar Szabo., pg. 69.

autonomy). Similarly, advocates of restricted abortion access often frame the issue of being one where the considerations of the fetus as a person or potential person are put front and center. This is true most of all in the slogans that are often used to classify each side on this issue, "pro-choice" and "pro-life" respectively. These different considerations and introduced valences are so varied that the idea of a single correct interpretation to a thought experiment might be untenable.

An important final consideration is that some thought experiments seem either unanalyzable because of their complexity, because they introduce so many devices, or they might bring up so many varied valences for consideration. Certain thought experiments, while almost certainly thought experiments in some sense, are not useful as either rhetorical or evidential instruments. A joke can make this vivid:

Consider the following case:

On Twin Earth, a brain in a vat is at the wheel of a runaway trolley. There are only two options that the brain can take: the right side of the fork in the track or the left side of the fork. There is no way in sight of derailing or stopping the trolley and the brain is aware of this, for the brain *knows* trolleys. The brain is causally hooked up to the trolley such that the brain can determine the course which the trolley will take.

On the right side of the track there is a single railroad worker, Jones, who will definitely be killed if the brain steers the trolley to the right. If the railman on the right lives, he will go on to kill five men for the sake of killing them, but in doing so will inadvertently save the lives of thirty orphans (one of the five men he will kill is planning to destroy a bridge that the orphans' bus will be crossing later that night). One of the orphans that will be killed would have grown up to become a tyrant who would make good utilitarian men do bad things. Another of the orphans would grow up to become G.E.M. Anscombe, while a third would invent the pop-top can.

If the brain in the vat chooses the left side of the track, the trolley will definitely hit and kill a railman on the left side of the track, "Leftie" and will hit and destroy ten beating hearts on the track that could (and would) have been transplanted into ten patients in the local hospital that will die without donor hearts. These are the only hearts available, and the brain is aware of this, for the brain *knows* hearts. If the railman on the left side of the track lives, he too will kill five men, in fact the same five that the railman on the right would kill. However, "Leftie" will kill the five as an unintended consequence of saving ten men: he will inadvertently kill the five men rushing the ten hearts to the local hospital for transplantation. A further result of "Leftie's" act would be that the busload of orphans will be spared. Among the five men killed by "Leftie" are both the man responsible for putting the brain at the controls of the trolley, and the author of this example. If the ten hearts and "Leftie" are killed by the trolley, the ten prospective heart-transplant patients will die and their kidneys will be used to save the lives of twenty

kidney-transplant patients, one of whom will grow up to cure cancer, and one of whom will grow up to be Hitler. There are other kidneys and dialysis machines available, however the brain does not *know* kidneys, and this is not a factor.

Assume that the brain's choice, whatever it turns out to be, will serve as an example to other brains-in-vats and so the effects of his decision will be amplified. Also assume that if the brain chooses the right side of the fork, an unjust war free of war crimes will ensue, while if the brain chooses the left fork, a just war fraught with war crimes will result. Furthermore, there is an intermittently active Cartesian demon deceiving the brain in such a manner that the brain is never sure if it is being deceived.

QUESTION: What should the brain do?<sup>62</sup>

Though this case is clearly a ridiculous joke, it can sharpen some of what has been introduced heretofore. First, it brings into focus a critique of thought experiments as contrivances which might not tell us important truths about the world. Second, even if thought experiments can tell us something useful about the world, it does not follow that all or even most thought experiments are going to be able to do so. This thought experiment seems like as good a candidate as any for the label "uninformative." However, the thought experiment is clearly informing my discussion here, although not in the way that typical thought experiments are supposed to do. I will try to assess each of these points in turn.

Thought experiments being a contrived method is no great new critique. Indeed, one wonders if all the talk of trolleys, banks, brains-in-vats, alien invasions, and regicide is constructed in such a way that it does not really correspond to the world outside of philosophical theorizing. I think that there are skeptical and non-skeptical versions of this point. The non-skeptical versions would probably say something like the devices and metaphors which thought experiments in philosophy use to make their points are reductive. They do not give us an accurate map of the landscape, for the landscape is not so easily reduced to the scale of the map. This way of thinking about thought experiments comports with one sort of view about

---

62 Patton, Michael F., "Can Bad Men Make Good Brains Do Bad Things?"

philosophy that philosophy is either (depending on which way you like the metaphor) taking a very small everyday problem and building it up, or that it is taking a very large set of issues and trying to examine them atomistically, outside of their context. I think that both Williamson and Malmgren's positive proposals are attempts to deal with this non-skeptical criticism of thought experiments by making the classic notion of thought experiment more robust. The skeptical challenge is that philosophy is being corrupted by thought experiment. Thought experiments take very complex situations and attempt to categorize or classify them in a way that not well-suited to saying anything useful. The skeptic might say that thought experiments do not track the world, they track our (maybe very fallible and wrongheaded) thinking about the world. Williamson's anti-mentalism argument deals with this sort of view in some more depth. Malmgren does not deal with this sort of argument. I do not intend to deal with this skeptical critique either. The skeptical critique is still important, but I just do not have the space here.

Could Patton's thought experiment be uninformative? I think both Williamson and Malmgren would answer that it is. In fact, I think most philosophers would think that it, along with our intuitions about it, gives us no strong reason to think that we could answer the question of what the brain should do. So many considerations are brought up that one wonders how you would even tell what your intuitions are. (Maybe this is what Williamson had in mind when talking about overly stipulated and heavy thought experiments?) Malmgren might be a bit more sensitive to the philosophical value of the joke thought experiment as a thought experiment, but constructing a really robust reading of the joke is certainly a bit boggling. Maybe, upon really understanding the example, our intuitive judgements about it might render whether it tells us something about the metaphysical possibility of the judgement. This could be true in spite of the

example being difficult to assess.

From a different perspective though, the joke thought experiment might be really informative. For instance, it is useful now as a device to demonstrate that thought experiments might need a certain amount of unstipulated lightness. Part of its philosophical value might come from the ways in which it demonstrates that certain constraints are necessary for good thought experiments. Or it might have a philosophical moral of showing that the peculiar devices and affectation philosophers adopt to talk about thought experiments are really just a contrived code. (Patton certainly thinks that all the people who truly appreciate this example need to have had "at least five years in philosophy graduate school.")<sup>63</sup> Let me be clear that I do not think Patton intended this joke thought experiment to have all these different considerations associated with it. I just think that when this example is examined critically that it has some value beyond the comedic.

### **Avoiding the Dilemma:**

Williamson and Malmgren work with the same spirit in a significant sense. Both seem to have some worries about skeptical critiques of philosophy tossing the baby out with the bathwater. Similarly, I think the naturalized leaning of Williamson's picture of philosophy also hold some sway for Malmgren. None of Malmgren's theorizing presents decisive evidence for the wrongness of Williamson's picture of philosophical methodology. But overall, I think Malmgren's points spell out that, at the current state of debate, there are a few plausible challenges to some of the details of Williamson's picture of philosophy. What fallbacks or refinements might be made to Williamson's current argument? I make one suggestion to refine

---

63 Ibid.

parts of his proposed philosophical methodology with evidence from empirical psychology on imagined rational choice cases and another that philosophical context matters in Williamson's account. To this end, I suggest that some ideas forwarded by Gendler, Earlenbaugh, and Molyneux are compatible with much of what Williamson and Malmgren have laid out.

Let us start with one way that Williamson might have to go into making his general capacity to consider counterfactual conditionals account of thought experiments more compelling. Here, I wish to sharpen a bit of what Williamson has to say on conversational context and stipulation's role in thought experiments. Instead of just conversational context, I think we should also be interested in philosophical context of thought experiments. In fact, I suspect Williamson might have sensitivity to philosophical context already, although any leanings he has are not as explicit as I make them here.

What is philosophical context then? Different sorts or subfields of philosophy often have very different stipulated or assumed elements than others. Much of this practice comes just from the way that modern sub-fields of philosophy have taken to doing their work. Certain principles are taken for granted in some branches of philosophy which might not be taken for granted in others. For instance, in philosophy of mind, few ever entertain the hypothesis that there is no external world. This is not to say the hypothesis that there is no external world holds no bearing on the topic; surely, if it were true that there was no external world, then many of the entities posited in many common theories in philosophy of mind would not obtain. And the same is true for many other disciplines in philosophy: hypotheses common in one branch of inquiry are not broadly considered in another, even if they have some bearing on the theorizing done therein. But that is not itself seen as a reason to weigh against some hypothesis, quite often. The skeptics

might dig in, saying to a philosopher of mind, "Look, none of this theorizing about language of thought matters, because all this might be an illusory experience. After all, the world might not exist." But such a claim does not really persuade us very often that philosophy of mind is a void field.

I think there are two reasons why such skeptical concerns do not sway us. First, philosophers seem quite comfortable allowing subfields of philosophy to have specific foundational principles. (Certain areas of discourse in certain sub-fields might have yet more specific givens. For instance, theorists in a specific branch of epistemological research might have similar precommitments; theorists in a specific branch of metaphysics might have different precommitments.) These givens should not be inarguable: sometimes philosophers might bring hypotheses from certain sub-domains of philosophy into a realm where they are not normally considered with great success. Further, these givens are not likely immutable either. As discourses and sub-fields develop, certain givens might change, and become more or less "popular." For instance, consequentialist approaches occupied the position of primary foil in a large portion of ethical discourse; arguments were structured quite often with "the consequentialist" as the primary opponent or consequentialism as the view ultimately affirmed. Now though, Rawlsian theory, bioethical intuitionism and arguments over relativism or objectivity of ethics are ascendant (though certain versions of consequentialism are still quite healthy). In metaphysics, over a longer period, Cartesian dualism went from being widespread to being fairly uncommon. That it was so bespeaks a certain orthodoxy, which has shifted.

Second, the seeming unattainability of philosophical first principles might be a worry. For instance, in order to do epistemology, you might need to first have certain tools in logic. But to



have those tools in logic, you might need to have a theory in metaphysics accounting for them. Our acceptance of a certain theory of metaphysics might depend on value theory, and how well that metaphysical theory accords with our notions of the way things should be categorized. That value theory might just be reducible to evolved views. And so on, all the way down, or around, if our justifications are circular. Where you would stop this regression is a hard sort of question. An important upshot is that we cannot expect every philosopher to get explicit in proving, or even in establishing, every single assumption vital to whatever they seek to have as part of their philosophical projects. Just as the physicist is not generally expected to defend their assumption scientific realism or that one and one make two, philosophers of specific sub-fields are not generally called to accounts to perfectly defend all the specific aspects of their given principals. Keeping (non-skeptical) philosophical context in mind can cut down on the stipulations, which make awkward and complex thought experiments (like Patton's joke).

I think that Williamson might actually be very inclined to believe something like my philosophical context suggestion already. First, he creates room for an account of counterfactual thought experiments which is nuanced in such a way that different counterfactual possibilities are evaluated more or less reliably due to the power of our general capacity.<sup>64</sup> If indeed we have a general capacity to consider counterfactuals, which developed evolutionarily, then it follows that are capacity might have some serious shortcomings in certain areas, like certain more extreme metaphysics thought experiments. Roughly, things closer to the actual world are probably more likely to be easily computable by our general capacity to consider counterfactuals. The philosophical context of subfields of philosophy which are more alien to our ability to consider counterfactuals are probably more open to debate than those fields which are more in line with

---

64 Williamson, Timothy., pg. 164.

our ability to consider counterfactuals.

I think that Williamson's philosophical anti-mentalism also makes philosophical context limiting the bounds of thought experiments a likely argumentative strategy for him. If we are concerned not just with ideas, concepts, or terms in language, but the thing in world, does it not make sense that we are interested in thought experiments which tell us about the way things are in the world? Williamson's strategy of resisting a conceptual analysis view of thought experiments is a good example of his anti-mentalist views making him sympathetic to philosophical context.<sup>65</sup> There he spells out that it is not *just* epistemological accounts that we are interested in when we are considering an area of philosophy around a particular subject; we do not *just* want to know the epistemic properties of moral truths, say. The metaphysics of the subject of a particular subfield are important in studying that field. To put a contextual gloss on this, the metaphysics of specific philosophical subfields can help clarify what is or is not contextually relevant.

Last off, it is worth considering that there is yet debate about whether or not we have a general capacity to consider counterfactuals, if there is a conceptual core to philosophy, or even if there is anything wrong with philosophical methodology. Or even if those are appropriate questions around which to frame a debate, they might not be the only ways that a plausible account of intuitive judgements around the subject of thought experiments might be framed. Tamar Szabó Gendler, Joshua Earlenbaugh and Bernard Molyneux make some points about thought experiments that I think Williamson could incorporate into his specific account for how thought experiments are supposed to work. I intend to just very briefly describe their works, and then to offer an account of how Williamson might incorporate them.

---

<sup>65</sup> Williamson, Timothy., pg. 207.

Earlenbaugh and Molyneux's primary aim in "If Intuitions Must Be Evidential then Philosophy is in Big Trouble" is to avoid a methodology with an evidential account of intuitions/intuitive judgements altogether. Earlenbaugh and Molyneux share Williamson's skepticism of mentalism, and similarly want to have an account of philosophy where philosophy tells us about things which are "troublingly extra-mental."<sup>66</sup> Their account of intuitions and intuitive judgements is that they should not be seen as giving evidence (giving reason to believe), but rather as pointing to psychological inclinations to believe.<sup>67</sup> In this way, intuitions might come into play as important elements of the "socio-rhetorical" aspect of making arguments appropriately plausible.<sup>68</sup> Thus, in all sorts of domains, including those beyond philosophy, aligning an argument appropriately with common intuitions might make it an "easier pill to swallow" even if the intuitively held elements are wrongheaded. Though this account deals with intuitive judgements beyond just thought experiments, if one follows Williamson and Malmgren in thinking that thought experiments are supposed to have an intuitive character, then I think that this account is plausible.

Tamar Szabó Gendler considers different thought experiments and many articulations of similar thought experiments in "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium." The basic idea that Gendler forwards is that seemingly similar thought experiments presented differently might inspire different conclusions. That is differently articulated thought experiments, or different ways of framing similar thought experiments might draw on different processing mechanisms, especially as we fill out relevant details left unstipulated.<sup>69</sup> Much of what Gendler considers is empirical evidence for this hypothesis from

---

66 Molyneux, Bernard., pgs. 39-40.

67 Molyneux, Bernard., pg. 44

68 Molyneux, Bernard., pg. 38.

69 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pg. 69

psychological research where similar thought experiments with different forms (such as different presentations of Rawls' "Original Position" or Foot's "Trolley Case") seem to lead to different *prima facie* correct conclusions.<sup>70</sup> An idea or example presented one way might inspire different conclusions than one presented a different way.

I think that part of what makes this hypothesis so compelling is that there has already been significant psychological research on rational choice and framing effects. Most of that evidence seems to suggest that environmental factors and background features of agents significantly affect the abilities of those agents to make decisions and assessments. For instance, one psychological study looked at errors in algebraic reasoning among students presented with identical problems depicted with different symbols. (One group might get the traditional "x" and "y" representing variables, while another gets a somewhat less utilized " $b_1$ " and " $b_2$ " or similarly nonstandard symbols.) In this study, the error rates for the groups using non-standard notation were twenty-eight percent, compared to standard notation group's six percent.<sup>71</sup> Similar studies using syllogistic reasoning with unfamiliar and familiar forms bore out similar results.

This is not to say that familiarity with the form is the only factor in determining the abilities of typical persons to accurately solve different problems. Some syllogistic tests had logically supported conclusions deemed wrong or illogical by test respondents because they had independently plausible, but logically unsupported conclusions.<sup>72</sup>

Consider the following syllogism test, with these two premises:

1. If police dogs work hard, then all highly trained dogs are police dogs.
2. Police dogs work hard.

---

70 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pgs. 76-79, 84-85.

71 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pg. 70.

72 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pg. 71.

Respondents might then be asked to pick which of the following two choices are logically supported by that syllogism:

3. All highly trained dogs are police dogs.
4. Not all highly trained dogs are police dogs.

Though claim 3 is logically supported by the syllogism by way of simple *modus ponens*, it is far less likely to be selected by test respondents than claim 4. This is true even after they are reminded that the test assessors are only interested in knowing which premise the person thinks is supported by the rules of syllogistic logic. Presumably, this is because claim 4 is more independently plausible than claim 3 (which is clearly empirically false, consider therapy dogs, for instance).

The inverse of the above finding from empirical psychology seems true as well: logically supported, but independently implausible conclusions are less likely to be selected. Consider this brief syllogism test, again with two premises:

1. If some vitamin tablets make you ill, then some vitamin tablets are not nutritional.
2. Some vitamin tablets make you ill.

Respondents might again be asked to pick which of the following two choices are logically supported by that syllogism:

3. Some vitamin tablets are not nutritional.
4. Vitamin tablets are nutritional.

Again, claim 3 is logically supported by the syllogism by way *modus ponens*, but it is far less likely to be selected by test respondents. Again, even after test respondents are reminded that the test assessors are only interested in which premise is supported by rules of syllogistic logic,

logical errors are common. So, the independent implausibility of a logically supported conclusion seems to make it a more likely candidate for selection. One might think that philosophers are exceptional in that they would be better at such tests. Maybe such a view is right. Theorists like Gendler wonder, with good reason I think, if philosophers might not be so paradigm rational as we might hope, though. Even assuming that philosophers can develop the capacities to exceed on those tests, one might still worry that they do so in spite of normal human psychological functioning.

Many other psychological experiments in "rational choice" have brought out similar conclusions: it seems that there are situations where normal people might make logical errors or not behave in a paradigm rational fashion. This is true even when those people are reminded of common errors, mistakes, or false assumptions which people make. The origin or causes of these logical errors is a great question, but not the one I will focus on here. The upshot for our purpose is coming up with a "good" solution in a rational choice sort of situation would likely involve some thinking about the "form" of the situation, and the specific capabilities and skills we might have in dealing with that form or the question itself. The subject of the question, the way the question is framed, our capabilities for understanding the question and dealing with the subject matter it raises all impinge on our decision-making.

These pieces of evidence are useful for theorizing about thought experiments in philosophy because seemingly similar psychological or even cognitive processes are at play in consideration of philosophical thought experiments. Let us consider the Philippa Foot's "trolley problem," as presented by Judith Jarvis Thomson:

Some years ago, Philippa Foot drew attention to an extraordinarily interesting problem (Foot 1978). Suppose you are the driver of a trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track. The track goes through a bit

of valley at that point, and the sides are steep, so you must stop the trolley if you are to avoid running the five men down. You step on the brakes, but alas they don't work. Now you suddenly see a spur of track leading off to the right. You can turn the trolley onto it, and thus save the five men on the straight track ahead. Unfortunately, Mrs. Foot has arranged that there is one track workman on that spur of track. He can no more get off the track in time than the five can, so you will kill him if you turn the trolley onto him. Is it morally permissible for you to turn the trolley? Everyone to whom I have put this hypothetical case says, Yes, it is.<sup>73</sup>

Thomson goes on to assess all sorts of different articulations and framings of the trolley problem, trying to find underlying moral principles. One is the "fat man" scenario, where instead of throwing a switch to change the path of the trolley such that it would kill the one, you must push a fat man into the path of the trolley if you wish to save the five. Commonsense intuitive theorizing seems to suggest that throwing the lever to save the five is acceptable, while pushing the fat man into the path of the trolley is not.<sup>74</sup> There could very well be cogent moral reason for this. For instance, a significant moral difference between killing and letting die in more or less direct ways might be able to explain the difference in intuitive judgement. But what about changes to seemingly morally irrelevant factors? Would they change the moral theorizing of people drawing conclusions about the thought experiment?

It turns out that changes in morally irrelevant factors might change intuitions, and that psychological experiments have been conducted which suggests that they do. Here is Gendler on one such experiment:

Psychologist David Pizarro presented subjects with "fat man" trolley cases that differed only in the nature of the sacrifice involved: In the one case, a man named Chip Ellsworth III could be thrown off a bridge to stop a trolley hurtling toward 100 members of the Harlem Jazz Orchestra; in the other, a man named Tyrone Peyton could be thrown off to save 100 members of the New York Philharmonic. Subjects were significantly more likely to consider it morally acceptable to sacrifice Chip to save the Harlem Jazz Orchestra than to sacrifice Tyrone to save the New York Philharmonic (presumably an overcorrection of an initial instinctively racist response).<sup>75</sup>

---

73 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pgs. 76-77.

74 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," 77.

75 Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium.," pg. 78.

There are a few things to say about this. First, one might wonder if perceived races of those saved or killed really is not morally relevant. That statistically significant numbers of people report different intuitions might point to an argument for genuine moral differences between the cases of Chip being pushed and Tyrone being pushed. Whether or not it really is a morally irrelevant difference, Pizarro's evidence definitely points to ways in which the framing of thought experiments can impact conclusions that thinkers might have about them.

What can philosophers thinking about thought experiments do with this information about framing? One solution might be something like Thomson's exercise with the trolley case: re-articulate thought experiments in new forms, and try to find out which differences intuition does and does not track. One problem which this method would likely not address is the way in which differently framed thought experiments might lead the same person to unconsciously make different realizations, because the person is not considering the thought experiment adjacent to every possible way of framing it. There is probably no way that a philosopher will have the cognitive power to consider all possible relevant different articulations of the same thought experiments. Another problem is that the consideration of different versions of the same thought experiments in a similar context is a frame of a sort! It is definitely plausible to me that if the subjects of Pizarro's experimenting were confronted with both the "throwing Chip" and the "throwing Tyrone" cases that they might have similar theorizing about both of them. The danger would come from a rationalizing view: whichever realization that a person might have about one would likely affect their realization about the other case framed with no significant moral difference. In fact, Malmgren makes the good point that we seem rationally bound to have similar judgements in relevantly similar cases, because we would almost certainly be using the



same set of general rules.<sup>76</sup> After all, how likely does it seem that someone might say on the basis of moral differences, "Well, I am not sure about throwing this Tyrone, but I might throw Chip." The worry I have, though, is that whichever case is presented first in a specific engagement might be the one to spur the creation of the rule. And, if the way that rule is created is contingent on the frame of the example, the intuitive judgments of the assessor, and numerous social or biological inclinations of the assessor, then it seems that we might worry about the rule created not according properly with each example on its own merits. To be sloganizing, we might be less paradigmatically rational, and more rationalizing.

Similar cases might abound in other philosophical considerations where when two appropriately similar cases are considered together: a person might reach a conclusion about both of them that is different than the conclusion they might have had about just one presented in isolation. One solution might be to try to find more "neutral" translations of thought experiments which do not have the same sorts of problems which more bias-provoking frames might. However, then there is the large problem of determining what is "neutral" and the problem of flattening certain philosophical situations so that they lose their original force. None of the issues which Gendler raises are insurmountable, but they do ask philosophers to reconsider their self-image as highly trained concept analyzers or paradigm rational explorers of seemingly intractable issues. Though philosophers might try to minimize the effect of thought experiment framing by performing Thomson's exercise of thought experiment iteration, or attempting to find more neutral forms of thought experiments, framing almost certainly affects philosophical conclusions those philosophers reach. These worries have consequences for any attempts Williamson might make to adopt more contextual sorts of details into his general ability to

---

76 Malmgren, Anna-Sara., pgs. 291-292.

consider counterfactuals account because some frames and contexts seem as though they would be more or less easily assessed by our alleged general capacity to consider counterfactuals.

### **Conclusion**

While it is not my goal here to examine Gendler, Earlenbaugh, and Molyneux to the same level of scrutiny as Williamson, I think their views' plausibility and empirical backing presents an interesting material for Williamson. Earlenbaugh and Molyneux's point is, I think, directly counter to Williamson's, but Williamson could almost certainly adopt a modified version where some intuitions are non-evidential, or say that his version of a general capacity to consider counterfactuals presents intuitions which somehow track situations in the world. In that way, the spirit might be similar to something that Williamson could accept. To me, Williamson's view that imagination might irreducibly warrant certain premises seems like it is a form of this view. And this seems *prima facie* plausible for many cases; if we are interested in saying what a thing would be like if it were a different color, then it seems pretty clear that we must somewhere in out reasoning using intuitions or intuitive judgements. I think Williamson could accept Gendler's view with fewer modifications. Different thought experiments of a similar sort engendering different realizations is definitely in line with his views on the boundedness of a general capacity to consider counterfactual. Further, Gendler's empirical details for this hypothesis could help neutralize objections that Williamson's hypothesizing about the general capacity to consider counterfactuals is too abstract. Then Williamson might be able to avoid some of Malmgren's criticisms that his abstractness hinders his argument against the rationalist. At any rate, my view is that incorporation of some views from framing or philosophical context might help Williamson against some of Malmgren's criticisms.

I think there are three interesting ways that this research could be taken. One might be looking more explicitly at the structures of thought experiments, and their content. Williamson's "general capacity to consider counterfactuals" and Malmgren's "metaphysical possibility" claim proposals are both interesting, but many other proposals could also be forwarded, including some where intuitions might not be implicated at all. A sort of apples-to-apples comparison of the many different accounts of thought experiments might be able to highlight some ways in which different accounts of thought experiments could better explain general features of thought experiments. Another way to go might be more consideration of psychological or neuroscientific literature on human cognition. If we are naturalists about our ability to do philosophy, then we might well find interesting points in such research. A third option might be to pitch the conversation into more general methodological discussions other than thought experiments. I expressed anxiety earlier that good "first principles" could ever be found, but I might be wrong. Maybe the best way to tighten up philosophical methodology is more consideration of what all philosophers must have in common methodologically, or what essential norms philosophy must follow. I think such a project is untenable, as it invites too much circularity and navel-gazing. But it would certainly be nice to be wrong.

#### Works Cited

- Malmgren, Anna-Sara. "Rationalism and the Content of Intuitive Judgements," *Mind* 120. (2011): 263-327.
- Gendler, Tamar Szabó. "Thought Experiments Rethought – and Repercieved." *Philosophy of Science*. 71. (2004) 1152-1163.
- Gendler, Tamar Szabó. "Philosophical Thought Experiments, Intuitions, and Cognitive Equilibrium." *Midwest Studies in Philosophy*. 36. (2007) 68-89.

Kevin G. Gilfether – Final Honors Thesis – pg. 59 of 59

Gettier, Edmund. "Is Justified True Belief Knowledge?" In *Epistemology: An Anthology*, edited by Ernest Sosa, Jaegwon Kim, Jeremy Fantl, and Matthew McGrath., 192-193. Malden: Blackwell Publishing, 2008.

Locke, John. *An Essay Concerning Human Understanding*. 1689. Accessed February 16, 2013. <http://www.gutenberg.org/cache/epub/10615/pg10615.html>.

Molyneux, Bernard, Joshua Earlenbaugh. "If Intuitions Must Be Evidential then Philosophy is in Big Trouble." *Studia Philosopha Estonica*. 2. (2009) 35-53.

Patton, Michael F. "Can Bad Men Make Good Brains Do Bad Things?" Accessed on March 6, 2013. <http://www.mindspring.com/~mfpatton/Tissues.htm>.

Thomson, Judith Jarvis. "A Defense of Abortion." *Philosophy & Public Affairs*. 1. (1971) 47-66.

Weinberg, Jonathan M., Shaun Nichols and Stephen Stich. "Normativity and Epistemic Intuitions." In *Epistemology: An Anthology*, edited by Ernest Sosa, Jaegwon Kim, Jeremy Fantl, and Matthew McGrath., 625-646. Malden: Blackwell Publishing, 2008.

Williamson, Timothy. *The Philosophy of Philosophy*. Blackwell Publishing: Malden, MA. 2007.

I affirm that I have adhered to the Honor Code in this assignment.  
– Kevin G. Gilfether