On the Nature of Thought Experiments
and a Core Motivation of Experimental Philosophy

appeared in Philosophical Psychology (2010), 23: 547 - 64.

Joseph Shieber¹
Lafayette College, Pennsylvania (USA)

Abstract: In this paper I discuss some underlying motivations common to most strands of experimental philosophy, noting that most forms of experimental philosophy have a commitment to the claim that certain empirical evidence concerning the level of agreement on intuitive judgments across cultures, ethnic groups or socioeconomic strata impugns the role that intuitions play in traditional “armchair” philosophy. I then develop an argument to suggest that, even if one were to grant the truth of the data adduced by experimentalists regarding the level of agreement – or lack thereof – regarding intuitive judgments among various groups, this would nevertheless not yet provide sufficient basis to reject the role of intuitions in traditional philosophical theorizing. Though this argument, if successful, will not prove fatal to all forms of experimental philosophy, it would limit the scope of experimental philosophical criticisms of traditional philosophical practice.

1. Introduction

Although less than a decade old, experimental philosophy must already be recognized as one of the most self-consciously iconoclastic movements to arise

¹ My thanks to Thomas Grundmann, Joachim Horvath, Ernest Sosa, Jonathan Weinberg, Thomas Nadelhoffer, the participants at the 2008 symposium “Armchair in Flames,” and an anonymous reviewer of this journal for helpful comments on an earlier version of this paper.
within philosophy at least since Richard Rorty published *Philosophy and the Mirror of Nature* in 1979. Indeed, the very name of a 2008 conference devoted to experimental philosophy, “Armchair in Flames,” attests to the fact that the movement of experimental philosophy—dubbed by Appiah [2007] as the “new new philosophy”—sees itself as posing a serious challenge to the project of traditional philosophical inquiry. If the proponents of this new movement have their way, the comfortable armchair from which philosophers have grown accustomed to issuing their pronouncements would be no more, having simply gone up in smoke.

As Appiah suggests, the key to this assault on traditional philosophy is an attempt to impugn the use of thought experiments as a source of evidence by proponents of traditional philosophy. In this paper, however, I suggest that this assault is unsuccessful, in that even if one accepts core assumptions regarding what experimental philosophers take to be the role of thought experiment in traditional philosophy, what experimental philosophers take to be evidence undermining the utility of thought experiments does not in fact impugn the use of thought experiments as data in philosophy.

I proceed as follows. I begin with a discussion of some underlying motivations common to most strands of experimental philosophy, noting that most forms of experimental philosophy have a commitment to the claim that certain empirical evidence concerning the level of agreement on intuitive judgments across cultures, ethnic groups or socioeconomic strata impugns the role that intuitions play in traditional “armchair” philosophy. I then develop an argument to suggest that, even if one were to grant the truth of the data adduced by experimentalists regarding the level of agreement – or lack thereof – regarding intuitive judgments among various groups, this would nevertheless not yet provide sufficient basis to reject the role of intuitions in traditional philosophical theorizing.

### 2. A Central Motivation of Experimental Philosophy

In order to examine the core assumption underlying experimental philosophy that is of interest here, it will be useful to introduce a working taxonomy of the branches of experimental philosophy. To do so, I will borrow from the account of “The Past and Future of Experimental Philosophy” offered by Nadelhoffer and Nahmias [2007]. There, Nadelhoffer and Nahmias distinguish among three main sorts of experimental philosophy: (1) experimental analysis, or A-experimentalism, (2) experimental descriptivism, or D-experimentalism, and (3) experimental...
restrictionism, or R-experimentalism.

As Nadelhoffer and Nahmias characterize it, A-experimentalism has as its primary goal the experimental exploration of what intuitions ordinary people normally express and the examination of the relevance of that data for philosophical debates. This results in a method for “[testing] philosophers’ claims that their positions align with common sense and [challenging] those claims that are not supported by the evidence.” (Nadelhoffer and Nahmias [2007], 126) Clearly, A-experimentalism comports with what is often taken to be the traditional view of the role of intuition in philosophy; where A-experimentalism is revolutionary is in the method endorsed for canvassing those intuitions. For, whereas traditional philosophy suggests that all of the evidence on folk intution may be acquired by the solitary philosopher reflecting in her armchair, the project of experimental analysis involves the employment of “methods borrowed from experimental psychology to probe folk intuitions in a controlled and systematic way.” (Nadelhoffer and Nahmias [2007], 126)

In contrast with experimental analysis, neither D-experimentalism nor R-experimentalism involves an endorsement of the relevance of intuitions for the proper conduct of philosophy. Whereas D-experimentalism might profitably be characterized as neutral with respect to the relevance of intuitions for philosophy, R-experimentalism, on the other hand, is actively critical of the use of intuition in philosophical theory-building. Nadelhoffer and Nahmias suggest that, unlike experimental analysts, who use their study of folk intuitions to afford a stronger evidential foundation to the use of intuitions in philosophy, experimental descriptivists use the data that they gather concerning “the underlying psychological processes and cognitive mechanisms that produce our intuitions” to understand more clearly how the mind works. (Nadelhoffer and Nahmias [2007], 127)

Of course, not all processes or mechanisms are sufficiently reliable in all of the situations in which they yield evidence. Thus, it should hardly be surprising that at least some of the data gathered by experimental descriptivists suggest that the sources of what seem to be “pure” intuitions are in fact sometimes influenced by subjects’ cultural or socio-economic backgrounds. If, however, traditional philosophical claims in fact rely on intuitive evidential support -- and if, furthermore, much of that evidence is infected by what the data show to be philosophically irrelevant factors -- then some might suggest that this would seem to provide a serious challenge to the project of traditional philosophy and its reliance on intuition. Raising these concerns and gathering further empirical evidence to
strengthen the case against traditional philosophy is the project that Nadelhoffer and Nahmias, following Alexander and Weinberg [2006], term *experimental restrictionism*. It will be helpful for us to summarize some of the features of these forms of experimental philosophy by means of a table:

<table>
<thead>
<tr>
<th>Branch of Experimental Philosophy</th>
<th>Stakes Claim on Importance of Intuitions to Traditional Philosophy</th>
<th>Position on Intuitions as Evidence in Philosophical Theorizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Analysis</td>
<td>Yes</td>
<td>Endorse</td>
</tr>
<tr>
<td>Experimental Descriptivism</td>
<td>No</td>
<td>Agnostic</td>
</tr>
<tr>
<td>Experimental Restrictionism</td>
<td>Yes</td>
<td>Oppose</td>
</tr>
</tbody>
</table>

To say that a position stakes a claim on the importance of intuitions to traditional philosophy is to say, following Weinberg, Nichols, and Stich [2008], “that a sizable group of [philosophical] projects ... would be seriously undermined if one or more of a cluster of empirical hypotheses about [philosophical] intuitions turns out to be true.” (Weinberg, Nichols, and Stich [2008], 17) Thus, as the table illustrates, A-experimentalism and R-experimentalism share a commitment to describing the traditional philosophical project as assigning a privileged place to intuitive evidence in philosophical theory-building. However, as the second column of the table illustrates, these two positions are diametrically opposed with respect to their positions on whether intuitions *ought* to have such a place. In contrast, D-experimentalism, given as it is to the exploration of intuitions and their sources as a tool for the exploration of the human mind, need stake no claim with respect to the role of intuitions in traditional philosophy.

In the discussion that follows, I focus primarily on two features common to most or all of the forms of experimentalism, one explicit and one implicit. The first is the notion, common to both A-experimentalism and R-experimentalism, that empirical evidence of intuitive disagreement impugns the core commitment of
traditional philosophy to the role of intuitive evidence as data in philosophical theorizing. The second is the commitment on the part of the experimentalists, explicit in D-experimentalism and tacit in A-experimentalism and R-experimentalism, to seeing sources of intuition naturalistically, in terms of, as Nadelhoffer and Nahmias put it, “the underlying psychological processes and cognitive mechanisms that produce our intuitions.” In what follows, the criticism of experimental philosophy that is developed here will suggest that, even if one adopts a naturalistic understanding of the source of intuitions, none of the evidence presented by the experimentalists is sufficient to impugn the appeal to intuitions as evidence in traditional philosophical argumentation.

Given this sketch of the criticism to be developed, the argument here clearly will not apply to all forms of experimental philosophy as delineated by Nadelhoffer and Nahmias; experimental descriptivism will remain untouched by these considerations. Thus, though the criticism advanced here, a criticism posing a challenge to both A-experimentalism and R-experimentalism, will be a serious enough challenge for the project of experimental philosophy as it is currently conceived, there is also room for optimism about the surviving form of experimental philosophy, experimental descriptivism. Indeed, as noted in the conclusion of this paper, it is difficult to see how a project like experimental descriptivism, with its clear implications for the epistemology of philosophy itself, could fail to be of value for traditional philosophy.

3. **The Role of Intuition**

As discussed in the previous section, both A-experimentalism and R-experimentalism agree in depicting traditional philosophy as committed to according intuitions an evidentiary role in theory-building. Where they differ is in the role that A-experimentalism and R-experimentalism themselves accord to intuitions. A-experimentalists think that, precisely because intuitions are important sources of data, philosophers should employ better – that is, empirical – techniques in gathering that data from the actual “man on the Clapham omnibus,” rather than simply assuming that the philosophers’ own intuitions are a good gauge as to what the average Clapham commuter thinks. In contrast, R-experimentalists think that, because of the variability in intuitions among various socio-economic, ethnic, and cultural groups, intuitions cannot even serve as important sources of data in philosophical theorizing.

Even given these differences between A- and R-experimentalists, one can see
them both as arguing against traditional philosophical projects’ employment of intuitions as follows:

1. A sizable group of philosophical projects and/or methodologies would be seriously undermined if one or more of a cluster of empirical hypotheses about philosophical intuitions turns out to be true.

2. Some of those empirical hypotheses are true.

3. Thus, a sizable group of philosophical projects and/or methodologies are seriously undermined.

A-experimentalists affirm (1) in that they see empirical evidence that the “folk” disagree with professional philosophers as bearing on those philosophers’ methodology of appealing to armchair canvases of the philosopher’s own intuitions to provide evidence for philosophical theorizing, while the R-experimentalists affirm (1) because they think that, if the empirical evidence they adduce regarding disagreement among different socio-economic, ethnic, and cultural groups is correct, then all appeals to intuitions as evidence for philosophical theorizing would be bankrupt.

In discussing this shared assessment of the bearing of empirical evidence regarding disagreement in intuitive judgments among various different ethnic, cultural or socio-economic groups on the role of intuitions in traditional philosophy, I proceed as follows. In the remainder of this section I examine more closely some of the ways in which intuitions in fact figure in philosophical theorizing. In particular, I focus here on thought experiments, appealing to a taxonomy of thought experiments developed by Sorensen to better structure the discussion. Then, in the following sections, I suggest that this way of analyzing the role of thought experiments allows us both to do justice to the data adduced by experimentalists and to allow for the role traditionally accorded to intuitions by traditional armchair philosophy. If this is the case, however, then premise (1) in the experimentalist argument presented above is false: the account presented here would provide a scenario according to which it could both be the case that even all of the cluster of empirical hypotheses about philosophical intuitions to which the experimentalists appeal turn out to be true, but that, nevertheless, the traditional employments of intuitions in philosophical projects would not be seriously undermined.

Of course, there are a number of ways in which intuitions may figure in theorizing in general, and in philosophical theorizing in particular; thought
experiments are merely one such deployment of intuitive judgments. For this reason, before proceeding it is important to address an immediate objection that one might raise to this narrowing of focus to thought experiments in particular. Indeed, arguably the core way in which intuitions yield evidence would be directly, as when one claims that a belief in a particular proposition -- e.g., the claim that no object has a surface that is red and green all over at the same time -- is believed due to its intuitive plausibility.

This may well be the case. Nevertheless, the method through which experimental philosophers of almost all stripes -- and in particular adherents of A and R-experimentalism -- assess folk intuitions is by means of the employment of thought experiments. Thus, the limitation of focus here to a discussion of thought experiments cannot be a source of criticism from the camp of experimental philosophers. Thus, for the purposes of the present discussion it will be sufficient to note that experimental philosophers will not be able to blunt the criticism advanced here by questioning the focus on thought experiments in this discussion of the evidentiary status of intuitions in philosophy.

It will aid discussion to have a framework with which to characterize traditional philosophical appeals to thought experiments in theorizing in a way that is broadly compatible with the experimentalists’ tacit philosophical naturalism. Adapting Sorensen [1992], those views will be characterized here in terms of four, partially overlapping, explanatory models: (a) the recollection model, (b) the transformation model, (c) the homuncular model, and (d) the rearrangement model. As Sorensen notes, none of the models discussed need be accepted as exclusive: “There are thought experiments conforming to each -- and some that fit all models simultaneously.” (Sorensen [1993], 109) Though the overlapping nature of the explanatory models may well disturb some readers’ desire for parsimony, by appealing to all four models it is hoped more fully to cover all potential cases that readers might bring to mind in considering the range of thought experiments.

As in the case of the narrowing of focus of the discussion here to a

---

2 Readers of Sorensen [1993] will recall that Sorensen also has a fifth model -- the cleansing model. Pace Sorensen, for whom the cleansing model is in fact to be preferred, I find the cleansing model obscure; regardless, the other four models seem sufficient to account for the range of thought experiments deployed in philosophy.
consideration only of thought experiments, some readers might object to the employment of Sorensen’s taxonomy, particularly given the fact that there are a number of other philosophical accounts of thought experiments that one could have employed instead.³ It is important to underscore that the argument presented here is intended to comport with the experimentalists’ stated presuppositions and to show that, even given these presuppositions, premise (1) of the experimentalists’ argument can be rejected. Thus, since the present dialectic requires only that I provide evidence that there is a framework capable of accommodating both the experimental data adduced by the experimentalists and the role of thought experiments as they are traditionally employed by armchair philosophers, however, the fact that there are alternate accounts of thought experiments not compatible with the account presented here is irrelevant to this discussion.

The recollection model takes its motivating idea from Plato’s notion of knowledge as recollection. According to this model, the use of thought experiment is to help retrieve information already possessed by the reasoner that is salient to the question at hand. One example cited by Sorensen to illustrate a case particularly well-suited to this model is the use of a thought experiment to illustrate the concept of air resistance to physics students. (Cf. Sorensen [1992], 90) The students are encouraged to imagine dropping playing cards onto the surface of a table, first with the edge pointing down toward the table and then with the flat surface of the card parallel to the table. Perhaps mindful of the example of a parachute, the students will visualize the cards falling faster edge-down, thus underscoring for them the importance of the surface area of a falling object in influencing the speed of descent. Though no new information has been acquired, the already-possessed information salient to the question at hand is made available to the students in a way that aids in their learning.

³ Some other influential discussions of thought experiment include Brown [1991], Haggqvist [1996], and Gendler [2000]. Although I employ Sorensen’s discussion for the purposes of this paper, the use to which I put it would be compatible with Haggqvist’s and Gendler’s discussions as well; only Brown’s discussion—relying, as it does, on a robust Platonist notion of thought experiment—is incompatible with the use to which I put thought experiments here. It is important to emphasize, however, that an exclusion of Brown’s robust Platonist notion of thought experiments is completely in line with the experimentalists’ tacit—or, on the part of the descriptivists, explicit—naturalism.
The term transformation model is the name that Sorensen applies to the method through which one can make one’s mastery of linguistic rules productive for the purpose of conceptual analysis by imagining what one would say in a given situation. Sorensen terms this the transformation model because, as it is traditionally conceived, this process of imagination involves transforming one’s knowledge how to speak, say, English, into knowledge-that, propositional knowledge concerning the concepts under investigation. As Sorensen describes such cases, “when the informant is a philosopher or linguist, the procedure is rationally persuasive. By monitoring your own performance, you become living testimony for the thesis at issue.” (Sorensen [1992], 93) An example Sorensen adduces to illustrate this model is Frankfurt’s use of the thought experiment of the evil mind-control scientist to refute the notion that “A person is responsible only if he could have done otherwise.” (Frankfurt [1969]) As Sorensen summarizes the lesson of that example, “Knowledge of how to use ‘responsible’ transforms into knowledge that responsibility is compatible with the inability to do otherwise.” (Sorensen [1992], 93)

The homuncular model is best described in terms of Daniel Dennett’s notion of cognitive autostimulation, the process by which one poses a question to oneself in order to activate the cognitive modules relevant to the question at issue so that the application of those modular capacities to the question at hand might deliver a response relevant to that question. As Dennett describes it, “pushing some information through one’s ears and auditory system may stimulate just the sort of connections one is seeking, may trip just the right associative mechanisms, tease just the right mental morsel to the tip of one’s tongue. One can then say it, hear oneself say it, and thus get the answer one was hoping for.” (Dennett [1984], 40)

Thus, the homuncular model makes use of cognitive autostimulation as a way of making the deliverances of subpersonal cognitive modules available to consciousness. In this way, we can appreciate Sorensen’s description of the homuncular theory of armchair enquiry as explaining “the possibility of newsless edification by modeling the intrapersonal case on interpersonal information processing. A piece of information that is useless to me may become useful when passed to someone who is better able to digest it or who already has information that can be combined with the new item.” (Sorensen [1992], 95) Furthermore, this helps us to see how the homuncular model may be understood as a more general model of which the transformation model is a special case, as it relies solely on the
activity of the language module or modules.\(^4\)

Thus, to consider merely one illustrative example plausibly understood as an instance of the homuncular model, consider Haidt, et al. [1993]. In that study, Haidt and his colleagues presented Brazilian and U.S. adults and children of high and low socioeconomic status with a number of scenarios involving victimless yet offensive actions. Haidt and his colleagues found that only one group -- the high socio-economic status U.S. adults, exemplified by college students at the University of Pennsylvania -- consistently judged the harmless acts not to be wrong, despite agreeing with the other groups in finding those acts to be disgusting. Furthermore, in all cases except those of the U.S. college students Haidt and his colleagues found that experimental subjects seemed to base their moral judgments on their automatic feelings of disgust and to justify those judgments only in a post-hoc fashion. For this reason, Haidt [2001] argues that default moral judgment is better understood as affective and intuitive rather than rational and deliberative. Of particular significance for our purposes here, the results of Haidt et al. suggest that, in the default case, subjects assessed the moral implications of the thought experiments under investigation by accessing the cognitive subsystems controlling for disgust responses; when disgust is triggered, the default response is one of moral reprobation.

There is still one remaining explanatory model for dealing with thought experiments: the rearrangement model. As noted above, the recollection model primarily involves the retrieval of already-possessed information according to its salience for the problem under consideration. Sometimes, however, one already is conscious of information, but is simply unable to recognize that it is in fact salient. In such cases, the application of armchair reflection that falls under the purview of the rearrangement model can be of use. As Sorensen puts it, “this model is inspired by situations in which the information at hand is made more digestible by changing its form. We have the information in the sense of possessing the data but lack it in the sense of not being able to draw needed inferences from it. The information is ungainly, like a plank grasped at an extreme end instead of at its center of gravity.” (Sorensen [1992], 99)

For an illustration of the rearrangement model from the history of science, consider one of Galileo’s thought-experiments that led to the establishment of the

\(^4\)I argue for the usefulness of the homuncular model in making intuition safe for philosophical naturalism in Shieber [forthcoming].
law of inertia, thereby overturning the central principle of Aristotelian dynamics. (See Figure 1.) As Park [1988] describes the thought experiment,

Galileo ... found by experiment that if a ball is allowed to roll down one incline and up another with friction reduced as far as possible, then, regardless of the slopes of the two inclines, the ball rolls up the second one until it reaches a height that is very nearly equal to the height from which it started. Suppose now that the second plane is made flatter and flatter. The ball will have to travel further and further to regain its initial height. Thus, if the second plane were made smooth and perfectly level, it would go on for a very long way, showing that once started, violent motion can persist indefinitely. (Park [1988], 205)

Sorensen glosses the significance of Galileo’s achievement with this thought-experiment thus: “According to Aristotle and common sense, it is natural for things to slow down and come to rest; continued movement is what needs explaining. But after Galileo’s thought experiment, continued movement seemed natural and slowing

5 Cf. Eric M. Rogers [1966].
required explanation.” (Sorensen [1992], 9) The rearrangement model thus encapsulates the sort of thought-experiments leading to the types of Copernican perspectival shifts illustrated by the story of the following conversation between Wittgenstein and G. E. M. Anscombe: “Anscombe remarked that it wasn’t surprising that people used to believe that the sun goes round the earth, because that is how it looks. And apparently Wittgenstein asked in reply: ‘How would it look if it looked as if the earth goes round the sun?’” (Glock [2003], 143)

As in the discussion of the taxonomy of experimental philosophy in the previous section, it will be helpful to summarize some of the features of these explanatory strategies of the role of thought experiments by means of a table:

<table>
<thead>
<tr>
<th>Explanatory Strategy</th>
<th>Inputs</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recollection Model</td>
<td>Information stored in memory</td>
<td>Retrieval according to salience</td>
</tr>
<tr>
<td>Transformation Model</td>
<td>Outputs of the language faculty</td>
<td>Conversion of linguistic know-how to knowledge-that</td>
</tr>
<tr>
<td>Homuncular Model</td>
<td>Outputs of cognitive modules</td>
<td>Intrapersonal communication as interpersonal communication</td>
</tr>
<tr>
<td>Rearrangement Model</td>
<td>Information available to consciousness</td>
<td>Perspectival shift</td>
</tr>
</tbody>
</table>

Although, as the table makes clear, all of the inputs of the forms of thought experiment explained by the various strategies are distinct, it is equally clear that in each case the method of thought experiment operates on information already present in the subject (actually present, in the case of stored propositional knowledge, and dispositionally available, in the case of encoded knowledge-how). In this way, all of the explanatory strategies respect one of Mach’s core motivations in introducing the notion of the Gedankenexperiment - viz., that such thought experiments are nourished from “a treasure-store ... uncomprehended and unanalyzed in our percepts and ideas ... which is ever close at hand and of which only the smallest portion is embodied in clear, articulate thought.” (Mach, quoted in Sorensen [1992], 51) Thus, there is a very real sense in which the employment of any of the forms of thought experiment accounted for by the four explanatory strategies does not involve
the acquisition of new information on the part of the subjects employing those thought experiments; rather, those subjects already possess the raw information forming the inputs of those thought experiments.

Of course, however, there is another sense in which all of these thought experiments may be seen to result in the acquisition of new information. As an analogy, consider the case of a library containing volumes of non-fiction. Organizing that library by labeling its holdings and cataloging them does not increase the store of information that those volumes already contain; it simply provides a more useful structure through which to access and employ that existing information. However, there is another sense in which providing such a structure does involve an increase in information. Thus, e.g., the recollection model may be compared to consulting with a reference librarian on a question of interest: one learns propositions such as that such and such a book (on such and such a subject) is relevant to the topic of concern, and that that book may be found in such and such a location. Or the rearrangement model may be compared to an instance of analogical reasoning: one learns that a certain phenomenon is analogous to some other phenomenon, perhaps heretofore considered unrelated, that the two formerly unlinked phenomena in fact share relevantly similar characteristics, and that problem-solving techniques for dealing with the one phenomenon might profitably be applied to problems arising with respect to the other phenomenon.

Evidence that the strategy of employing thought experiments may profitably be thought of in analogy to other -- and very often non-deductive -- methods of belief revision comes from the way in which we assess thought experiments. Rather than judging a thought experiment in terms of truth or untruth, we speak of the elegance of a thought experiment, or in terms of its suggestiveness or fruitfulness. In this way, it would seem that thought experiments belong more to Reichenbach’s [1938] context of discovery, as opposed to the context of justification -- as, indeed, Hempel explicitly suggests in his discussion of the role of thought experiments in theory construction (Hempel [1968], 165). For those of us who are post-positivists, however, and who no longer require that our theories derive solely and deductively from logical posits and sense data, there is no reason to deny that the outputs of such heuristics can have evidentiary value -- to deny, that is, that heuristics may, to borrow a line from Gigerenzer, et al. [2000], “make us smart.”

Note furthermore that, if understood in this way, thought experiments must be assessed in terms of their epistemic utility in ways similar to those employed to assess other strategies of theory development involving considerations of, say,
simplicity, elegance, explanatory coherence, suggestiveness, etc. Given this, however, it should be obvious that using thought experiments to assess and develop theories -- like using those other considerations of simplicity, elegance, etc. -- will sometimes yield different results for at least two reasons. Different agents could have different responses when employing the same strategy (considering a particular thought experiment, say, or judging a particular theory with respect to its elegance) either because they are employing the strategy differently or because they are employing the strategy in an identical way but they are employing it on a different set of inputs (i.e., a different set of beliefs, say, or a different background theory).

Given this, however, it should be clear that the fact that two subjects could employ a strategy and thereby achieve disparate results in reasoning is not an indictment of the strategy. In the latter case, the case in which two people are using the same strategy, but on different inputs, this should be obvious. Thus, if I use reductio ad absurdum to weed out inconsistencies within my beliefs, arriving at a set of beliefs containing \{P_1, \ldots, P_n, Q_{n+1}\} and some other person uses the same strategy to arrive at a set of beliefs containing \{Q_1, \ldots, Q_n, \neg Q_{n+1}\}, nobody in their right minds would take this as evidence for an indictment of the method of using reductio ad absurdum to weed out inconsistencies within one’s set of beliefs.\footnote{One might complain that the comparison to the employment of reductio ad absurdum is inapt, given that the comparison we are drawing here, between the employment of thought experiments and other strategies of theory development involving considerations of, say, simplicity, elegance, explanatory coherence, suggestiveness, etc., would suggest that we take the employment of thought experiment to be a form of abductive inference or inference to the best explanation, while reductio ad absurdum is a form of deductive argument. Of course, reductio is deductive, but the choice of which beliefs to abandon, from a set of beliefs that one has shown—by means of reductio—to be inconsistent, is not dictated by logic, as Quine, and, before him, Duhem, long ago suggested.}

4. Implications for Experimental Restrictionism

Given this way of understanding the role of thought experiment in traditional philosophical inquiry, however, it would seem that the motivation for much of what is conducted under the rubric of experimental philosophy is due for significant reconsideration. I will trace this out by considering the implications of this
reassessment of the evidentiary role of thought experiments for each of the types of experimental philosophy canvassed in section 2 above.

Thus, consider, for example, the use to which the experimental restrictionists, Weinberg, Nichols, and Stich, put the fact of diversity in intuitive epistemic judgments in arguing against the employment of intuitive evidence in epistemology (Weinberg, Nichols and Stich [2008], 22):

i. There are people who “have epistemic intuitions which ... yield the conclusion that their strategies of reasoning and belief formation [strategies that are “significantly different from ours”] lead to epistemic states that are rational.”

ii. “If this is right, then it looks like the [intuition-based] strategy for answering normative epistemic questions might sanction any of a wide variety of regulative and valutational norms.”

iii. “And that sounds like bad news for an advocate of the [intuition-based] strategy, since the strategy doesn’t tell us ... how we should go about the business of forming and revising our beliefs.”

Note, however, what happens when we see the role of thought-experiments - - the sorts of intuitive inputs that Weinberg, et al., explicitly test in their establishment of the fact of diversity in epistemological intuitive judgments -- as we have suggested in the previous section. If thought-experiments in fact function in a way analogous to that of -- deductive, inductive, or abductive -- inferential mechanisms, then the argument of Weinberg, et al., is in fact structurally identical to the following one:

i.’ There are people whose employment of the practice of maximizing internal consistency among their beliefs yields the conclusion that their maximally consistent set of beliefs (a set that is significantly different from ours) is the set that they ought to endorse as correct.

ii.’ If this is right, then it looks like the strategy of maximizing internal consistency among one’s beliefs as a method of answering the question of which set of beliefs to endorse as correct might sanction any of a wide variety of maximally consistent sets of beliefs.

iii.’ And that sounds like bad news for an advocate of the strategy of
maximizing internal consistency among one’s beliefs, since the strategy doesn’t tell us how we should go about the business of endorsing our beliefs.

As should be clear, though, this argument is a terrible one. The mere fact that the strategy of maximizing internal consistency yields different outcomes based on different inputs does not speak against the value of employing that strategy. Of course, it does demonstrate that mere employment of the strategy, irrespective of the set of beliefs or background theories that serve as its inputs, cannot guarantee reliably true results. However, as the discussion of externalist positions in epistemology has demonstrated over the last four decades or so, no purely internalistically described strategy of belief adjudication or modification by itself could offer such a guarantee.

There are two points here that deserve emphasis. The first is that the truth-conduciveness of the strategy of employing thought-experiments always depends on the context in which that strategy is employed—and, in particular, on the set of beliefs or background theory of the subject entertaining the thought experiment. For this reason, the mere fact that a subject in a different context—e.g., with a different set of beliefs or background theory—achieves a different result when employing the strategy is irrelevant to the truth conduciveness of the strategy as deployed in the original context.

Second—and perhaps more significantly, given the dialectic of the experimental restrictivists’ criticism of the use of thought experiments—the fact that someone else comes to a different result as a consequence of considering a particular thought experiment need not, on the face of it, have any bearing on my employment of the thought experiment in question. For, as our discussion of the nature of thought experiments demonstrated, thought experiments serve as a method for singling out information from a specified belief set—viz., the belief set of the subject employing the thought experiment. Singling out this information might serve many purposes for the subject, including making his salient doxastic commitments more explicit, directing his attention to relevant information, and so on. Given this, the fact that others would arrive at different results in employing the same thought experiment has as yet no bearing on the original subject’s deployment of the thought experiment.

5. A Reconsideration of the Gettier Case
One might object, however, that our treatment of thought experiments so far has been too broad. In particular, one might note that the experimental restrictionists ought not to be taken as arguing against the use of intuition in philosophy *tout court*, but rather as arguing against a more restrictive class of intuitions. Though such a rejoinder on behalf of the restrictionists seems to suggest that the onus should be on the restrictionist, then, to better delineate which intuitions are the target of her critique, we will leave aside such issues here. Instead we will focus our discussion on one of the thought experiments that has been a focus of much of the experimental philosophical literature discussing epistemological intuitions, viz., Gettier cases.

As detailed in Weinberg, et al. [2008], East Asians and subjects from the Indian subcontinent are demonstrably far more likely to attribute knowledge to subjects in Gettier cases than are Westerners. Indeed, with respect to subjects from the Indian subcontinent, Weinberg, et al., characterize their results by noting that “it seems that what counts as knowledge on the banks of the Ganges does not count as knowledge on the banks of the Mississippi!” (Weinberg, et al. [2008], 30) These results, in fact, were one of the central pieces of evidence supporting the claim of Weinberg, et al., that there are people who “have epistemic intuitions which ... yield the conclusion that their strategies of reasoning and belief formation [strategies that are “significantly different from ours”] lead to epistemic states that are rational,” cited as premise (i) in the argument discussed in section 4, above.

Of course, if the argument sketched in section 5 is correct, it ought not be surprising that there are people who have epistemic intuitions that are “significantly different from ours.” Thus, suppose – in analogy to Sorensen’s discussion of Frankfurt’s use of thought experiment to argue against the Principle of Alternate Possibilities – we understand the consideration of Gettier cases as involving an instantiation of the *transformation model* of thought experiment, in this case a process by which knowledge of how to use “knows” transforms itself into knowledge that knowledge is incompatible with certain species of luckily formed justified true belief.

Much empirical research on concept acquisition, possession, and use, however, has demonstrated the influence that a subject’s prior beliefs have on her acquisition and deployment of other concepts. Given this, the transformation

---

7 Indeed, Weinberg has raised this objection to me in conversation.

8 For a good summary discussion, cf. Murphy [2002], particularly chap. 6.
model of thought experiments would predict that subjects with differing background beliefs relevant to the concept of knowledge will acquire or deploy that concept in differing ways, some of which may well result in one or the other subject’s becoming prone to error in her conceptual intuitions.

Note that the point here isn’t to recapitulate a response to experimentalism that Weinberg et al. have anticipated – viz., that there might be different senses of “knowledge.” Rather, the point here is that, due to differing background beliefs, some might use our concept of knowledge improperly in a certain range of cases. This is a point that Weatherson has recently made quite forcefully:

At one stage it was thought that whales are fish, that Mars is a star, the sun isn’t. These are beliefs, not intuitions, but there are clearly related intuitions. Anyone who had these beliefs would have had the intuition that in a situation like this (here demonstrating the world) the object in the Mars position was a star, and the objects in the whale position were fish. The empirical errors in the person’s belief will correlate to conceptual errors in their intuition. To note further that the kind of error being made here is conceptual not empirical, and hence the kind of error that occurs in intuition, note that we need not have learned anything new about whales, the sun or Mars to come to our modern beliefs. ... The factor we had thought to be the most salient similarity to the cases grouped under the term, being a heavenly body visible in the night sky for 'star', living in water for 'fish', turned out not to be the most important similarity between most things grouped under that term. So there is an important sense in which saying whales are fish, or that the sun is not a star, may reveal a conceptual (rather than an empirical) error. (Weatherson [2003], 6)

Given this possibility that differing background beliefs may be a source of error in conceptual intuition, the transformation model can explain differing Gettier intuitions by suggesting that one – or both – of the groups with differing intuitions are simply in error.

Note also that a simple review of the history of philosophy would have provided at least one example of a highly sophisticated philosophical school that rejected the Gettier intuition so dominant in late 20th century analytic epistemology: Gangesa’s (12th – 13th century C.E.) Natya-Nyaya school. Thus, e.g., B. Matilal discusses a case in which Gangesa considers an accidentally true belief and explicitly identifies it as an instance of knowledge:

Gangesa speaks, interestingly enough, of the case of a misinformed liar where
what the speaker says (although he says it to deceive his audience) just happens to be true. The question is: since the hearer has a true belief generated by an utterance in this case, are we entitled to call it knowledge? Assuming that the hearer does not have any prior idea that the speaker is a compulsive liar, Gangesa would claim that the hearer would have knowledge from the utterance in this case. (Matilal [1990], p. 70)

Interestingly, this suggests that one needn’t have appealed to experimental philosophy to discover that there is disagreement—even among communities of sophisticated, highly trained philosophers—as to the interpretation of thought experiments.

Not only does this case suggest that experimental philosophy hasn’t, in this instance, told us anything that we didn’t already know about the possibility of disagreement—even about such widely agreed upon cases such as the Gettier cases. Furthermore, the example of Gangesa suggests that disagreement—even about such core cases—does not in fact invalidate the use of thought experiment in philosophy. Indeed, as J. Ganeri [1999] discusses in the case of Gangesa’s rejection of Gettier-cases as examples of non-knowledge, other philosophical schools at the time vehemently disagreed with the Navya-Nyaya position; they did so by appealing to higher-order considerations, explanatory power, etc.,—precisely the sorts of argumentative strategies that one would employ when an appeal to intuition fails to resolve a disagreement. (Cf. Ganeri [1999], pp. 77-8)

6. Implications for Experimental Analysis

Similar implications hold for the case of experimental analysis as well. Experimental analysts share restrictionists’ conception of traditional “armchair” philosophy as committed to the use of intuitions as evidence. As noted in section 2, above, where experimental analysts differ is in their evaluation of the evidentiary value of intuitions. Where the two schools also coincide, however, is in their acceptance of the fact that the disagreement between the intuitions of philosophers and those of their folk counterparts does pose a problem for traditional philosophy; whereas experimental restrictionists see this fact as proof of the disvalue of intuitions as evidence, experimental analysts instead argue for the -- at least potentially -- greater value of folk intuition vis-a-vis the intuitions of philosophers.

Thus, Jesse Prinz [2008] would seem to have experimental analysis in mind when he suggests that “experimental philosophy uses the traditional philosophical
method of reporting intuitions together with statistical analysis to criticize philosophers’ claims about *what the authoritative intuitions should be.*” (Prinz [2008], 199; my italics) The central point of dispute between experimental analysis and armchair philosophy, according to this reading, isn’t whether intuitions may serve as evidence, but *which intuitions* -- or *whose* intuitions -- should serve as evidence.

That is, experimental analysts seem to argue as follows:

iv. Philosophers and the folk sometimes disagree with respect to their intuitive judgments as measured by their responses to thought experiments.

v. If philosophers and the folk sometimes disagree with respect to their intuitive judgments as measured by their responses to thought experiments, we ought to appeal to folk intuitive judgments, as measured by their responses to thought experiments, as a corrective to philosophers’ judgments.

vi. Thus, we ought to study folk intuitive judgments, as measured by their responses to thought experiments, as an -- at least potential -- corrective to philosophers’ intuitions.

Note again, however, the change that results when one reconceives the role of thought-experiments as suggested above. The problem, clearly, is with premise v: if in fact thought experiments involve selections from, reconceptions of, or implications from one’s existing beliefs, then the fact that others would arrive at other beliefs when selecting from, reconceiving, or deriving implications from *their* beliefs neither suggests that one’s own results are unreliable, nor can it, without further knowledge regarding those others’ existing beliefs, serve as a corrective to one’s own results. As in the discussion of experimental restrictionism, the comparison with the strategy of achieving maximal consistency within one’s own set of beliefs is instructive. The fact that others arrive at a different set of beliefs when following the strategy of pursuing maximal consistency does not impugn that strategy, nor does it suggest that one’s own employment of the strategy requires correcting.

### 7. Implications for Experimental Descriptivism

Many experimental philosophers perceive a need to remedy the traditional philosopher’s reliance on intuition -- either by rejecting intuitions altogether, in the case of the experimental restrictivist, or by using folk intuitions as revealed in surveys
as a corrective to the armchair intuitions of the traditional philosopher, in the case of the experimental analyst. I have argued that, in both of these cases, the experimental philosopher has as yet been unsuccessful in establishing that the way that thought experiments figure in philosophical argument requires correction in the first place. When one appreciates the role of thought experiment in philosophical argument, or so I have argued, one recognizes that the disagreements between traditional philosophers and survey subjects to which experimental restrictivists and experimental analysts point as evidence against the traditional philosopher’s use of thought experiment are in fact not appropriate evidence at all.

It is important to make it clear that this is not to say that no experimental evidence might bear on the traditional philosopher’s use of thought experiment. Thus, e.g., I have suggested that thought experiments are best understood as analogous to forms of -- deductive, inductive, or abductive -- reasoning. But cognitive and social psychologists have long been studying the ways in which one’s affect, health, diet, social situation, etc., may bear on one’s reliability in employing various forms of reasoning. Given this, however, it would be wise for philosophers to apply the lessons from those fields to their employment of thought experiments, relying only on those instances of the consideration of thought experiment that have been conducted under appropriate conditions.

Furthermore, if thought experiments are forms of reasoning, then studying subjects’ responses to thought experiments would aid in studying how it is that subjects reason, a topic of obvious interest to philosophers. Indeed, it should be little wonder that the use of surveys to study subjects’ intuitive judgments has long been recognized by experimental psychologists as a valuable means of exploring the ways in which subjects reason.

Thus, as the example of Haidt’s research discussed in illustration of the homuncular model of thought experiment demonstrates, the use of surveys of intuitive judgments to thought experiments may provide valuable evidence bearing on the nature of human thought in a wide variety of domains. To say this, however, is simply to grant the point recently made by Knobe and Nichols [2008], “that the patterns to be found in people’s intuitions point to important truths about how the mind works.” (Knobe and Nichols [2008], 12) However, granting this point would seem to involve granting the value of experimental descriptivism, which, in Nadelhoffer and Nahmias’s [2007] definition, involves striving “to better understand the nature of the underlying psychological processes and cognitive mechanisms that produce our intuitions and explore the relevance of this research to philosophical
questions.” (Nadelhoffer and Nahmias [2007], 127) Indeed, though Haidt himself is a social psychologist, it is clear that his work—and similar work conducted by other cognitive and experimental psychologists—would fit equally well within the field of experimental descriptivism, as characterized in section 2, above.

Given this, then, there are at least two ways in which the work of experimental descriptivists might be of value to traditional philosophers. First, as noted above, the results of experimental descriptivist investigations might shed light on the epistemology of philosophy, helping us better to understand ways in which our use of thought experiment in deliberation might be affected by situational effects—thereby aiding us in better weighing the outcomes of those deliberations. Second, as in the example of Haidt’s research, experimental descriptivist work can illuminate the workings of human reasoning more generally, thus yielding insights into human cognition that could be relevant to a wide swath of programs within traditional philosophy.

On balance, the present discussion suggests a mixed result for experimental philosophy. If thought experiments have the epistemic role that the discussion here has indicated, then one would have good reason to doubt the claims made by experimental analysts and experimental restrictionists that traditional philosophers ought to reject or, at the very least, correct their intuitive judgments in light of the research conducted by experimental philosophers and social psychologists. However, as has become clear, the prospect for experimental descriptivism seems far brighter. Though philosophers needn’t perhaps abandon their armchairs, it probably wouldn’t hurt them to get up once in a while to stretch their legs, either.

Sources:


