
Experiments in Thought

Walter Hopp

Boston University

What are thought experiments, and how do they generate knowledge? More specifically, what sorts of intentional acts must one perform in order to carry out a thought experiment, what sorts of objects are such acts directed toward, and how are those objects made present in such acts? I argue on phenomenological grounds that the proper objects of thought experiments are, in certain cases, uninstantiated universals and relations among them. I will also argue that, in the best of cases, we intuit or “see” these universals and their relations to one another, and respond to some objections to this view.

Thought experiments in science, philosophy, and everyday life give rise to all of the following: belief, conviction, justification, perplexity, and, sometimes, knowledge. How? More specifically, what sorts of intentional acts must one perform in order to carry out a thought experiment, what sorts of objects are such acts directed toward, and how are those objects made present, or not, in the carrying out of those acts? My view is that a careful and initially metaphysically unbiased phenomenological description will support the view that the proper objects of thought experiments are, in certain cases, uninstantiated, often mind-independent universals and relations among them. I will also argue that, in the best of cases, we intuit or “see” these universals and their relations to one another, and respond to some objections to this view.

I.

Following Sorensen, I will understand an experiment to be “a procedure for answering or raising a question about the relationship between variables by varying one (or more) of them and tracking any response by the other or others” (Sorensen 1992, p. 186). Every *actual* experiment requires that the experimenters undergo at least some perceptual experiences that

are appropriately related to the variables whose relationship the experiment is designed to test. In the best case, the experimenter is able to perceive the variables themselves. Quite often, however, we must rest content with perceiving things that we know to stand in quite intimate informational relationships with the variables being tested. We cannot directly perceive the readiness potential, for instance, nor can we reliably perceptually discriminate among very short periods of time. But we can perceive electroencephalographs and clocks, and we have good reasons to believe that the states of those instruments co-vary in lawlike ways with the phenomena that elude our unaided senses. Our confidence in the obtaining of informational relationships between our instruments, on the one hand, and the target phenomena, on the other is typically itself grounded in other perceptual experiences.

It is not enough to perceive the variables or items suitably related to them to acquire empirical knowledge by means of an experiment. The relevant perceptual experiences must also be appropriately related to the contents of the experimenter's beliefs and judgments. I have a view on what this relationship consists in, which is really just a modification of Husserl's, and which I will simply lay out here. Suppose I merely think that the window in my office is clean. By itself, merely thinking this provides absolutely no evidence whatsoever for the proposition <my window is clean>. Suppose that I turn to the window and take a look at it. After a sufficiently but not obsessively close visual inspection, I judge that the window is clean. What was, before, a mere thought with no epistemic worth has given way to knowledge. This is fulfillment. In such an act, I "experience how *the same* objective item which was 'merely thought of' in symbol is now presented in intuition, and that it is intuited as being precisely the determinate so-and-so that it was at first merely thought or meant to be."¹

Acts of fulfillment are quite often cases of knowing. In them, an act of perceiving and an act of thinking—a judgment, in this case—converge upon the same object. More generally, in every act of fulfillment (a) the object is intuited, (b) the very same object is thought about, and (c) there is an appropriate synthesis between the acts of intuiting and thinking. The object in this example—that which the act is about—is not just the window. Rather, what is presented in perception and then represented in the judgment is a state of affairs—the being-clean of the win-

1. Husserl 1970, p. 694. Also see Husserl 1970, p. 720: "What the intention means, but presents only in more or less inauthentic and inadequate manner, the fulfillment—the act attaching itself to an intention, and offering it 'fullness' in the synthesis of fulfillment—*sets directly before us*, or at least more directly than the intention does. In fulfillment our experience is represented by the words: '*This is the thing itself*.'"

dow. Finally, these acts do not merely occur side-by-side. They are related in such a way that I am aware of the window's being the way it is represented as being in the judgment.

Although the case of fulfillment above resulted in me knowing that my window is clean, not all acts of fulfillment confer justification on empirical beliefs. Had I imagined my window as clean, my thought that it is clean would have been intuitively fulfilled, but imagining would not have conferred any justification on the judgment that the window is clean. This would be a case of merely *illustrative* fulfillment. The reason is that while imaginative acts present, rather than merely represent, their objects, they do not present actual objects and states of affairs in the flesh. Perception is, while imagination is not, an "originarily presentative" act (Husserl 1982, p. 327) with respect to actuality; it "is that mode of consciousness that sees and has its object itself in the flesh" (Husserl 2001, p. 140). Perception is a *self-giving* act with respect to actual individuals and states of affairs.

Fulfillment is, in every case, completely unlike inferential justification. My belief that P cannot be inferentially justified solely on the basis of another belief that is about exactly the same state of affairs that P itself is about. Rather, inferential justification involves coming to believe the world is one way because of other ways one believes it to be; reasoning is a matter of determining what must, might, or cannot be the case given that something else is the case. In fulfillment, however, the fulfilled act and the fulfilling act in virtue of which it is justified must be about the same state of affairs; we believe the world is a certain way because *it is that way*. Epistemic fulfillment is that unique intentional act in which we confirm a truth-bearer by consulting its truth-maker.

This does not mean that experimental results are cases of non-inferential knowledge. Most experiments, as is well known, depend upon a wide range of background assumptions in order for the theoretical importance of their results to be interpreted, and often also depend on beliefs about the absence of potential defeaters—and the more complex and theory-laden the experiment, the more vast the array of possible defeaters. James McAllister has argued persuasively that the evidential value of an experiment is not an intrinsic property of it. This should not force us into his position, however, according to which "the evidential significance of experiment is conferred on it in particular areas of science at particular times by the persuasive effort of scientists" (McAllister 1996, p. 237). Rather, we can and should insist that all experiments do, at some level or other, involve the noninferential verification of at least some propositions. We may not be able to determine anything noninferentially about free will by observing EEG readings while a subject reports the time he or she

felt a conscious decision to press a button, since the interpretation of the results depends on beliefs about such things as the relationships between EEG readings and the readiness potential, the readiness potential and neural activity, neural activity and conscious intentions, conscious intentions and the self-conscious noticing of conscious intentions, and so on. Nevertheless, we can acquire noninferential knowledge of such propositions as “The subject pressed the button” and “The EEG reads 5 microvolts.” Knowledge of this sort does not evidentially depend on the persuasive efforts of scientists, but is presupposed by it. If scientists themselves did not encounter such evidence in experimentation, it would be difficult to see where they could summon the rational motivation for engaging in any persuasive efforts or experimentation at all.

Finally, because fulfillment is a more complex act than mere perception (or merely thinking), it is susceptible to a wider array of failures. One obvious type of failure stems from nonveridical perceptual experiences. For instance, one’s thought that there is a car coming down the street might seem to be fulfilled by a hallucination of a car coming down the street. But there are other ways fulfillment can fail. In particular, the object one is thinking about might not actually be identical with the object one is perceiving. Twin Earth cases exhibit this possibility. If transported to Twin Earth, your singular thoughts about earthly individuals would seem to be fulfilled by your experiences, but they would not be. Your perception of twin-Kripke might be perfectly veridical, yet your thought “there’s Kripke” would not be genuinely fulfilled. Other familiar sorts of errors can occur. One’s thought that a student is cheating might seem to be fulfilled on the basis of his sidelong glances, when, after all, one has only observed sidelong glances. One might hear the note C veridically, but misidentify it as a G. And so on.

With respect to actual individuals and states of affairs, only perception and introspection are self-giving and are, accordingly, the only ones that can serve as constituents in acts of epistemic fulfillment directed upon such objects.² Actual experiments essentially involve *perceptual* or *introspective* experiences, and those experiences must figure as constituents of acts of epistemic fulfillment. As we will see, however, perception and introspection are not the only kinds of self-giving acts. Rational intuition and even imagination and hallucination can be self-giving with respect to

2. Or, more precisely, in acts of primary epistemic fulfillment, which alone concern us here. Intuitive memories can also fulfill empirical beliefs, as can certain acts of image-consciousness (looking at pictures, videos, and maps), but they are each, in differing ways, dependent upon perception. They provide derivative epistemic fulfillment. For more on these distinctions, see Hopp 2011, §7.2.

properties or essences and the relations among them, and that is what sometimes occurs in the case of thought experiments.

II.

What, then, about thought experiments? Is a thought experiment just like an actual experiment minus the perceptual experiences and acts of fulfillment that are essential to the latter? Is a thought experiment a case in which those experiences of fulfillment are replaced by mere thoughts, devoid of fulfillment? I am not sure anyone holds this view, but Sorensen's claim that thought experiments are a type of unexecuted experiment might suggest something along those lines (Sorensen 1992, pp. 213–4). One obvious problem with this claim is that thought experiments themselves can be executed or unexecuted. The view is also suggested as a real possibility by the leading question that frequently figures in this debate: how can merely thinking about something provide us with knowledge about it? The answer, I suggest, is that it cannot, at least if by “merely thinking” we intend to contrast that with “having an experience of epistemic fulfillment.” A thought experiment is neither a mere thought about an actual experiment, nor is it a mere thought about the entities observed and manipulated in the course of an actual experiment. Merely thinking about carrying out an experiment in which one measures the timing of the readiness potential and subjects' reports of becoming conscious of making a decision provides no evidence whatsoever about the results of that experiment. More generally, merely thinking about actual states of affairs provides no reason to believe that they obtain. But thought experiments do, or at least can, provide evidence for propositions. So either thought experiments provide some other sort of access, distinct from mere thinking and from fulfillment via perception (on pain of being actual experiments), to actual states of affairs, or they do not have actual states of affairs as their principal objects.

What might thought experiments be if the first option is correct—if, that is, they are directed towards actual states of affairs by means other than mere thinking or fulfillment by means of perception? The only answer I can think of is that they are really, despite appearances, *arguments*. This view has been endorsed by John Norton, and is susceptible to some rather serious objections. Different conclusions can be drawn from the same thought experiment (Bishop 1999). Thought experiments are not formally valid or invalid (Häggqvist 2006). And many thought experiments look an awful lot like actual experiments rather than arguments (Sorensen 1992).

I suggest we provisionally take this similarity at face value and see where it leads us. I further suggest that we take the experiences involved

in thought experiments at face value. What I mean by that is that we provisionally set aside whatever metaphysical or empirical presuppositions, however established we might take them to be, and take a frank look at the experiences involved in performing thought experiments. That is, we should exercise the phenomenological *epoché*, which, despite Husserl's long-winded and often foggy descriptions of it, is no great mystery. We examine our experiences and determine what they are aiming at and what they've acquired (Husserl 1969, p. 177) or seem to have acquired, and *in that order*. What we do *not* do is determine in advance what they *must* or *cannot* acquire, and then deduce what they must have been aiming at.

An example might illustrate the difference. We can imagine a philosopher convincing himself that it would be irrational to attempt to do something unless one knew for sure that one was going to succeed. Suppose such a philosopher observes me shooting free throws and asks me what I am trying to do. "I'm trying to make free throws," I answer. "But I see you miss quite a few of them. And yet you're rational, right? So you must not be trying to make free throws. You must really be trying to do something else—maybe release the ball from your hands." This philosopher has not consulted the experiences of trying to do something to determine what they are aiming at and then determine what they have acquired. Rather, he has determined in advance what such acts must acquire, and then inferred what they must be aiming at. The results speak for themselves.

It is this latter, non-phenomenological approach that generates such counterintuitive theories as the sense-datum theory: since perception and sensation simply *must* get things right, they *cannot* be directed towards physical objects and states of affairs. The corresponding danger in this arena should be obvious, and it consists in reasoning as follows: since platonic objects do not exist, and since we simply could not have a faculty for accessing them if they did, thought experiments must not be aiming at those. Naturally these are objections one must deal with, and if it turns out that we really do aim at physical objects in perception, universals in thought experiments, or objective moral values in our moral discourse, and that such aimings simply could not hit their mark, perhaps an error theory of those domains is called for. But such philosophical views should not be allowed to distort the content and character of our intentional experiences.

III.

My view is that thought experiments, at their best, are in fact founded on acts of fulfillment in which we intuit universals and the relations among them, and that the actual instantiation of those universals and relations is immaterial. I begin with a thought experiment intended to show that this

is indeed possible. Suppose that I am faced with the task of moving a rather bulky couch out of a cramped apartment and through a narrow doorway. I know it won't be easy, but I also know it can be done. After all, I got it in here. As I survey the scene before me, something remarkable happens: the couch begins to float through the air of its own accord, makes a surprisingly simple sequence of twists and turns, and exits through the doorway and down the stairs. In disbelief, I rub my eyes, and upon opening them again find the magical couch sitting in its original position, stubbornly unmoved.

It's easy to say what went wrong here. I underwent a hallucination. My experience presented the world to be a certain way, and it was not that way. But something went right, too: I now know, or at least have a justified belief, that the couch will fit through the doorway, and I know just how to make it fit. How is that possible? What was I right about?

The knowledge I acquired on the basis of this hallucination concerns the spatial properties of the couch and the doorway. I know that something with the shape and size of my couch can fit through something with the shape and size of my doorway if maneuvered this way and that. These spatial properties and relations, moreover, were genuinely presented in the experience. I did not merely think about those properties and relations—a procedure which, before the hallucination, did not get me anywhere. I *beheld* them.

What more can we say about what I beheld and learned? One thing we cannot say is that I veridically saw a physical individual with the same properties as my couch moving through a physical space with the same properties as my doorway. Perhaps, then, I beheld different kinds of individuals, individuals that are non-physical or at least not “external”—a mental picture or an idea, say, or maybe a sense datum. I am fairly confident that any answer along these lines is incorrect. First, hallucinations are errors. But if my experience was veridically of actual particulars that genuinely possessed the properties that I beheld, then my experience was not erroneous after all. I cannot be wrong about a couch or the state of things in my apartment if my acts are not even about them.

Second, the experience could not possibly have presented such individuals veridically anyway, since my experience presented the relevant spatial properties as instantiated by middle-sized physical objects out there in physical space, not as instantiated by little mental pictures in a little mental space or by sense data in a phenomenal space. Perhaps sense data can be located in physical space. Even so, my perceptual experience was directed at something with properties that no sense datum could have. For instance, it looked as though it could be sat on. Most importantly, it appeared, as all physical objects do, as though it had more to it than what

was presented in my experience—it looked like it had hidden sides that could come into view, and presented sides that could become hidden from view. As it rotated in the air, it looked like the *same thing* that was changing its orientation and position and whose parts were coming into or receding from view. But sense data do not have more to them than what is manifest, and, in stark contrast to the experience of physical objects, any change in one's experience of a sense datum is a change in the sense datum one experiences. If we insist that some actual individuals must bear the properties that are genuinely presented in hallucinatory experiences, then we must insist that there is some actual individual that has the same size, shape, and spatial location as my couch, something that can only be partially perceived at any one time, which can be presented as identical across experiences that are not identical, and which can be perceived simultaneously by many perceivers. But what besides something physical and “external” could possibly instantiate those properties?

To recap, then: first, in this hallucination I am genuinely *presented* with spatial properties and relations. Secondly, there are no individuals that could plausibly be thought to bear the spatial properties and relations that are presented to me. When I hallucinated, nothing seven feet long came into existence. No light was absorbed or reflected by a new gray thing. No doorway, whether physical or mental, was momentarily rendered impassable by a seven-foot long gray obstruction. My experience undoubtedly involved a number of mental individuals, perhaps even mental models or images. But the object of my hallucination was not a mental model or image, but what the model or image (supposing one exists) is a model or image *of*.

The third point is that I learned something on the basis of this presentation. My knowledge was not of any particular couch, since there was no individual veridically presented to me in that act. And yet I know something about my couch. The reason is that my experience provided me with the knowledge of any possible individual bearing the properties that were presented to me, and my couch is one of those individuals. Though nonveridical with respect to any individual, my hallucination was veridical with respect to *something*.³

The acquisition of knowledge on the basis of this hallucination was not a thought experiment, any more than perceptual experiences and fulfill-

3. Compare this with Mark Johnston's view: “When we see we are aware of instantiations of sensible profiles. When we hallucinate we are aware merely of the structured qualitative parts of such sensible profiles” (Johnston 2004, p. 137). My view, which is inspired by this one, differs from it insofar as uninstantiated sensible qualities and relations are among the objects we are veridically aware of in hallucination. But hallucination is also nonveridical insofar as it presents those qualities and relations as instantiated.

ments are actual experiments. But the difference between it and a thought experiment is just this: in an experiment, one purposefully performs a procedure in order to answer or raise a question by varying and tracking the relations among variables. Experiences of the sort I had are among those that enable one to track those relations. Suppose that I could hallucinate like this at will. Then, when struck with the question of how to get the couch through the door, I could induce an experience like this on demand. If the couch doesn't get through one way, I repeat the hallucination. I thereby deliberately vary one variable, the orientation of the couch, while keeping others, such as its shape and size and the shape and size of the doorway, constant, and track their relations. Now I will have acquired knowledge on the basis of a thought experiment. The nonveridicality of the act with respect to actual individuals and states of affairs is immaterial, provided it is veridical with respect to the properties presented in the hallucinatory experiences.

If this sounds suspiciously similar to simply saying that I have conducted an experiment by means of my imagination, that's because it is. At this point, now that I am savvy to the hallucination, the chief difference between it and a vivid imaginative act is that the hallucination cannot exist side-by-side with a set of veridical experiences. A hallucination of a couch in a doorway cannot coexist harmoniously with a veridical experience of that same couch resting in the room, for instance. Imaginative acts, however, can: I can imagine anything whatsoever, and what is thereby presented to me does not compete with what is presented in my veridical experiences. This is because the imaginative act is similar to the hallucination, minus the latter's positing of actual individuals and states of affairs. If I imagine a seven foot-long couch fitting through a doorway, I do not thereby posit, in either thought or imagination, any actual individuals. The object of imagination is not a mental image, either, contrary to some views that, like mine, stress the observational character of thought experiments (see, for instance, Gendler 2004). If what I were imagining were a real, mental particular—an image or model in my mind, say—then I would not be imagining but *introspecting*, and would not be imagining a seven foot-long couch, since nothing mental is a couch or seven feet long and I know this. What imagining and hallucinating have in common is that in such acts, properties and relations can be veridically presented.

Not only can properties and relations be veridically presented in both hallucination and imagination, but those acts are *self-giving* with respect to them, just as perception is with respect to actual individuals and states of affairs. In other words, intentional acts directed towards properties, relations, and other universals can be epistemically fulfilled on the basis of such presentations. While hallucination purports to be self-giving with

respect to actual individuals and states of affairs but is not, and while imagination does not even purport to present us with actual individuals, both are acts in which we can, at least in principle, behold properties. Furthermore, as Mark Johnston points out, hallucination can be an original source of *de re* knowledge of qualities (Johnston 2004, p. 131). I can acquire the concept of a square in a hallucinatory experience, for instance.

Acquiring knowledge on the basis of thoughts which are fulfilled, but not fulfilled on the basis of our experience of actually existing individuals, is utterly commonplace, and lies, to some extent or other, within virtually everyone's ken. A good chess player can determine how to force his opponent to make various moves, how to fork a queen, how to threaten mate, and so forth by imagining what will happen if he moves thus and so. He can and probably will perform various thought experiments by determining what will, could, or could not occur if he moved this way or that. This is not because he is playing a game of chess on a little mental chessboard. Perhaps he is. The important point is that it does not matter. What he learns in imagining possible chess moves is invulnerable to defeat by the factual existence or nonexistence of any individual chessboards, whether physical or mental. It is this arbitrariness with respect to the identity and even the existence of the individual exhibited in intuition that allows it to fulfill thoughts about any possible individual.

Earlier I argued that if thought experiments give us knowledge—and they sometimes do—then either they do so by means other than mere thinking or fulfillment via perception, or they do not have actual states of affairs as their objects. I rejected the first view, and have defended the second: thought experiments provide knowledge of properties and their relations—relations such as compatibility, incapability, and necessary co-instantiation. In at least some cases, moreover, thought experiments are founded on self-giving intuitions of those properties. Just as actual experiments are founded on presentations of actual states of affairs, thought experiments are founded on presentations of ideal states of affairs.

Before turning to some of the problems with my account, let me point out two benefits. The first is that it explains why thought experiments resemble actual experiments so closely: both are attempts to raise or answer a question by manipulating one or more variables and tracking what happens to the others, *on the basis of a self-giving presentation of those variables or something suitably related to them*. Thought experiments and actual experiments, then, differ not insofar as one involves “perception” and one involves mere thought. They both involve a kind of “perception,” at least if by “perception” we mean an act which presents, rather than emptily represents, its object, and both involve the fulfillment of various thoughts on the basis of those presentations. They differ, rather, insofar as actual exper-

iments are founded on presentations of actual states of affairs, while thought experiments are founded on presentations of states of affairs whose factual existence is irrelevant.

A second virtue of this account is that it does *not* conform to a characterization of thought experiments that does make them look genuinely magical. John Norton, for instance, writes, “Thought experiments are supposed to give us information about the physical world. From where can this information come?” (Norton 1996, p. 333) Roy Sorensen claims that, according to rationalism, “we can learn about the world without experience” (Sorensen 1992, p. 15). And that sounds downright spooky. On my view, we cannot know anything about the actual world on the basis of thought experiments *alone*, since thought experiments are founded on experiences that either contain no reference to actual individuals, or whose reference to those individuals is such that the individuals are treated as mere *possibilia*. The best they can do is tell us what is true in some possible world, or any possible world, or any possible world which has such-and-such features, properties, or kinds. For instance, Galileo’s cannonball thought experiment might be able to inform us that Aristotle’s theory of motion cannot describe any possible world. But it cannot tell us that our world contains cannonballs, or even that it contains any objects with mass. A thought experiment concerning how one can mate in three moves given a certain configuration cannot give us any information about our world by itself, since it contains no information about actualities at all. It cannot, without further ado, tell us how Smith can beat Jones. Thought experiments do, however, tell us about the world in conjunction with information that is about the actual world. Once I know that there is an actual chessboard configured in the same way as the one I have imagined, then I can know something about the actual world, namely that I can beat an actual opponent by moving actual pieces thus and so.

IV.

So far I have emphasized thought experiments founded on acts with sensuous or quasi-sensuous content—acts of hallucination and imagination. Not all thought experiments rely upon imagination, and almost none rely upon hallucination—though many *could*, as the thought experiment in the previous section shows. Many, and perhaps most, thought experiments succeed in spite of whatever imaginative acts we happen to carry out while performing them. Think, for instance, of Twin Earth thought experiments, or Maxwell’s Demon, or the notorious trolley cases in ethics. Not only do imaginative acts seem inessential to such thought experiments, they might sometimes prove to be a dangerous distraction. How does that

square with my quasi-perceptual model of rational intuition, according to which we can “intuit” universals and their relations?

Well, as a stab at an answer, I think it is because not all “intuitions” or self-giving acts are sensuous or quasi-sensuous, nor are all things which can be given capable of being presented in sensuous or quasi-sensuous acts. Fulfillment, at its best, is a matter of having the objects one thinks about given with the highest degree of adequacy which objects of that type are capable of being given. This often occurs without anything resembling imagery. And sometimes it occurs without doing anything more, seemingly, than understanding a proposition. When I think that $2+3=5$, I find it instantly and primitively compelling, just as, when I look at and listen to the television which is on and consider the proposition that the TV is on, I find myself unable to refrain from believing it. I can also merely think that the TV is on and feel no compulsion to believe. There is no contrasting case with respect to the proposition that $2+3=5$, however. Provided I actually *think about* the entities represented by the proposition, I find myself convinced that it is true.

Does this show that in certain cases intuition is dispensable and mere thinking is sufficient for rational insight? And if so, then why not settle for a more modest rationalism, such as Christopher Peacocke’s, that grounds a priori insight into the very fabric of concept possession, and do away with the problematic notion of rational intuition?

In the first place, a subject’s inability to think a given proposition without also finding it self-evident might show that merely thinking that proposition is sufficient for finding it evident. What it does not show, however, is that the complex phenomenon of fulfillment—the wedding of thought and intuition—is not taking place, or that finding the proposition evident is *just* a matter of understanding it. Consider the cogito. A proper phenomenological account of why the proposition that I exist is self-evident whenever I think it is not that merely thinking it *just is* to find it evident, or that various other known capacities of mine would be lacking if I did not. Rather, I find it evident because the state of affairs that it represents is always present to me in a privileged way.

In the second place, we can draw the contrast between merely thinking about something and having it given to consciousness in the case of logical and mathematical propositions too. I know there was a time when I did not find certain logical equivalences—DeMorgan’s Laws, for instance—evident. I had memorized DeMorgan’s Laws, and I could use them correctly in proofs, but without insight. Eventually I gained insight into their truth. Now I cannot consider them with due attention without finding them evident. This is not a matter of mere memorization or repetition. I also memorized and repeatedly applied the quadratic formula cor-

rectly while in school, and no behaviorist would be able to tell the difference between my understanding of it and my understanding of DeMorgan's Laws. But I never achieved insight into its truth. When it comes to certain propositions, it is not that I have acquired the ability to know them just by thinking them. It is that I have lost the ability to think them without having those thoughts fulfilled.

Again, I think my four year-old son understands the proposition that $2+3=5$. But he does not appear to find it self-evident. When I ask him whether $2+3=5$, he goes through a procedure—counting fingers—to determine the answer. That he understands the proposition is part of the best explanation for his successful use of this procedure. So even though I cannot produce the contrast between merely thinking that $2+3=5$ and having insight into it in my own case, that contrast does seem to show up between my four year-old son and me. I possess something that he, as yet, lacks, namely rational insight into the truth of that proposition, even though both of us possess the ability to think that proposition and, therefore, possess the concepts that make it up.

Third, even if mere concept possession fully explains our ability to find certain propositions evident, this is compatible with an account which insists that we bear direct epistemic relations to platonic entities. Suppose that every time I think that nothing can be red and green all over, I find it evident. Suppose, further, that this is best explained by my possession of the concepts 'red' and 'green'. This only constitutes an alternative to Platonism if possessing those concepts does not give me epistemic access to platonic entities. Similarly, if someone found it mysterious that we can stand in direct perceptual relations with distal objects, appealing to our possession of a visual system would not constitute, by itself, a demystifying move. If spelled out in the most plausible way—a way, for starters, that did not maintain that we see our visual systems—it would very likely simply constitute an explanation of what, in us, allows us to directly see distal objects. But what do I possess when I possess the concepts 'red' and 'green'? What I possess, I suggest, is the ability to carry out intentional acts directed at red and green and the various actual and possible states of affairs into which they enter. In exercising those concepts, I am not thinking *about* the concepts or the propositions in which they figure (which, to complicate matters, might themselves turn out to be platonic—there aren't as many concepts of red as there are acts of thinking of the color red). I am thinking *with* them. A very simple argument for this is that concepts and propositions represent things. But what I think and know *about* when I know that nothing can be red and green all over does not represent something. What I come to know about in thinking that nothing can be red and green all over are the colors red and green and

their inability to be jointly instantiated. Indeed, I think I know quite a bit more about red and green, and circles and bachelors, than I do about the concepts of them. I bet you do too. List everything you know about circles. Now list everything you know about the concept of circles. Which list contains more information?

Furthermore, the knowledge we have of properties and essences is not epistemically indirect, since we do not typically base our knowledge of those concepts' referents on prior knowledge of the concepts themselves.⁴ We do not find out about circles by examining the concept 'circle', but by using the concept 'circle'. Most of what we know about the concept 'circle', rather, depends on prior knowledge of circles. But these referents are the very platonic entities access to which is supposed to be problematic, and to which we *do* have access on any theory of concept possession that gets the objects of our a priori knowledge right. If a theory maintains that the possession of concepts gives me direct insight into universals and necessary states of affairs, then rational intuition, or something very much like it, has not been dispensed with, but folded right into the structure of concept possession. Such a theory is hardly a more conservative position than my own, according to which concept possession and rational intuition are distinct. And if it renders me incapable of being conscious or knowledgeable of such entities—if, for instance, it substitutes the properties red and green with the concepts of red and green—then it is objectionable on phenomenological grounds for replacing the genuine objects towards which I am directed with completely different ones.

V.

The account I am putting forward is most similar to that endorsed by Platonists such as James Brown. We shouldn't overstate the similarity between my view and Brown's. For one thing, as Harald Wiltsche has pointed out, I endorse Platonism for phenomenological reasons: they show up in experiences of fulfillment, and, unless extremely compelling reasons are presented not to do so, we ought to take such experiences at face value. Brown's Platonism, according to Wiltsche, is more theoretically motivated. For Brown, it provides the best alternative to Humean accounts of the laws of nature, which counterintuitively hold them to be nothing more than regularities. Furthermore, while Brown and I both hold that we

4. Even moderate rationalists do not think this. For Peacocke, possessing a concept amounts to "knowing what it is for something to be the concept's semantic value (its reference)." But, as he explains, this does not require "that someone can first grasp a concept and then go on to raise the question of what it is for something to be its semantic value. On the present theory, grasp consists in knowing the answer to that question, so such a state of ignorance is not possible" (Peacocke 1992: 22).

can be acquainted with universals in “some sort of intellectual perception” (Brown 2002, p. 1131), I do not think any contingent laws of nature are among the abstract entities that we can intellectually perceive in thought experiments, or at least not in thought experiments alone. Thought experiments tell us nothing about which things or laws are actual.

Despite these differences between my view and Brown’s, the metaphysical commitments of mine are liable to draw similar objections from similar quarters. One of the most common objections to this view is that, as Rachel Cooper puts it, “there is no account of *how* the Platonic universals are ‘perceived’” (Cooper 2005, p. 333). Although Cooper does not consider a Husserlian account like that above, the real objection appears to be the one raised by John Norton to the effect that “The mechanism of perception of Platonic laws is essentially completely mysterious” (Norton 1996, p. 360).

One response to this argument comes from Brown, who raises the same worries for ordinary sense perception. True, we know a great deal of the physical story that begins with the stimulation of the retina and ends with signals being sent to the visual cortex, but there ends our knowledge. Exactly how this process culminates in belief is something concerning which we are completely ignorant (Brown 1991, p. 65). This, however, should not make us skeptical about the existence of sense perceptual experiences giving rise to beliefs, nor should we be skeptical of the deliverances of sense perception itself. Similarly, the absence of an account of precisely how platonic perception works should not make us doubt its existence or results either.

Brown’s response illustrates something important about mysteriousness: a phenomenon might be mysterious insofar as we lack an understanding of how it comes about or operates, even while its existence is fairly uncontroversial. The existence of consciousness ought to be uncontroversial, for instance, as should the connection between consciousness and events in the brain, even though we have, as yet, no widely agreed upon account of the nature of that connection. The charge that seeing universals or essences is mysterious, then, had better amount to more than the mere observation that we lack an account of how it works, since that, by itself, gives us very little reason to doubt that it exists.

The real force of the objection, as Brown acknowledges, is that there is not, and cannot possibly be, any acceptable causal story—one couched in terms of things appropriately thought of as “mechanisms”—about how our minds (or brains) interact with platonic entities. At least that is Cooper’s objection to Brown’s account. And if there is no possibility of causal interaction with such entities, then we cannot even refer to them, much less know them. As Paul Benacerraf expresses it:

I favor a causal account of knowledge on which for X to know that S is true requires some causal relation to obtain between X and the referents of the names, predicates, and quantifiers of S. I believe in addition in a causal theory of *reference*, thus making the link to my saying knowingly that S *doubly* causal. (Benacerraf 1973, p. 671)

The argument in favor of this view is that we would not accept the claim that X knows that p if we had compelling evidence that X could not have stood in the right causal relations with those portions of reality in virtue of which p is true. Nostradamus, we reason, could not have known anything about or even made *de re* reference to Hitler, because he did not stand in the right causal relations to Hitler. Since we don't ever stand in causal relations to such *abstracta* as numbers and, presumably, platonic objects, reference to and knowledge of them is impossible.

As plausible as this position is, I think causation is a red herring here. Certainly causal relations aren't nearly sufficient to ground either knowledge or reference. Gamma rays and mitochondrial DNA have causally interacted with every human who has ever lived, but knowledge of them is a recent achievement. What really seems to matter is that knowledge of something requires that it be either *present* to you, present to someone whose testimony can be relied upon, or inferable from what is present to you. In the case of empirical objects and states of affairs, it is typically the case that something can be present in the required way—through perception—only if one is causally related to it. Causation is, with respect to these sorts of entities, necessary for the type of presence-to-consciousness that, I suggest, really makes Benacerraf's argument plausible.

For the causal theory of knowledge to provide an objection to the possibility of perceiving platonic entities, we would need a good reason to think that, just as in most cases of perception, causal contact is required to secure the right kind of knowledge- and reference-grounding presence-to-consciousness to the affairs in question. But I do not think that's especially obvious. In the first place, it is not clear to me that all of the entities to which a causal theorist is committed are things to which we stand in causal relations, including causation and theories. Does causation itself have causal powers? And do we really stand in causal relations with such things as *theories*? True, we stand in causal relations with the signs and symbols in whose terms theories are expressed, but so do the windows that vibrate when we utter them and the sheets of paper on which we print them, without thereby standing in causal relations to *theories*. A theory isn't identical with this or that inscription or utterance. If I wad up a piece of paper on which a theory is written and throw it in a fire, I will not have

incinerated a theory. We also stand in causal relations with our acts of thinking theories, but these aren't theories either; there are not as many theories of Special Relativity as there are acts of thinking it. Furthermore, the mental acts involved in thinking any theory whatsoever are part of the subject matter of psychology, but not every theory whatsoever is part of the subject matter of psychology. Absent an explanation of the "mechanism" in virtue of which we stand in causal relations to causation and to theories, I remain unconvinced that a causal theorist can consistently claim to know or even think about the causal theory of knowledge or reference.

In the second place, there is some reason for thinking that hallucinatory and imaginative experiences qualify as counterexamples to the causal theory. There is no actual instance of couch-hood, or even shape or color, before me when I hallucinate or imagine a couch, and insofar as causal relations hold only among actualities, I do not stand in any causal relations, at least at that time, with the properties revealed to me in those experiences. And yet not only do I refer to them, but they are presented to me.

One response to this argument is that even if we do not bear causal relations to what shows up in hallucinatory experiences and acts of imagining, we at least bear causal relations to actual instances of relevantly similar types of objects. That might be right, but any defense along these lines threatens the other premise of the anti-Platonist argument, namely that we do not have causal contact with platonic objects. If what explains my ability to hallucinate the color red is that I have previously been causally related to the color red, then I am, after all, capable of being causally related to the color red.

Furthermore, even if we set aside the question of whether we causally interact with platonic entities or not, seeing features or properties or universals is a thoroughly mundane occurrence in both ordinary perception and actual experimentation. Suppose that I think, emptily, about the color red. What kind of experience will fulfill my empty intentions towards the color red? Here is one kind: I perceive a red tomato, and then turn my attention to another red tomato, and then to another. In the course of this series of acts, I perceive three different individuals. But I also perceive *one* identical color, and perceive it *as* one identical color. My intention towards the color red is fulfilled throughout a unified sequence of experiences. And even though I was aware of some individual at each stage in that sequence, there was no individual such that I was aware of it across that sequence. I was aware of an identical something, even though there was no actual or intended individual that was identical. If I were only aware of individuals—whether individual tomatoes or individual property instances—I would not have beheld one thing. Though I might, at this

point, only have acquired the consciousness of a universal that is common to this and this and this tomato, it is not difficult to bring redness itself into view. One way is to realize that this property common to the three tomatoes can be instantiated by “an infinity of possible single instances” (Husserl 1977, p. 59) or, more simply, that the “universal is not bound to any particular actuality” (Husserl 1973, p. 329). Our understanding that universals are not bound to particular actualities is manifested in such trivialities as wanting to paint one’s kitchen red; we understand that the color red is not bound to what is presently red, and that the present color of one’s kitchen is not bound to it.

This itself is just one arbitrarily chosen example illustrating the pervasiveness of our awareness of universals. “The truth,” says Husserl, “is that all human beings see ‘ideas,’ ‘essences,’ and see them, so to speak, continuously” (Husserl 1982, p. 41). Sometimes the particularity of the objects of perception and thought becomes almost transparent, and they count in our cognitive and practical economy only insofar as they bear the properties that really concern us. I don’t particularly care *which* \$20 bill I receive from the ATM, as long as I receive *a* \$20 bill. When I look up the meaning of a word, I don’t care *which* token definition I encounter. I just want to encounter *a* token, never mind which, of the relevant type. And when I want to know whether heavy objects accelerate at the same rate as light ones, I don’t care which individuals I drop from which tower. I only care that they instantiate the properties whose relationships I want to discover. In all of these examples, it matters that I encounter actual individuals—I don’t want an imaginary \$20 bill—but the individuals are only relevant insofar as they instantiate the relevant properties, not in virtue of being the very individuals that they are. Any other qualitatively identical individual would have served just as well.

As the example of falling objects illustrates, the variables tracked and varied in actual experiments are almost always properties and relations, and experiments are typically construed as answering or raising questions about the relations among those sorts of variables. Suppose we discover that a 5 lb. weight—let’s call it “Timmy”—and a 10 lb. weight, “Jimmy,” fall at the same rate in conditions C. The variables that this experiment, if it is a sensible one, is designed to track are not Timmy and Jimmy, but the properties that they instantiate. The knowledge we acquire on the basis of this experience, moreover, is not confined to Timmy and Jimmy, but applies to any qualitatively identical objects in identical circumstances, and we know this. Here is a patently absurd follow-up question to the experiment: “Sure, Timmy and Jimmy fell at the same rate. But would two individuals exactly like Timmy and Jimmy also fall at the same rate in conditions identical to C?” Experimentally answering this question is a waste of

time. If we tried out two different weights and they did not behave in the same fashion, we would simply conclude that either those two individuals were not qualitatively identical with Timmy and Jimmy, that the conditions were not the same as those in C, or that we performed the experiment incorrectly on one of the trials. The reason, moreover, is due to our understanding that individuals do not fall under laws in virtue of being the individuals that they are, but in virtue of having the properties they have and standing in the relations that they stand—an insight that quite obviously depends on our ability to grasp properties. There are several versions of the problem of induction which are indeed problematic. This is not one of them. Determining whether qualitatively identical objects will behave in the same way in qualitatively identical circumstances is not like determining whether all crows are black on the basis of our knowledge of all observed crows. It's more like determining whether all crows which are qualitatively identical to the black ones are black.

The laws and regularities that scientists attempt to discover, and in whose terms the antics of individuals are explained, hold for individuals in virtue of their properties and the relations they bear to other propertied individuals. This holds in the cognitive sciences as well. According to most scientific accounts of perception, it is properties rather than individuals which are detected by our perceptual systems. When, for instance, it is claimed that people attuned to popular culture have a “Jennifer Aniston cell,” it is clear that this cell does not detect *Jennifer Aniston*. It does not, in standard conditions, fire if and only if Jennifer Aniston is visually present. It fires when subjects are presented with *pictures of Jennifer Aniston* (Quian Quiroga et al. 2005). No doubt it would also fire in the visual presence of the real Jennifer Aniston, or a twin of Jennifer Aniston, or a hallucination of Jennifer Aniston. What it quite plainly detects is Jennifer Aniston-esque features. Indeed, our access to features or properties is precisely what is *not* mysterious on most contemporary naturalistic theories of content. As Ruth Millikan puts it, “Failure to account for our capacity to represent individuals in language and thought has been, perhaps, the most serious failing common to contemporary naturalistic theories of content” (Millikan 2002, p. 43). Informational relations are lawful relations, and “there are no natural laws just about individuals” (Millikan 2002, p. 35). I will not try to determine whether Millikan’s own efforts to solve this problem are successful. What is noteworthy is that it is a problem at all.

Our mundane perceptual experiences, as well as those underlying scientific experiments, are saturated with reference to universals. Universals show up in both thought and experience, and our cognitive achievements, including those involved in carrying out empirical science, would be incomprehensible if they did not. What genuinely sets thought experiments

apart from actual experiments is not, then, that in the former case we are aware of universals. What sets them apart is that thought experiments, and the intuitive acts on which they are founded, have lost the connection to actuality that is essential to actual experiments. By allowing us to fulfill thoughts which are not rendered true or false by the actual existence of this or that individual, and therefore by the actual instantiation of this or that property, they allow us to discover relations that hold, or not, among instances of the relevant universals in any possible world. If there is a mystery involved in a priori insight, that mystery is not (just) our access to universals—which, far from relying upon a special faculty in addition to thinking and perception, is built into them—but the ability to “purify” our thoughts of any positing of actuality, to “treat the factual actuality of the single cases attained in variation as completely irrelevant” (Husserl 1977, p. 55; also see Kasmier 2010). Even if we do not know just how this ability works, the existence of the ability is beyond question. It is routinely exercised by the producers and consumers of fiction, lucid dreamers, savvy hallucinators, chess players of virtually all skills levels, and those with imaginations.

VI.

To sum up, then, on the view defended here, actual experiments and thought experiments are both genuinely experiments. Both are controlled attempts to determine the relation between variables on the basis of acts of epistemic fulfillment and, therefore, on the basis of self-giving presentations of those variables or objects suitably related to them. The difference is that actual experiments are directed at actual states of affairs, while thought experiments are directed at ideal states of affairs. What distinguishes thought experiments from actual experiments is not that thought experiments involve the awareness of universals. That’s true of actual experiments too. What sets thought experiments apart is that they, and the intuitive acts on which they are founded, have lost the connection to actuality that is essential to actual experiments.

References

- Benacerraf, Paul. 1973. “Mathematical Truth.” *Journal of Philosophy* 70: 661–79.
- Bishop, Michael. 1999. “Why Thought Experiments are Not Arguments.” *Philosophy of Science* 66: 534–41.
- Bokulich, Alisa. 2001. “Rethinking Thought Experiments.” *Perspectives on Science* 9: 285–307.
- Brown, James Robert. 1991. *The Laboratory of the Mind*. New York: Routledge.

- Brown, James Robert. 2004. "Peeking Into Plato's Heaven." *Philosophy of Science* 71: 1126–38.
- Cooper, Rachel. 2005. "Thought Experiments." *Metaphilosophy* 36: 328–47.
- Gendler, Tamar Szabo. 2004. "Thought Experiments Rethought—and Reperceived." *Philosophy of Science* 71: 1152–63.
- Häggqvist, Sören. 2009. "A Model for Thought Experiments." *Canadian Journal of Philosophy* 39: 55–76.
- Hopp, Walter. 2011. *Perception and Knowledge*. Cambridge: Cambridge University Press.
- Husserl, Edmund. 1969. *Formal and Transcendental Logic*. Translated by D. Cairns. The Hague: Martinus Nijhoff.
- Husserl, Edmund. 1970. *Logical Investigations*, 2 vols. Translated by J. N. Findlay. London: Routledge & Kegan Paul.
- Husserl, Edmund. 1973. *Experience and Judgment*. Edited by L. Landgrebe. Translated by J. S. Churchill and K. Ameriks. Evanston, IL: Northwestern University Press.
- Husserl, Edmund. 1977. *Phenomenological Psychology*. Translated by J. Scanlon. The Hague: Martinus Nijhoff.
- Husserl, Edmund. 1982. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book: General Introduction to a Pure Phenomenology*. Translated by F. Kersten. The Hague: Martinus Nijhoff.
- Husserl, Edmund. 2001. *Analyses Concerning Passive and Active Synthesis*. Translated by A. J. Steinbock. Boston: Kluwer Academic Publishers.
- Johnston, Mark. 2004. "The Obscure Object of Hallucination." *Philosophical Studies* 120: 113–83.
- Kasmier, David. 2010. "A Defense of Husserl's Method of Free Variation." In *Epistemology, Archaeology, Ethics: Current Investigations of Husserl's Corpus*. Edited by Pol Vandavelde and Sebastian Luft. New York: Continuum.
- McAllister, James W. 1996. "The Evidential Significance in Thought Experiments in Science." *Studies in the History and Philosophy of Science* 27: 233–50.
- Millikan, Ruth Garrett. 2004. *The Varieties of Meaning*. Cambridge: The MIT Press.
- Norton, John. 1996. "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26: 333–66.
- Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, Mass.: The MIT Press.
- Quian Quiroga, R., L. Reddy, G. Kreiman, C. Koch, and I. Fried. 2005. "Invariant Visual Representation by Single Neurons in the Human Brain." *Nature* 435: 1102–07.

- Sorensen, Roy. 1992. *Thought Experiments*. Oxford: Oxford University Press.
- Wiltsche, Harald. 2012. "Platonism and Thought Experiments: Phenomenology, How Deep is Your Love?" Paper presented at the Workshop "Thought Experiments in Science: Four Blind Spots," University of Toronto, March 2012.