

Philosophical Methodology and its Implications for Experimental Philosophy

by

Benjamin A. Keil

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy

---

Chairperson Ben Eggleston

---

John Bricke

---

Erin Frykholm

---

Dan Hirmas

---

Tom Tuozzo

Date Defended: April 27, 2015

The Dissertation Committee for Benjamin A. Keil certifies that this is the approved version of the following dissertation:

Philosophical Methodology and its Implications for Experimental Philosophy

---

Chairperson Ben Eggleston

Date approved: April 27, 2015

## Abstract

Since most philosophers accord *some* role to intuitions in the practice of philosophy, my dissertation's first paper addresses an important subsidiary question: *Whose* intuitions should be allowed to play a role in proper philosophical practice? My paper critiques Steven Hales' view when he argues that the intuitions of philosophical laymen lack philosophical significance. I rebut the main arguments he gives in support of the "expertise defense" and then provide an Aristotelian-style argument in favor of the significance of lay philosophical intuitions.

My second paper extends John Norton's work into the realm of experimental philosophy. Norton argues that scientific thought experiments are arguments; I develop his work and show that his position entails that philosophical thought experiments are also arguments. I consider Thomson's "Trolley Problem" and demonstrate that it contains an implicit argument, even if the argument's conclusion is often omitted when presented in a classroom setting. Since my position entails that philosophical thought experiments are non-neutral devices for eliciting intuitions, I suggest two key implications for practitioners of experimental philosophy.

Conflicting ethical intuitions are nothing new to philosophers, but a new way of resolving some of those intuitional conflicts is my third paper's topic. AJ Ayer famously argues that unless some criterion for deciding between conflicting intuitions exists, appeals to intuition are worthless. I partially answer Ayer's challenge by drawing on Steven Hales' defense of foundationalism. Hales argues that at least one self-justifying proposition exists. If true, and if one self-justifying *ethical* proposition exists, I argue that this provides us a partial way towards answering Ayer's challenge. Since self-justifying propositions must be justified *a priori*, where a

conflict exists between an ethical intuition ultimately justified *a priori* and another ethical intuition ultimately justified *a posteriori*, the latter intuition should be rejected.

## Acknowledgments

To list all my debts (both intellectual and personal) which resulted in this dissertation's completion would be a hopeless task. But first among those on such a list would be my wife Elizabeth. Without her patience, love, and support I could not have completed graduate school. My children James, Joy, Gabriel, and Hope helped remind me that a dissertation is not the most important component of a good life. My parents Steve and Carole Keil provided the formative home-educational environment which enabled me to arrive at this educational summit – I could not be where I am without their significant efforts. Concerning my academic mentors, I am proud to recognize the roles played by Drs. Ben Eggleston, James Stephens, and Donald Turner. Whatever academic success I accrue is, I hope, an honor to their efforts spent improving my philosophical abilities. Finally, I wish to thank my friends and officemates Brian Clarke and Andrew McFarland. Their conversations throughout the years of graduate school have made me a better philosopher and a wiser person – I could not ask for better academic companions than them. *Virtus tentamine gaudet.*

## Table of Contents

<b>Experimental Philosophy and Lay Intuitions</b> .....	1
<b>I: Preliminary Observations</b> .....	2
<b>II: The Debate in the Literature</b> .....	3
<b>III: Hales’ Argument &amp; Response</b> .....	6
<b>IV: An Aristotelian-style Argument in Favor of Lay Intuitions</b> .....	13
<b>V: Objections and Replies</b> .....	19
<b>Works Cited</b> .....	24
<b>Philosophical Thought Experiments are Arguments – and What This Means for Experimental Philosophy</b> .....	25
<b>I: Scientific Thought Experiments are Arguments</b> .....	26
<b>II: Philosophical Thought Experiments are like Scientific Thought Experiments</b> .....	30
<b>III: Demonstration</b> .....	37
<b>IV: Replies to Objections</b> .....	40
<b>V: Implications</b> .....	46
<b>Works Cited</b> .....	49
<b>Sorting Through Our Ethical Intuitions</b> .....	50
<b>I: The Problem</b> .....	51
<b>II: Hales on Foundationalism</b> .....	52
<b>III: Answering Ayer</b> .....	55
<b>IV: Objection Concerning First Principles</b> .....	65
<b>V: Objection Concerning Reflective Equilibrium</b> .....	67
<b>Works Cited</b> .....	73

## Experimental Philosophy and Lay Intuitions

I begin with the assumption that intuitions have some role to play in the practice of philosophy. This, obviously, is not wholly uncontroversial. Still, it is an assumption shared by a sufficient number of philosophers that it may function as this paper's "first principle". But it immediately suggests two further queries: First, *what role* should intuitions play in philosophical practice and, second, *whose* intuitions should play that role?

Although both topics are important, this paper shall only address the second: *Whose* intuitions have a role to play in the practice of philosophy? Some philosophers have argued that the intuitions of philosophical experts should play little (or no) role in our discipline's practice. In reply, others have gone to the opposite extreme and argued that *only* the intuitions of philosophical experts have a role to play; they thereby significantly minimize or exclude entirely the intuitions of philosophical laymen. In contrast, I will defend the middle ground between these two extremes: I shall argue that the intuitions of *both* philosophical experts *and* philosophical laymen have an important role to play in the practice of philosophy, although it may well be the case that the experts' intuitions are ultimately more valuable.

To demonstrate this thesis my arguments will be as follows. First, I shall examine some preliminary matters. Second, I will situate this paper within an ongoing philosophical debate concerning the topic. Third, I will consider the strongest argument for excluding lay intuitions and demonstrate its falsity. Fourth, I will develop a positive Aristotelian-style argument in favor of including lay intuitions. Fifth and finally, I shall consider and refute certain objections against the positions I develop.

## **I: Preliminary Observations**

I begin with two preliminary concerns. The first involves my paper's scope and the second involves the purpose of experimental philosophy surveys.

My first preliminary concern is a limitation of this paper's scope: I shall explicitly limit my work to methods of experimental philosophy involving surveys. For example, a researcher might conduct surveys dealing with areas of ethical inquiry (such as trying to discover which philosophical arguments against abortion are viewed as being strongest) or areas within the history of philosophy (which interpretation of a particular Platonic claim is most compelling), etc. Of course, while there are no doubt other kinds of experimental philosophy not necessitating the use of surveys, my points shall not directly address any of those. I do not argue that such methods are not proper forms of experimental philosophy, then, but rather that I am only concerned with defending lay intuitions with regard to experimental philosophy surveys.

The second and final preliminary concern deals with the purpose of experimental philosophy surveys. Obviously, one cannot address the relevance of lay philosophical intuitions unless the purposes of experimental philosophy surveys themselves are known. There are at least two distinct purposes for which experimental philosophy surveys may be administered.

The first purpose is *informative* with regard to discovering what intuitions members of a given group have about a certain philosophical question. For example, Weinburg, Nichols, and Stich's 2001 survey was administered for purposes of trying to determine what various peoples believe about Gettier cases. Of great interest, they discovered that research subjects from East Asian and Indian nations displayed intuitions divergent from those of Western subjects, even when controlling for various socioeconomic criteria (Weinburg et al. 448).

For informative research the relevancy of laymen is entirely determined by the groups which the researchers wish to study. If researchers wish to study laymen's intuitions about a particular philosophical thought experiment, then of course laymen's opinions will be philosophically relevant. Contrariwise, if researchers wish to study experts' intuitions about Gettier cases, then surveying laymen would not be philosophically relevant. The relevancy of philosophical laymen's intuitions to these surveys is thus contingent upon the information being sought - if the intuitions in question can be provided by laymen, then they ought to be included in the survey group and if not then they ought to be excluded.

The second purpose of administering surveys is *truth-ascertaining*. Surveys of this sort are designed to help the truth be better understood than it would be if no survey were conducted. For example, one might present two opposing courses of action and gather the respondents' intuitions concerning which one is ethical and why. This could not only help shed light on the ethical course of action and its justification but also could implicate which, if any, ethical theory is truest.

For present purposes I am concerned only with those surveys administered for purposes of ascertaining the truth concerning some or other philosophical question. And it is on precisely these grounds that the relevance of lay philosophical intuitions is challenged. Let us now briefly consider the present state of philosophical literature on the topic.

## **II: The Debate in the Literature**

This paper contributes to a heated philosophical debate concerning the role of intuitions in philosophy. Recent experimental philosophy work has drawn criticism from some for using

intuitions at all; from others, it has been criticized for its substantial reliance upon laymen's philosophical intuitions.

To briefly summarize this debate's status, there are some philosophers who believe that intuitions themselves have no (or very little) role to play in the practice of philosophy. This opposition to intuitions is largely absolute (rather than opposing intuitions on some contingent empirical grounds). A standard defense of this view is offered in Earlenbaugh and Molyneux (2009). To oversimplify they argue in favor of a sharp distinction between *perceptions* and *intuitions* and conclude that, while perceptions can serve as philosophical evidence, intuitions cannot.

This paper will assume (without arguing) that this wholesale opposition to intuitions in philosophy is wrong – intuitions can (and in fact do) serve as philosophical evidence.<sup>1</sup> Given this (widely shared) assumption, there then exists a secondary dispute concerning *whose* intuitions may play an evidentiary role.

Weinberg *et al.* (2010) argue that the intuitions of philosophical experts should not play such a role. They understand the greater trustworthiness of expert philosophical intuitions (over against lay intuitions) as an empirical claim. So after canvassing the psychological literature on expertise, they argue that the empirical claim concerning expert trustworthiness has been falsified. Consequently, they believe that the intuitions of philosophical experts should not be given much (if any) weight.

(Somewhat confusingly Jonathan Weinberg also coauthored a 2001 paper arguing that *epistemic* intuitions lack normative import [Weinberg *et al.* 2001 434], and it is unclear why, in

---

<sup>1</sup> The best response to Earlenbaugh and Molyneux's arguments can be found in Hales 2012 193-196.

principle, the objections could not be expanded to other non-epistemic philosophical intuitions as well. This 2001 paper thus expresses something closer to an absolute rejection of philosophical intuitions, albeit on different grounds than those given by Earlenbaugh and Molyneux. However in the 2010 paper Weinberg and his coauthors oppose expert philosophical intuitions on contingent empirical grounds drawn from the psychological literature on expertise. Put another way, *if* philosophical experts' intuitions were shown to be empirically trustworthy, Weinberg *et al.* 2010 would be committed to allowing them a greater evidentiary role in philosophical practice, unlike Earlenbaugh and Molyneux's opposition to expert intuitions.)

In response to this and similar contingent challenges against philosophical expertise, the "expertise defense" was originally developed in Williamson 2007 and, after Weinberg *et al.*'s later 2010 reply, Williamson developed it further in his 2011 article. The expertise defense, in brief, defends the intuitions of philosophical experts on grounds that their education has significantly improved relevant skills (such as careful attention to details and understanding their relevance to the situation) (Williamson 2007 191). Williamson thus defends what I take to be the majority view: That intuitions have some evidentiary role to play in philosophy and, at minimum, the intuitions of philosophical experts can play this role.

Williamson, however, never squarely addresses what role (if any) the intuitions of laymen have. This is not surprising given that the purpose of his 2011 article is to defend expert intuitions rather than addressing the role of lay intuitions. In his earlier 2007 book Williamson only addresses the topic in the briefest of terms: "Although the philosophically innocent may be free of various forms of theoretical bias, just as the scientifically innocent are, that is not enough to confer special authority on innocent judgment, given its characteristic sloppiness" (Williamson 2007 191).

Promisingly Williamson notes that philosophical laymen can be free of some biases possessed by experts – thus potentially providing space for some role to be played by lay philosophical intuitions. But it is unclear how Williamson intends his larger point to be understood: He might mean only that the “characteristic sloppiness” of lay intuitions does not confer *special* authority on lay philosophical judgments. But he could also mean that their characteristic sloppiness removes *any* authority from lay intuitions. Given that he speaks on the topic in only the briefest of terms, and tangentially at that, I hesitate to draw strong conclusions concerning his views of lay intuitions. It should suffice to say that if Williamson’s view is closer to the former position that I describe then there exists a significant amount of overlap between his views and my own. (As shall be shown later, I do not argue that freedom from theoretical biases confers special authority on lay intuitions and I recognize that sometimes lay intuitions can be “sloppy”.)

Although Williamson’s views concerning the role of lay intuitions in philosophy are ambiguous, other philosophers have addressed the topic specifically. Hales has affirmatively argued that intuitions of philosophical laymen ought to be excluded from the practice of philosophy. Since his arguments in favor of this position are the best on offer, my second section will consider and then refute such a view.

### **III: Hales’ Argument & Response**

I will begin with the arguments contained in Hales’ 2006 book *Relativism and the Foundations of Philosophy*, taking notice where appropriate of his later argumentative developments. To be clear, Hales’ initial 2006 arguments were not expressly applied to experimental philosophy (he initially addressed only whether or not laymen’s intuitions can help resolve thought experiments). But as clarified in his 2012 article his arguments unambiguously

apply to all experimental philosophy surveys involving lay intuitions. (Experimental philosophy surveys of this sort no doubt comprise the vast majority of work done in experimental philosophy at present.)

The relevant section of Hales' book was written in response to Weinberg, Nichols, and Stich's controversial 2001 paper demonstrating that there exists significant variance in the epistemic intuitions in Gettier cases among members of different socioeconomic groups and cultures. By itself, this intuitional variance would be noteworthy but need not be understood as a significant challenge to the justification of rational intuition. Weinberg *et al.*, however, use their data to undercut the justification for rational intuition itself. They write:

It may well be that upper middle class Westerners who have had a few years of graduate training in analytic philosophy do indeed all have strong, modality-linked intuitions about Gettier cases. But since most of the world's population apparently does not share these intuitions, it is hard to see why we should think that these intuitions tell us anything at all about the modal structure of reality, or about epistemic norms or *indeed about anything else of philosophical interest* (Weinberg *et al.* 2001 452, emphasis added).

In response to this radical claim that intuitions tell us *nothing* of philosophical interest, Hales is understandably enthusiastic to provide a defense of intuition's role in philosophy. Thus he addressed the topic in his 2006 book and subsequent 2012 article.

Hales ultimately defends intuitions by distinguishing between lay and expert intuitions and arguing that the latter possess a significance lacked by the former. In his 2006 book he begins by noting that "...[N]ot all intuitions are created equal" and that "Intuitions are and should be sensitive to education and training in the relevant domain" (Hales 171). He ultimately vindicates expert intuitions via an analogy between the sciences and philosophy: Just as the scientific intuitions of scientific laymen are not significant but the scientific intuitions of

scientific experts are, similarly the intuitions of philosophical experts are significant in a way that the intuitions of philosophical laymen are not. Hales writes:

For example, the physical intuitions of professional scientists are much more trustworthy than those of undergraduates or random persons in a bus station. Scientists have and rely on physical intuitions, intuitions that are trained, educated, and informed and yet are good indicators of truth for those very reasons. In the same way, the modal intuitions of professional philosophers are much more reliable than either those of inexperienced students or “the folk” (Hales 171).

Hales here suggests that the intuitions of professional scientists (and philosophers) are good indicators of truth inasmuch as professional scientists (and philosophers) possess greater expertise – that is, greater training, education, and information concerning their respective domains.

But Hales’ position is muddled somewhat in his 2012 article where he writes: “I mean ‘expertise’ in an internalist sense, one that does not require any connection at all to gaining the truth. In this sense, Joshua Bell is an expert violinist and physicians are experts about the human body even if we are all brains in vats” (Hales 2012 190). What is one to make of the assertion that expert intuitions are “good indicators of truth” (Hales 171) but yet that expertise “does not require any connection at all to gaining the truth” (Hales 2012 190)? One possibility would be that Hales has simply changed his views between these two publications. But another more consonant interpretation exists: That while expert intuitions are good *indicators* of truth, expertise does not *require* a connection to the truth. This interpretation is bolstered by Hales’ claim that physicians are still experts on the human body even if global skepticism were to be true – so while expertise is a good indicator of truth (if anything is), it does not require a connection to the truth (just in case a global skeptical scenario were true).

Now that we understand Hales' defense of intuitions, we must turn to my arguments against it. To be clear, Hales' arguments are not problematic insofar as they defend the role of philosophical experts' intuitions for the practice of philosophy – both he and I agree that they have a significant role to play. The problematic portion of his defense is its derogatory implication concerning the value of *non-expert* intuitions. For example, Hales writes that some have held "...that it is the expert intuitions of professionally trained philosophers that have epistemic merit, not the uninformed reactions of the unwashed masses" (Hales 2012 199). (Hales does not intend this quotation merely to demonstrate that *others* have held this view but to note that he himself holds the view. Thus, in a footnote immediately following the previous quotation, Hales writes "I made this argument" with a citation to his 2006 book [Hales 2012 199.]

Note the strong nature of his conclusion: He not only argues that the intuitions of philosophical experts have significant value but also alleges that the intuitions of philosophical laymen are "uninformed" (while disdainfully referring to philosophical laymen as "the unwashed masses"). It is my present purpose to argue that Hales' arguments do not justify his strong exclusion of lay intuitions (and in the paper's next section I will argue that we have good reason to include them). There are two grounds on which Hales' argument fails. First, Hales' arguments for excluding lay intuitions are unsuccessful. Second, discarding lay intuitions abandons an important check for philosophical truth. These shall be addressed in turn.

In order for Hales to show that expert intuitions are significantly valuable while lay intuitions possess little (or no) value, Hales must argue that there exist some properties  $p$ ,  $q$ , and  $r$  which are possessed by expert philosophical intuitions but lacked in lay philosophical

intuitions. These properties, then, are what render expert philosophical intuitions significantly more valuable than lay intuitions.

As previously cited, Hales believes that expert intuitions are more valuable in virtue of their greater training, education, and information. Thus, implicatively, he must hold that the “uninformed reactions” of the “unwashed masses” lack training, education, and information in the relevant senses. (If he did not so hold, then those same properties which make expert intuitions valuable would also make lay intuitions valuable.) But do philosophical laymen in fact lack training, education, and information in the relevant sense?

I concede the (near tautological) claim that philosophical laymen often lack much philosophical training. But if education and information refer to the education and information necessary to have justified intuitions concerning the thought experiment, it might well be the case that philosophical laymen possess enough education and information for their intuitions to be worthy of consideration. And if so, then by Hales’ own arguments lay intuitions should be included.

For example, if an experimental philosophy survey were to ask which argument against a coherentist epistemic system were the most successful, it is safe to assume that few laymen possess sufficient education or information to even have an intuition on that topic. But experimental philosophy surveys can and often do concern rather more pedestrian philosophical matters (such as whether promises made to the dying ought to be kept or whether beauty is objective or subjective). These types of questions either do not require much education and information in order for one to have a justified intuition, or the amount of education and information they do require is widely dispersed among philosophical non-experts as well as

experts. Hales' own argument, then, does not justify discarding all lay philosophical intuitions on grounds that they are merely "uninformed reactions" – if anything, his arguments might well *require* the inclusion of lay intuitions.

In his later 2012 article Hales seems to modify his position somewhat: The greater training, education, and information philosophical experts possess results in philosophical expertise, and this philosophical expertise "...is partly manifested in the considerable agreement about philosophical cases that one does not find among nonprofessionals" (Hales 2012 190). Hales thereby places some emphasis not only on the philosophical training, education, and information itself but also the fact that it results in considerable *agreement* among professionals in contrast to the supposed disagreement one finds among nonexperts.

However if considerable philosophical agreement is a partial manifestation of philosophical expertise, then there is no reason that philosophical laymen could not also exhibit a considerable degree of philosophical agreement as well – and if they did, we would have good reason to include their intuitions in our philosophical analysis.

The only presently-available evidence provides some support for my contention. Thus far only one published study on intuitional variance between philosophical laymen and experts exists and, while it showed some divergence between expert and lay philosophical intuitions, it also showed a considerable amount of lay intuitional agreement on the topic being studied.

Although this might be obvious to some, it is nonetheless worth providing the following *caveat*: At present the data are insufficient to draw any strong conclusion as to whether or not there exists widespread philosophical agreement among non-experts. (This particular study suffers from the confounding factor that philosophical expertise was operationally defined as at

least the pursuit of an undergraduate degree in philosophy – so it is likely that Hales would not even regard the “experts” picked out by this study as experts in the sense relevant to his argument).

Nonetheless, Sytsma and Machery surveyed philosophical experts and laymen concerning a phenomenological question. They found that laymen (but not philosophical experts) were “...willing to ascribe the perceptual state of seeing red to a simple robot” (Sytsma and Machery 13). Significantly, the standard deviation for nonexperts’ responses was *lower* than for philosophers in this scenario. This indicates that the nonexperts’ responses were *closer* to the mean (*i.e.*, closer to each other) than the responses of the philosophical experts were (Sytsma and Machery 12). So, to oversimplify, inasmuch as philosophical agreement manifested in this study, it manifested to a greater degree among nonexperts than among the philosophical experts.

Although the data state is highly preliminary, there is thus at least some very initial reason to doubt the attribution of widespread philosophical disagreement among lay philosophical intuitions. But since Hales excluded lay philosophical intuitions based on their lack of training, education, information, and inter-group agreement, and since I have shown that philosophical laymen’s intuitions can well be grounded in education and information while exhibiting inter-group agreement, Hales’ argument justifying the exclusion of lay intuitions is fatally undercut.

Second and finally, if laymen’s intuitions were disregarded as Hales suggests, one potential check for truthfulness will be discarded. To begin, remember that it is always worthwhile to ask oneself the following question: “If I were wrong, how would I know?” If we

extend this question to the present situation, we might wonder how we would know if the philosophical experts' intuitions were wrong.

Let us consider the analogy to the sciences which Hales so frequently draws upon. In the natural sciences, the intuitions are checked not only against the intuitions of other experts (as is also the case in philosophy) but, ultimately, by the data themselves. After all, a scientific theory is ultimately not answerable to experts and their scientific intuitions but to whether or not the theory explains and/or predicts the data. ("And yet it moves..." Galileo is apocryphally reported to have said of the Earth, even when the reigning Aristotelian cosmology predicted its motionlessness.)

Philosophy, however is dissimilar from the natural sciences in that expert philosophical intuitions are not ultimately answerable to the data – and it is precisely this fact which enables one man's *modus ponens* to be another man's *modus tollens*. So if laymen's intuitions were excluded in the manner Hales wishes them to be, this would entail the exclusion of one potential check on expert intuitions – and thus entail one fewer potential way for us to know if the philosophical experts were wrong. Because Hales' analogy is flawed, and because we do not wish to exclude an additional check against philosophical expert error, we should reject Hales' argument entailing exclusion of lay intuitions.

#### **IV: An Aristotelian-style Argument in Favor of Lay Intuitions**

I have now demonstrated that Hales' arguments against including lay intuitions in experimental philosophy surveys are unsuccessful. Obviously, however, this does not by itself suffice to demonstrate that the intuitions of philosophical laymen ought to be included – as I think they should. It is to that task that this section will be devoted. I shall develop an

Aristotelian-style argument and its conclusion will show that the intuitions of philosophical laymen ought to be included in the experimental philosophy survey process.

Before doing so, however, a brief preliminary note is in order. I want to emphasize that this is an Aristotelian-*style* argument; it is, I believe, broadly consonant with Aristotelian philosophy (as well as the beliefs of other philosophical systems), but I do not here assert that my arguments strictly describe or encompass all of what Aristotle actually believed. My goal will be satisfied if a good argument from broadly (but not exclusively) Aristotelian premises is given in favor of including lay intuitions in experimental philosophy work.

I begin with the premise that humans (*homo sapiens*) are, among their other properties, *rational* creatures. As such, humans are capable of questioning, seeking knowledge, and coming to understanding in a manner beyond the capacities of non-rational creatures such as birds. Obviously this rational faculty may be damaged or otherwise defective in some particular humans, but this no more disproves that humans are rational than other damages or defects disprove that humans are bipedal.

(Although I have phrased this premise in fairly Aristotelian terms, the belief that humans are rational is not an assumption unique to Aristotelians. Many other non-Aristotelian philosophers would accept this premise. Indeed the set of *philosophers* who believe that human rationality either has not been established or has been conclusively disproven is, I imagine, quite small.)

Let us follow, then, an entailment of our assumption that humans are rational. Inasmuch as any creature is rational, this entails that it is oriented towards (and apt for) knowing the truth. In other words, humans both desire and are capable of coming to know the truth. This does not

entail, obviously, that we humans always *will* find the truth (particularly in our all-too-short lifetimes). But our ordinary human faculties (when combined with standard truth-gathering processes such as perception, experience, and experimentation) are adequate to know the truth, or at minimum to not deviate too far from the truth (Kraut 78).

If by virtue of our rationality humans are apt for knowing the truth, this provides us with good reason to expect that much of what is believed is actually true, or at least contains a significant element of the truth in it. But among that which is believed, what beliefs should receive preference?

One possibility would be to consider every view on some topic *t* that has ever been held. But this poses practical as well as theoretical problems. Practically it would be impossible to catalogue all the beliefs that all people have had about *t* in any reasonable amount of time. Theoretically, some number of beliefs about *t* will have been held without adequate justification, so those beliefs ought not be considered in the first place. (Thus Aristotle, perhaps wryly, notes that “To examine all the opinions that have been held [on the given topic] were perhaps fruitless” [Aristotle NE 1095a28-9, trans. McKeon.]

Another possibility is to prefer all the *uncontroversial* views that have been held concerning *t*. But a given view’s uncontroversiality with respect to *t* cannot be known without a complete catalogue of beliefs concerning *t* being made first (which raises the same practical difficulties discussed in the previous paragraph). Moreover there is no antecedent reason to believe (and plenty of empirical reason to deny) that the uncontroversiality of a view bears any necessary relationship to its truth. Indeed (as the history of science thoroughly bears out) restricting our heuristic to consider *only* uncontroversial beliefs would result in our having

significantly fewer true beliefs than we presently do! Thus, considering all views about *t* or only the uncontroversial ones are suboptimal.

But if those are less preferable ways to go about acquiring true beliefs, what would be better? Aristotle recommends that we consult the *endoxa* (ordinarily translated as the “reputable beliefs” or “reputable opinions”) on a topic in order to help discover the truth concerning it. More specifically Aristotle writes that the *endoxa* are the opinions accepted by every person, or by the majority of persons, or by wise persons (among which they may be accepted by all the wise persons, or by the majority of the wise persons, or by the most notable and illustrious persons among the wise) (Aristotle Top. I.1.100b21-3, trans. McKeon).

But why consult the *endoxa* at all concerning a given topic? They should be consulted because, as previously established, ordinary human faculties are adequate for knowing the truth, and if human rationality is adequate to know the truth (or not deviate far from it), this provides us with good reason to expect that some or all of what is reputedly believed is true (Cooper 288-289). As a heuristic, then, we should consult the *endoxa* because those beliefs are likely to be true, or to contain more truth than we would ordinarily find by consulting other sources – and that is the purpose of a truth-acquiring heuristic.

Although phrased in Aristotelian terms, much of the forgoing would be acceptable to non-Aristotelians. It is not significantly controversial that, in order to discover the truth, reputable beliefs ought to be consulted. But if a good way to acquire truth is to consult reputable beliefs, we must then decide *whose* opinions are reputable.

Aristotle classifies two groups as having reputable opinions: the wise and the many (sometimes translated as “the masses”). If Aristotle is correct that both the opinions of the wise

and the many constitute reputable beliefs, then experimental philosophers will be justified in collecting the intuitions of both philosophical experts (the wise) and the intuitions of philosophical laymen (the many). Cooper explains Aristotle's beliefs thusly:

If the *endoxa* are 'beliefs we hold' this has to be taken to refer, not exclusively to the beliefs we ordinarily hold, but to those together with whatever the wise may have to say: in other words, the 'we' means the whole lot of us, ordinary people with their opinions, and the 'wise,' or the relevant experts, with their sometimes, to the ordinary person, surprising or even outlandish opinions (Cooper 285).

It is important to note, then, that *endoxa* are comprised of what both ordinary persons say and what experts say about a topic. The views of these two groups need not be compatible with each other and, obviously, *endoxa* are not automatically true just by virtue of their being *endoxa*. But Aristotle, at least, holds that the beliefs of both the wise and the many can constitute "reputable opinions" that ought to be consulted.

Is he right to say this? I think it uncontroversial to say that the opinions of the wise (that is, the relevant experts) constitute reputable opinions that ought to be consulted when attempting to discern the truth about a topic. But what of the opinions of the many? Should their opinions count as reputable? I argue that their opinions, if they have any, should indeed be included among the set of reputable opinions and thus should be included in good experimental philosophy survey work. But why? There are two reasons why the opinions of the many ought to be included among the set of reputable beliefs.

First, the many lack the willingness to defend their philosophical theories at whatever cost. Aristotle implicitly gestures at this when he considers and denies that great misfortunes are compatible with happiness. Aristotle writes "A man who was living [in that way] no one would

call happy, unless he were maintaining a thesis at all costs” (Aristotle NE I.5.1096a2, trans.

McKeon). Kraut, speaking of this passage, writes:

When arguing with each other, philosophers have been known to persist in defending, at great length, propositions that, to most people, lack all plausibility... There is, in other words, a danger that those who specialize in a subject will become so eager to win points over other specialists, or to achieve prominence, that they lose their ability to tell what is reasonable to believe. That is perhaps why Aristotle’s method requires a student of a subject to pay attention not only to what seems to be the case to specialists in a field but also to what seems to be the case to ordinary people... Philosophers *can* be right when they hold views that conflict with common opinion (that is why Aristotle’s method requires us to consult their opinions) but they can also be wrong. And so it is part of proper method to pay attention to the views of both specialists and non-specialists (Kraut 79-80).

What Kraut addresses here is surely common experience to all professional philosophers: As much as we wish it never occurred, it occasionally happens that philosophers (including myself) defend wildly implausible views due to considerations of professional prominence or point scoring rather than the truth of the matter. Of course, *just because* a view is implausible does not entail its falsity (as Kraut notes). But a good way to keep a vice common to professional philosophers in check is to consider the views of non-experts. In this way *both* the opinions of experts *and* of laymen (the many) can be reputable opinions worthy of consideration.

The second reason why the many’s opinions can be included among the set of reputable beliefs (that is, beliefs worthy of attention and consideration) is that the many, no less than the experts, are oriented towards knowing truth. As Hales points out they might lack some degree of training, education, or information. But as I have already argued, depending on the topic in question, the many need not lack education or information. So inasmuch as humans are oriented towards and capable of knowing the truth, the opinion of the many is reputable and thus worthy of consideration, albeit to an extent different than that of experts.

To speak more precisely, I am defending the view that the opinions of experts and laymen can both be reputable and that both should be consulted. This is to say that the views of both sets of persons are worthy of consideration when attempting to discern the truth. But I am not defending the thesis that the many's views are equally reputable to those of the experts in all circumstances. In some cases we should (and do) reject the many's opinion in favor of what the experts believe; other times we should (and do) reject the expert's opinion in favor of what the many believe. Unfortunately it is impossible to give a precise account of *when* the expert view should be rejected in favor of the many's view (or vice versa); such a choice is necessarily contextual and contingent upon a host of situational factors. (Inasmuch as I believe the situation resists specificity, it is importantly similar to Aristotelian virtue ethics.) Sometimes a wise person will accept the many's views, sometimes the wise person will accept the experts' view, but I do argue that the opinions of both groups are worthy of *consideration*, even if what one or both groups believe is ultimately rejected.

I have thus shown that both the views of the experts and the views of the many are worthy of consideration – that both, in essence, are constituent parts of the reputable beliefs about a topic. Thus we have good philosophical reason to conduct experimental philosophy work accordingly: We should seek to discover both the intuitions of philosophical experts as well as philosophical laymen when we are attempting to discover the truth concerning a philosophical topic.

## **V: Objections and Replies**

Before closing I would like to briefly consider and respond to two potential objections. The first concerns the nature of my reliance upon Aristotle and the second concerns potential circumstances of excluding lay intuitions. These shall be examined in turn.

Given my extensive reliance upon Aristotelian premises (as well as Aristotelian secondary sources), one might regard my entire argument as committing the informal fallacy of Appeal to Authority (in this case, an Appeal to Aristotle). Since arguments containing informal fallacies ought to be rejected, my argument's premises would be substantially undermined if this were the case.

However, no such fallacy occurs because I do not ask readers to accept any part of my argument on the basis that "Aristotle said so" (particularly since, as mentioned earlier, my argument is only intended to be broadly consonant with Aristotelianism in the first place). At each step where I appeal to Aristotle's views, I provide *arguments* showing why Aristotle was correct to think as he did. Moreover, although Aristotelians will find my premises plausible, so too will many other persons with non-Aristotelian philosophical inclinations. Thus, although my argument is broadly Aristotelian in its generation and consonance, it does not constitute a fallacious Appeal to Authority.

Second, and perhaps more significantly, what is one to make of cases where it seems like lay intuitions ought not to be consulted? Two such scenarios might arise. In the first kind, one might think that laymen lack adequate knowledge concerning the topic. For example, suppose one conducts a survey concerning whether or not the personalist criticism successfully refutes utilitarianism. One might reasonably think that laymen would have nothing to contribute to a question like this requiring a relatively high level of background knowledge.

It could well be the case that, given the inquiry's nature, laymen might have no beliefs on the subject altogether. This would particularly be the case when the amount of education and information necessary to comprehend the topic is relatively high. Nothing in my argument

suggests that the many must have an opinion on every question (regardless of how specialized the question is). So in cases such as this it is most accurate to regard the many as having no opinion on the topic and proceeding to examine the views of those who do have an opinion (typically, the experts). What we cannot do, I have argued *contra* Hales, is exclude laymen's intuitions altogether from good philosophical methodology. But neither must they be categorically included if they do not have (or are sufficiently unlikely to have) views on the topic under consideration.

A second kind of scenario involves an instance where the many have an opinion but it is an unreflective one. Aristotle himself seems to recognize the existence of such a scenario when he writes the following in the *Eudemian Ethics*:

It would be superfluous to examine all the opinions about happiness that find adherents. Many opinions are held by children and the diseased and mentally unbalanced, and no sensible man would concern himself with puzzles about them; the holders of such views are in need, not of arguments, but of maturity in which to change their opinions, or else of correction of a civil or medical kind... Similarly, neither need we examine the views of the many [about happiness]; they speak in an unreflective way on almost any topic, most of all when they speak about this; only the opinions of the wise – on this subject at least – should be examined; it would be strange to present argument to those who need not argument but experience (Aristotle EE I.3.1214b28-9, trans. Woods).

Certainly a detailed exegesis of particular Aristotelian texts is beyond this paper's scope inasmuch as my argument is only intended to be broadly consonant with Aristotelianism in the first place. However this passage is worth briefly lingering over since Aristotle writes here of an occasion when, seemingly, he thinks the many's views ought *not* be considered.

His ground for this conclusion is that the many "speak in an unreflective way on almost any topic, most of all when they speak about" happiness (that is, *eudaimonia*) (*ibid.*). It seems

that Aristotle believes that if the many's opinions are "unreflective", then they ought not be considered. But what is an "unreflective" opinion?

First, it is important to note that instead of "unreflective", other translators suggest that the opinions of the many be described as "random" (Aristotle EE I.3.1214b28-9, trans. Rackham). So understood, then a parallel passage in the *Nicomachean Ethics* could help illustrate Aristotle's overall meaning:

...[W]ith regard to what happiness is [the general run of men and people of superior refinement] differ, and the many do not give the same account as the wise. For the former think it is some plain and obvious thing, like pleasure, wealth, or honour; they differ, however, from one another – and often even the same man identifies it with different things, with health when he is ill, with wealth when he is poor... (Aristotle NE 1095a28-9, trans. McKeon.)

Here, it seems, Aristotle is addressing the same topic as he addresses in the *Eudemian Ethics* (the many's views of happiness) and suggests that the views of the many on the topic shift depending on the circumstances of their lives. Perhaps, then, their views are "unreflective" or "random" inasmuch as they change whenever life's circumstances change. So rather than suggesting an occasion where the many's views are not to be considered at all, instead this seems to be an occasion where the many's views are to be considered but ultimately discarded.

So construed, the view Aristotle expresses is consonant with my own. Neither Aristotle nor I believe that the many's views are always correct. In fact, a mark of a wise person is knowing precisely when to discard the views of the many (or the views of the wise) in order to obtain true beliefs. Just as Aristotle correctly discards the views of happiness coming from children and mentally ill persons, it seems likewise correct to discard views of happiness coming from "random" persons (that is, persons whose philosophical views shift radically depending on their personal circumstances).

With regard to experimental philosophy in particular, this suggests the importance of longitudinal research (which studies repeated observations of a variable over a period of time). If sound longitudinal studies verify Aristotle's empirical claim (that the many's opinions of happiness shift with regard to their personal circumstances but presumably the opinions of philosophical experts do not), then this could indeed be a good reason to discard the many's views on this topic. Certainly as experimental philosophy matures as a discipline the importance of longitudinal studies will grow for precisely this reason. But both Aristotle and I agree on this fundamental point: While the opinions of the masses and experts should both be considered, sometimes the views of one or both groups ought to be rejected – and if the many's views on happiness are “random” in the relevant sense, discarding their views can constitute responsible philosophical practice.

In conclusion, I have shown that Hales' reasons for excluding lay intuitions are faulty. Further, I have given a positive Aristotelian-style defense with regard to why lay intuitions ought to ordinarily be included in experimental philosophy research. If successful, then, I have not only rebutted the presumption against lay inclusion based on Hales' arguments but have also shown positive reasons *for* the inclusion of lay intuitions. The next steps fall on either philosophers (who may refute my arguments) or on researchers (who should continue or expand their inclusion of philosophical laymen in survey-based experimental philosophy research).

## Works Cited

- Aristote. *Aristotle in 23 Volumes*. Trans. Harris Rackham. Vol. 20. Cambridge (Mass.): Harvard UP, 1981. Print.
- Aristotle. "Ethica Nicomachea." *The Basic Works of Aristotle*. Ed. Richard McKeon. New York: Random House, 1941. 935-1126. Print.
- Aristotle. *Eudemian Ethics*. Ed. Michael J. Woods. Oxford: Clarendon, 1992. Print.
- Aristotle. "Topica (Topics)." *The Basic Works of Aristotle*. Ed. Richard McKeon. New York: Random House, 1941. 188-206. Print.
- Cooper, John M. "Aristotle on the Authority of 'Appearances'" *Reason and Emotion: Essays on Ancient Moral Psychology and Ethical Theory*. By John M. Cooper. Princeton, NJ: Princeton UP, 1999. 281-91. Print.
- Hales, Steven D. "The Faculty Of Intuition." *Analytic Philosophy* 53.2 (2012): 180-207. Web.
- Hales, Steven D. *Relativism and the Foundations of Philosophy*. Cambridge, MA: MIT, 2006. Print.
- Kraut, Richard. "How to Justify Ethical Propositions: Aristotle's Method." *The Blackwell Guide to Aristotle's Nicomachean Ethics*. Ed. Richard Kraut. Malden, MA: Blackwell Pub., 2006. 76-95. Print.
- Sytsma, Justin, and Edouard Machery. "Two Conceptions of Subjective Experience." *Philosophical Studies* 151.2 (2010): 299-327. Web.
- Weinberg, Jonathan M., Chad Gonnerman, Cameron Buckner, and Joshua Alexander. "Are Philosophers Expert Intuiters?" *Philosophical Psychology* 23.3 (2010): 331-55. Web.
- Weinberg, Jonathan M., Shaun Nichols, and Stephen Stich. "Normativity and Epistemic Intuitions." Ed. Edward Minar. *Philosophical Topics* 29.1 (2001): 429-60. Web.
- Williamson, Timothy. *The Philosophy of Philosophy*. Malden, MA: Blackwell Pub., 2007. Print.
- Woods, Michael J. "Chapter 6." *Eudemian Ethics*. By Aristotle. Ed. Michael J. Woods. Oxford: Clarendon, 1992. 58. Print.

## **Philosophical Thought Experiments are Arguments – and What This Means for Experimental Philosophy**

A great deal of the practice of philosophy relies on thought experiments. But while the usage of thought experiments is widespread, this paper shall focus on their application within the contemporary practice of experimental philosophy. Drawing on John Norton's seminal 1996 article "Are Thought Experiments Just What You Thought?", I will give arguments for the following two claims. First, that philosophical thought experiments are fundamentally arguments. Insofar as this is correct, second, I will argue that this status of thought experiments will entail a significant rethinking of the role given to thought experiments in the practice of experimental philosophy.

I outline my arguments as follows. My paper's first section will overview Norton's own argumentation that scientific thought experiments are arguments. In the second section, I will show that Norton's argumentation with respect to scientific thought experiments also applies to philosophical thought experiments. Consequently, if scientific thought experiments are arguments, then philosophical thought experiments are also arguments. The third section will be devoted to supporting the notion that philosophical thought experiments are arguments by looking at a famous philosophical thought experiment and showing how it can be reconstructed and, from this, illustrating some general principles of reconstructing philosophical thought experiments into arguments. The fourth section will deal with alleged difficulties faced by my view.

Having done so, I shall then turn to my second claim: Since philosophical thought experiments are arguments, experimental philosophers must reevaluate the role of thought experiments in the practice of philosophy. So my paper's fifth section will explore two

implications: that my thesis' correctness could explain certain experimental divergences and provide a way towards potentially rectifying these divergences.

### **I: Scientific Thought Experiments are Arguments**

Norton argues that scientific thought experiments just are arguments – and, if correct, this entails that any scientific thought experiment can be replaced by a scientific argument lacking the characteristics of a thought experiment. But his claim that scientific thought experiments are arguments occasions two significant questions. First, what is a scientific thought experiment? Second, in saying that scientific thought experiments *are* arguments, what sense of “are” is being invoked? These both shall be explored in turn.

Although philosophers might suppose that thought experiments are limited to their own discipline, such a supposition would be false. In fact, thought experiments can be found not only in other humanities disciplines but also in the fields of pure and applied sciences. Let us briefly examine a famous example.

Many famous scientific thought experiments were developed by Albert Einstein – and one which we shall examine in particular is known as “Einstein’s Elevator”. Einstein’s project of establishing the scientific correctness of special relativity theory required multiple steps. First, Einstein showed that an implication of special relativity theory was that *uniform* motion was relative. Thus, all inertial frames of reference were intrinsically indistinguishable from each other. Having shown that uniform motion was relative, Einstein next wanted to demonstrate that *accelerated* motion was relative.

To demonstrate this, Einstein presented the famous elevator thought experiment. Imagine, Einstein says, an opaque chest (such as an elevator) in a remote region of space that has no gravitational masses nearby. A rope is attached to the elevator and a “being” pulls on the

chest such that it accelerates uniformly. An observer inside the chest will see that all free bodies fall with equal acceleration inside the elevator – so, for example, the person inside the elevator will not be able to tell by dropping a ball whether the elevator is at rest in a gravitational field or whether the elevator is being uniformly accelerated. Since the two states of affairs (being in a gravitational field and being uniformly accelerated) are observationally indistinguishable, and since Einstein held that no good scientific theory should distinguish between two states which are observationally indistinguishable, this led Einstein to conclude that accelerated motion was indeed relative to the observer (Einstein 66-70).

This example should aptly demonstrate some key aspects of scientific thought experiments. While their domain is clearly different from that of philosophical thought experiments, their methodologies are surprisingly similar. A scenario is presented and the readers are asked to follow along, with the presenter, to the desired conclusion. The desired conclusion, of course, falls within the domain of the natural sciences rather than the humanities – but that is inessential to understanding the fundamental methodologies of scientific thought experiments.

Now that a good example of scientific thought experiments has been provided, I can clarify what more precisely is meant by Norton's claim that "scientific thought experiments are arguments". As philosophical readers know, the "to be" verb in English can take multiple senses. Among many, the three most significant are the "are"s of identity (bachelors *are* unmarried males), predication (those people *are* angry), and class membership (whales *are* mammals). So, when Norton writes that scientific "thought experiments are arguments" (Norton 1996 335ff), what sense of "are" is he relying upon?

Although Norton does not explicitly address which sense of "are" he is relying upon, we

may fairly judge that Norton uses the “are” of class membership to describe his own position. As a key indicator, Norton writes:

When we evaluate thought experiments as epistemological devices, the point is that we should evaluate them as arguments. A good thought experiment is a good argument; a bad thought experiment is a bad argument. For these reasons, in an earlier paper, I characterized thought experiments as belonging to a *subclass* of arguments (Norton 1996 336, emphasis mine).

Although initially murky, I believe Norton’s intent is sufficiently clear so as to understand his claim. Norton notes that, epistemologically speaking, thought experiments should be evaluated as arguments. Consequently, good thought experiments are equivalent to good arguments (and the same holds with bad thought experiments). Then, tellingly, Norton notes that in an earlier paper he held that thought experiments are a *subclass* of arguments. (Further, Norton’s purpose in context is not to *repudiate* his earlier views, but rather to *recapitulate* them before providing further argumentation as to what thought experiments are.) But insofar as thought experiments are truly a “subclass” of arguments, Norton must be intending the “are” of class membership – that is, that scientific thought experiments constitute one subclass (among many, presumably) of arguments. So, the sense of ‘are’ in which “thought experiments are arguments” is the same as the sense of ‘are’ in which “whales are mammals”.

Now that we understand what scientific thought experiments are and what Norton means to denote by holding that “thought experiments are arguments,” we shall turn our attention to why Norton believes that thought experiments are just a subclass of arguments. Norton begins his argument by considering the epistemic status of scientific thought experiments: “Thought experiments are supposed to give us information about our physical world. From where can this information come?” (Norton 1996 333) He explores two opposing answers to this question; they are referred to as the “Platonic conception” of thought experiments (Norton 1996 337) and his

own view that thought experiments are arguments. (Norton refers to his own view as the “argument view” of thought experiments (Norton 2004 1144).)

The Platonic conception of thought experiments, primarily advanced by J.R. Brown, responds to the above question by holding that scientific thought experiments draw on some special source of knowledge which transcends our ordinary epistemic resources. This fact, says Brown, explains why scientific thought experiments can yield significant scientific and theoretical advancements without gathering any new empirical inputs (and, occasionally, without having empirical inputs in the first place) (Brown 271). Brown describes his own position as “Platonism” (*ibid.*) and seems to straightforwardly embrace the view that scientific thought experiments result in immediate perceptions which enable us to “see” the truth of positions we would not otherwise have been able to (Brown 278).

This view is opposed by Norton’s own argument view of scientific thought experiments. Norton embraces a moderate empiricism such that “Insofar as [scientific thought experiments] can tell us about our world, they do so using our standard epistemic resources: ordinary experiences and the inferences we draw from them” (Norton 1991 334). As such, scientific thought experiments do not somehow transcend our empirical sources of knowledge but instead draw only from the kinds of ordinary epistemic resources whose existence is noncontroversial. This position, Norton argues, entails that scientific thought experiments are arguments. In order to make his position clearer, I shall explain and illustrate it in some detail.

Norton believes, to begin, that scientific thought experiments involve no new experimental data. That is, they neither collect new data nor, *contra* Brown, could they succeed absent empirical input. But since scientific thought experiments involve no new experimental data, they can only reorganize or generalize and make explicit what is already known about the

world.

This prior knowledge, according to Norton, rests upon our experiences and can only enter thought experiments as assumptions. “Therefore,” says Norton, “thought experiments are devices that reorganize or generalize these assumptions to yield the outcome of the thought experiment. That is, these devices are arguments that carry us from our assumptions to a conclusion” (Norton 1996 335). Insofar as a thought experiment reorganizes our assumptions and necessitates its own conclusion, the thought experiment is a deductive argument and insofar as it generalizes our assumptions and supports (without necessitating) its conclusion, it is an inductive argument.

So, Norton believes, when we evaluate scientific thought experiments as epistemological devices, they should be evaluated as arguments. This, Norton argues, entails two theses. First, the elimination thesis: any scientific thought experiment can be replaced by an argument without the character of a scientific thought experiment (Norton 1996 336). Second, the reconstruction thesis: All scientific thought experiments can be reconstructed as arguments based on tacit or explicit assumptions which were already present in the thought experiment (Norton 1996 339).

## **II: Philosophical Thought Experiments are like Scientific Thought Experiments**

Now that we better understand Norton’s own argumentation as to why scientific thought experiments are arguments, I shall offer my own extension of his project. More specifically, I shall demonstrate that philosophical thought experiments are relevantly analogous to scientific thought experiments. Thus, I shall argue, if scientific thought experiments are arguments, so too philosophical thought experiments are arguments. My arguments to demonstrate this claim will be twofold. First of all, Norton offers some necessary conditions of what it is to be a scientific thought experiment – and I will argue these necessary conditions are also fulfilled by

philosophical thought experiments. Second, Norton's conception of thought experiments is sufficiently broad that he correctly recognizes that scientific thought experiments can include philosophical premises – and if so, the differences between philosophical thought experiments and scientific thought experiments are not sufficient to prevent Norton's argumentation from applying to both philosophical and scientific thought experiments.

To begin, Norton does not specifically define what he means by a scientific thought experiment. That is, Norton does not offer necessary and sufficient conditions of a thing's being a scientific thought experiment. However, he does offer two necessary conditions of a thing's being a scientific thought experiment. Thus, if I can show that philosophical thought experiments fulfill the two necessary conditions Norton proposes, this will show us that philosophical thought experiments *could* be relevantly similar in the manner required for my argument to be successful.

I begin, then, by presenting Norton's two necessary conditions of a thing's being a scientific thought experiment and then showing that philosophical thought experiments also fulfill these necessary conditions. As Norton originally wrote in 1991, and has repeatedly reaffirmed, "Thought experiments are arguments which: (i) posit hypothetical or counterfactual states of affairs, and (ii) invoke particulars irrelevant to the generalities of the conclusion" (Norton 1991 129).

Before examining whether philosophical thought experiments fulfill these necessary conditions proposed by Norton, I should clarify one aspect of Norton's argumentative project. Norton does not intend to give a stipulative definition such that thought experiments are merely *defined* as being arguments; this would trivialize his argument's correctness. Rather, when Norton first offered this definition in 1986 he did not anticipate his views being particularly

controversial. Later, having encountered scholarly disagreement about whether or not scientific thought experiments actually are arguments, Norton responded by offering argumentation in favor of that aspect of his definition. However, neither Norton nor his various respondents have disagreed about (i) and (ii) being individually necessary conditions for a thing's being a scientific thought experiment.

Now we are in a position to evaluate whether or not (i) and (ii) apply to both scientific and philosophical thought experiments or only to scientific thought experiments. Fortunately, (i) is straightforward: Both scientific thought experiments and philosophical thought experiments rely on positing hypothetical or counterfactual states of affairs. Indeed if one mentally runs through a list of the most philosophically important thought experiments, one will find that they all involve hypothetical or counterfactual states. This is, in fact, precisely what makes them *thought* experiments rather than empirical experiments. Thus we are justified in concluding that one necessary condition of scientific thought experiments is also met by philosophical thought experiments.

What of (ii), then? Norton explains this second necessary condition of a scientific thought experiment as follows:

The presence of these particulars is what makes thought experiments experiment-like. Thus, in one version of the thought experiment in which Einstein sought to demonstrate that the effects of acceleration mimic those of gravitation, he asked us to imagine a physicist-observer who has been drugged and reawakens closed up inside a box. That there is an observer, that the observer is a physicist, that the physicist has been drugged, that he is enclosed within a box – all these are particulars which are irrelevant to the generality of the conclusion which Einstein seeks to draw. Without particulars such as these, however, thought experiments would not have their experimental appearance (Norton 1991 130).

When explained in such a way, it should be clear that philosophical thought experiments also invoke particulars irrelevant to the generalities of the conclusion. The tale of Gyges' Ring, as

related by Glaucon in Plato's *Republic*, is filled with particulars which are irrelevant to Glaucon's conclusion: that upon removing consequences for immoral actions, one's moral character would evaporate. But the particulars involved in the tale (that a shepherd finds a cave containing a ring and then uses the ring to seduce the Queen and commit regicide) are irrelevant to the conclusion Glaucon reaches. This same principle is displayed in philosophical thought experiments whether we consider trolley problems, Gettier cases, evil demons, or any other philosophical thought experiment.

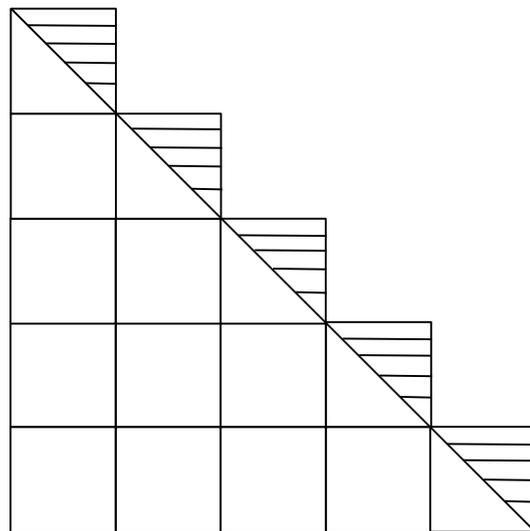
We are justified in concluding, then, that philosophical thought experiments fulfill both necessary conditions which Norton gives for scientific thought experiments. Thus, we are justified in concluding that philosophical thought experiments and scientific thought experiments could indeed be analogous in the sense required for my argument to succeed.

However, we need not rest with the conclusion that philosophical thought experiments could be relevantly analogous to scientific thought experiments. As I shall first argue, Norton's conception of scientific thought experiments is more broad than one might imagine – and this only increases the chances that scientific and philosophical thought experiments are fundamentally similar rather than dissimilar. Second, Norton correctly recognizes that scientific thought experiments can contain philosophical premises. This entails that the differences between philosophical and scientific thought experiments are not sufficient to prevent both from being arguments.

First, Norton's own conception of scientific thought experiments is broad – and the broader his conception of scientific thought experiments is, the more reasonable it is to believe that philosophical thought experiments will be relevantly analogous with respect to his argumentation regarding scientific thought experiments. When thinking of scientific thought

experiments, one is likely to only think of examples like the ones mentioned earlier. While Norton does hold that these kinds of thought experiments are scientific in nature, such a picture would be incomplete.

In addition to these “usual” scientific thought experiments, Norton also holds that thought experiments in the domain of pure mathematics can constitute “scientific” thought experiments. Brown, for example, proposes the following example as a scientific thought experiment which purportedly cannot be equivalently replaced with an argument. Consider the theorem derived from number theory that  $1 + 2 + 3 + \dots + n = n^2/2 + n/2$ . Brown offers the following “visual” proof of this theorem which (purportedly) by itself allows one to just *see* that the theorem is true:



In response to this, Norton writes:

Brown’s idea is that we immediately see the truth of the theorem in this figure without supplement of any text and that we do so without elaborate mathematical inferences. The powerful suggestion is that this moment of mathematical revelation coincides with the grasping of a Platonic law. The trouble with Brown’s case is that we do not immediately see the truth of the theorem upon being confronted with the figure (Norton 1996 352).

Norton then proceeds to explain how precisely we, in fact, “see” the truth of this theorem not by accessing a Platonic law but through step-by-step argumentation. Norton renders this thought experiment as the following argument:

- [1. Assumption] Each little square corresponds to an arithmetic unit.
- [2. From figure] The figure consists of  $(n=\text{five})$  columns of squares of height one, two, ...,  $(n=\text{five})$ ,
- [3. From 1 and 2] so that the total number of small squares is the sum we seek, one + two + ... +  $(n=\text{five})$ ...
- [4. From figure] The total number of squares is the sum of the number of squares in the shaded and unshaded portions.
- [5. From figure] The unshaded portion is half of an  $(n=\text{five})$  by  $(n=\text{five})$  square.
- [6. Assumption] An  $(n=\text{five})$  by  $(n=\text{five})$  square has  $(n=\text{five})^2$  unit squares in it.
- [7. From 5 and 6] The unshaded portion has  $(n=\text{five})^2/2$  unit squares.
- [8. From figure] The shaded portion consists of one half square for each column.
- [9. From 2 and 8] There are  $(n=\text{five})/2$  shaded squares.
- [10. From 4, 7, and 9] The total number of unit squares is  $(n=\text{five})^2/2 + (n=\text{five})/2$  which is the result sought (Norton 1996 352-353).

But, of course, *how* Norton renders this argument as a thought experiment is perhaps less significant for present purposes than *that* he does so. That is, Norton believes that what Brown offered was in fact not only a thought experiment but a scientific thought experiment in a sense sufficient for Norton to believe that it could be replaced with an argument lacking the characteristics of a thought experiment. Thus, while it is correct to imagine that Einstein’s Elevator is a scientific thought experiment, Norton also believes that the category of scientific thought experiments is at least broad enough to capture categories not usually considered to fall within its domain such as number theory.

Keeping this in mind, then, will help widen our understanding of just what it is to be a scientific thought experiment in Norton’s sense of the term. Obviously, the more expansive his understanding of scientific thought experiments, the more likely it is that those same arguments will be relevant to philosophical thought experiments.

Indeed, these implications could well prove decisive. For, assuming Norton is correct, this argument alone may well be sufficient to establish a relevant similarity between philosophical and scientific thought experiments. For if scientific thought experiments can contain empirical and theoretical premises as well as standard forms of argumentation, then it is difficult to see what the relevant dissimilarity between philosophical and scientific thought experiments could be. More precisely, given that scientific thought experiments can contain empirical premises (about the appearances of things in a uniformly accelerating elevator, say) and theoretical premises (about what states of affairs should be distinguished in legitimate scientific theories, say), and standard argument forms (*reductio ad absurdum*, say), then there does not appear to be any interesting philosophical difference between Einstein's Elevator and, say, Thomson's Violinist. Granted, the *conclusions* will be of scientific and philosophical natures, variously, but that will not suffice to demonstrate that scientific thought experiments are arguments while philosophical thought experiments are not. Rather, it helps demonstrate that *if* scientific thought experiments are a subclass of arguments, *then* so too are philosophical thought experiments another subclass of arguments.

To further this argument that I am making, I shall conclude by noting that Norton correctly demonstrates that scientific thought experiments can contain philosophical premises – and if this is the case, then it seems no good reason remains for holding that scientific thought experiments are arguments while philosophical thought experiments are not.

Consider Norton's earlier treatment of Einstein's Elevator. Although I offered my own summary of this thought experiment, Norton offers his own in the more standard premise/conclusion form as follows:

1. An observer in an elevator cannot empirically distinguish between being accelerated and being in a uniform gravitational field.

2. This situation is typical; the details of the observer in the elevator are not relevant.
3. Verification Principle: States of affairs which are not observationally distinct should not be distinguished by the theory.
  - ∴ Being uniformly accelerated and being at rest in a uniform gravitational field should not be theoretically distinguished.
  - ∴ Principle of Equivalence: Being uniformly accelerated is identical to being at rest in a gravitational field (Norton 1991 136-138).

Brown notes, and I agree, that this argument (which Norton believes is equivalent to Einstein's original thought experiment) actually contains a philosophical premise (Brown 1992 273).

Notice the third premise. It is straightforwardly taken from philosophy of science and it holds that if two states of affairs are not observationally distinct, a scientific theory should not distinguish between them. Norton himself (in an earlier paper) notes that Einstein's "interesting point is the success of the introduction of the philosophical principle" which is labeled as 3 above (Norton 1991 139).

So since Norton believes that scientific thought experiments can contain philosophical premises, and his belief is correct, this strongly suggests that the arguments Norton makes with respect to scientific thought experiments also apply to philosophical thought experiments. Indeed, then the main difference between scientific and philosophical thought experiments is not that one has no philosophical premises while the other does; rather, since both can contain philosophical premises it must be the nature of the accompanying premises which distinguishes philosophical from scientific thought experiments. If this is indeed the case, then I have shown that philosophical and scientific thought experiments are sufficiently analogous such that, if scientific thought experiments are arguments, so too are philosophical thought experiments arguments.

### **III: Demonstration**

To make my views more plausible, I shall now give an example of and comment on how

exactly philosophical thought experiments can be understood as (and indeed formulated as) philosophical arguments. I shall begin by considering how one can render the quintessential philosophical thought experiment as a philosophical argument: Gettier cases.

Gettier cases, of course, are designed to disprove the traditional “justified true belief” understanding of knowledge (where justification, truth, and belief are regarded as being individually necessary and jointly sufficient conditions for knowledge). Gettier asks us to consider a thought experiment of the following sort: Imagine Jones and Smith have applied for a job and that Smith has formed the following conjunction: “Jones will get the job and Jones has ten coins in his pocket”. On this basis, Smith infers “The man who gets the job has ten coins in his pocket”. Further, let us imagine that Smith has strong justification for his belief that the man who gets the job has ten coins in his pocket.

As Gettier indicates, Smith clearly has a justified belief that the man who gets the job has ten coins in his pocket. However, it is Smith (not Jones) who gets the job, and as it turns out Smith also has ten coins in his pocket. In this case, then, Smith had a justified true belief that the man who gets the job has ten coins in his pocket – but, according to Gettier, we are unwilling to attribute knowledge to Smith. As such, then, this disproves the notion that justification, truth, and belief are individually necessary and jointly sufficient conditions for knowledge (Gettier 122).

I rely here on Timothy Williamson’s book *The Philosophy of Philosophy* where he spends no little effort reconstructing Gettier cases as formal arguments rather than thought experiments (Williamson 181-187). Therein, Williamson offers a symbolic (modal) logic proof to demonstrate the insufficiency of the traditional conditions for knowledge. Instead of laboriously reproducing Williamson’s proof, I shall instead narratively outline Williamson’s work.

Williamson notes that one can be justified in believing what is in fact false. Further, if one is justified in believing something and correctly uses deduction, one is justified in believing what is so deduced. Since any truth can be deductively entailed by various falsehoods, one can believe a truth on the basis of correct deduction from a justified falsehood one believes. On that basis, one would be justified in believing the deduced truth too. Nevertheless knowledge does not result, because a belief no matter how justified may be relying on deduction from a false belief, and one's conclusion cannot be epistemically better off than one's premises. Therefore justified true beliefs are not sufficient for knowledge. So while it may seem that Gettier cases point out an instance where justified true beliefs are not sufficient for knowledge, the "pointing out" requires, implicitly, one to recognize that a justified deduction is taking place from a false belief.

Williamson's reconstruction aptly illustrates two principles which should be kept in mind when reconstructing thought experiments into arguments. Both of the principles I shall examine stem from a larger commitment to philosophical charity – that is, the disciplinary expectation that philosophers will, insofar as it is possible, treat their sources with the same degree of charity we would hope our own works would be given.

Consequently, when rendering thought experiments as arguments, we should adopt a principle of *validity*. This means that if a thought experiment may be rendered as at least two philosophical arguments, one valid and the other invalid, philosophical charity requires that we adopt the interpretation of the thought experiment which forms a valid argument. For example, when interpreting an author we should strive to understand her argument in some manner which preserves logical validity. Doing so need not require formalizing each and every argument as formally valid – but an argument should be presented such that if it *were* formalized, its

formalization would result in a valid argument.

A second principle of replacing thought experiments with arguments is one of *plausible soundness*. This means that, not only should a thought experiment be rendered as a valid argument, but this should be done in the most plausibly sound formulation possible. Where there is ineradicable ambiguity between two equally valid and plausibly sound arguments, this fact should be highlighted and analyzed appropriately. I suspect these principles are implicit in the proper practice of philosophy itself. Nonetheless it seems reasonable to highlight their importance particularly when discussing replacing another's thought experiment with an argument lacking the characteristics of a thought experiment.

#### **IV: Replies to Objections**

Let us now address a seeming difficulty with my thesis that all philosophical thought experiments are philosophical arguments. One could allege that, in some philosophical thought experiments, the conclusion to be drawn is unclear or wholly absent – but if all philosophical thought experiments are arguments, they cannot have unclear or absent conclusions. This objection, then, aims to disprove the soundness of my arguments.

Although I certainly cannot refute all potential thought experiments which are alleged to falsify my view, I shall address the most commonly proposed counterexample: the Trolley Problem. I shall summarize the Trolley Problem and then show how, in fact, it can be reconstructed as a philosophical argument.

Let us begin by precisifying what the Trolley problem is. One of its main versions is presented in Thomson's article "Killing, Letting Die, and the Trolley Problem". Part of her project in this article is to consider whether there is a moral difference between killing and letting die (and thus, obliquely, to consider the distinction between *doing* harm and *allowing* harm to

come to another). Thomson contrasts, then, the intuitions of the “transplant” case with the intuitions of the “fat man” trolley problem. Thomson first presents the transplant case as follows:

David is a great transplant surgeon. Five of his patients need new parts. One needs a heart, the others need, respectively, liver, stomach, spleen, and spinal cord but all are of the same, relatively rare, blood-type. By chance, David learns of a healthy specimen with that very blood-type. David can take the healthy specimen’s parts, killing him, and install them in his patients, saving them. Or he can refrain from taking the healthy specimen’s parts, letting his patients die (Thomson 206).

This scenario, Thomson believes, demonstrates that killing is worse than letting die since it would be immoral for David to cut up one patient to save the five. Thomson then recapitulates Foot’s version of the Trolley problem and argues that the trolley driver may redirect his trolley so that it kills one rather than five. Thomson then considers the “fat man” trolley case:

George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die (Thomson 207-208).

Thomson then notes that, presumably, George may not shove the fat man into the trolley’s pathway. But then, she wonders, why is it wrong to shove the fat man on the tracks but not wrong to alter the trolley’s path to hit one man rather than five people? She thinks this is because of some similarity between David’s Transplant case and George’s Fat Man case which does not hold with respect to the original trolley problem. This factor, Thomson holds, is that there is a morally relevant difference between that to which one has a claim (broadly construed) and that to which one hasn’t a claim. In the Transplant and Fat Man cases, both of those individuals have a claim to the goods in question (their organs and body, respectively).

However, in Foot's Trolley problem, the one has no more claim to his life than the five have claim to their lives – and consequently it is morally allowable to redirect the trolley onto the one person while it's not morally allowable to push the fat man into the trolley's path.

Taken together, these cases also clearly suggest an argument (although Thomson emphasizes that an argument is not entailed but only *suggested* by the situations she analyzes – and thus, it is possible that Thomson intended her argument to be inductive in nature rather than deductive). Regardless, we have seen that Thomson's Trolley Problem is a thought experiment which can be reconstructed as an argument, and hence is no counterexample to my claims.

But what if what is meant by the Trolley problem is not Thomson's version of the problem but perhaps its more common classroom variant? In it, students in introductory-level philosophy and introductory-level ethics classes are presented with the Transplant case by their instructor. This presentation is then followed by Thomson's "fat man" trolley scenario. Students, frequently, are then directed by their instructor to note the similarity between the two cases – and if they think that taking organs (in the Transplant case) would be immoral but pushing a fat man onto the tracks would not be immoral, then they are asked what the relevant moral difference between these situations is. Presented in this manner, it might seem harder to see what "argument" is being made – after all, aren't the students merely presented with the scenario and then asked to draw their own conclusions?

Such a teaching methodology in no way prevents the thought experiment proffered from being an argument. In fact, it is often the case that conclusions (or premises) to arguments may be left unstated. I myself have often done this when teaching introductory critical-thinking classes and, while teaching validity, asked my students to identify what conclusion necessarily follows from the premises "All men are mortal" and "Socrates is a man". The same may be done

in classroom presentations of the various trolley problems (or all other philosophical thought experiments). But failing to present a conclusion (or all premises) does not prevent something from being an argument; it suffices to prevent it from being a *complete* argument (and incomplete arguments are frequently bad), but they are still arguments nonetheless.

I argue, then, that when students are given (say) the Trolley problem and asked what conclusion follows from it they are actually *reconstructing* an argument. If a person believes, say, that pushing the fat man in front of the trolley is moral, that person is recreating some or other argument whose conclusion is that “this method of sacrificing one to save five is morally allowable”. The premises may guarantee the truth of the conclusion or may only indicate its truth – and this reconstruction process may be conscious or unconscious – but nonetheless the student is doing what we all do with respect to arguments. We complete them where they are incomplete (either in their premises or conclusion), we decide how strongly the premises support (or necessitate) the conclusion, and we accept the conclusion (or not). But the mere fact that the presenter allows his audience to reconstruct the last (or initial) steps of the argument does not entail that what is being given is not an argument.

Other possibilities exist, however. If some students believe the correct answer to the trolley problem is shoving the fat man onto the tracks and other students believe that the correct answer is to let the five men die, which argument has been presented? These circumstances do not change the argument that Thomson has presented – namely, that one may not morally push the fat man onto the track. After all, thought experiments elide over some premises which would be present in a full argument. So, when the students reconstruct the thought experiment as an argument, they might well be reconstructing different arguments based on the ambiguities present in the original thought experiment – and this would suffice to explain why some students

reach one conclusion and others the opposite. Alternatively, even if the reconstructed premises are largely the same, some students might believe a *modus tollens* operation is necessary to make the argument sound while others think *modus ponens* is. This also would suffice to explain why two persons considering many of the same premises reach opposing conclusions; this phenomenon is not unique to evaluation of philosophical thought experiments.

One might also object to my thesis on the following grounds: Suppose one considers the genesis of a thought experiment – that is, when a philosopher makes up a thought experiment but does not yet have an argument in mind. Perhaps she presents a thought experiment to raise an issue about which she is not presently presenting an argument one way or the other. This might seem to disprove my thesis since, if all thought experiments are a subclass of arguments, it would seem impossible to raise a thought experiment without giving an argument. What can we say about this?

Rather than replying to this purported counterexample directly, I observe that this type of reply is part of a broader class which asserts that there is at least one thought experiment which cannot be reconstructed as an argument. If this claim were true, it would falsify my thesis – so rather than addressing this example specifically, I shall address the kinds of strategies I could adopt in response to any purported counterexample of this type.

One option is to deny that any purported counterexample involves a *complete* thought experiment. For example, when a thought experiment is being invented it may lack crucial details. The person developing it may be unsure what the correct resolution to the thought experiment is – and may then seek input from colleagues who will point out ways to strengthen the thought experiment.

Now in its very earliest stages (whether thought experiment or argument), the “argument”

might be so radically incomplete that it is not, properly speaking, even an argument. Certainly not all things which *purport* to be arguments are, in fact, arguments. But as the argument or thought experiment moves beyond the very initial state of being radically incomplete, it becomes more and more complete – and hence our ability to reconstruct it as a standard argument grows. As mentioned earlier, my thesis does not entail that all thought experiments (or arguments) are *valid* or *complete*. Incomplete thought experiments may, then, only be reconstructed as incomplete arguments – but *upon their completion* my thesis entails that they can all be reconstructed as complete arguments.

A second option is to deny that the counterexample really involves a thought experiment at all. Certainly not all abstract or hypothetical scenarios are, properly speaking, thought experiments. For example, consider a “thought experiment” containing a simple scenario wherein you find a dead body – and then you are asked what you would do. If this is all there is to the “thought experiment”, and if this were presented as a thought experiment which purportedly could not be reconstructed as an argument, I would deny that this were, properly speaking, a thought experiment at all. This scenario, it seems to me, involves nothing more than a hypothetical circumstance and a query whose response would ordinarily lack philosophical significance. Thus I would argue that this scenario does not actually constitute a thought experiment. This second strategy of denying that a purported counterexample involves a thought experiment will be viable unless one implausibly holds that *all* abstract or hypothetical scenarios in philosophy are thought experiments.

I have shown good reason to think that all philosophical thought experiments are arguments – and two strategies one could use in response to any purported counterexamples. Having given some reason for thinking my thesis is true, then, I shall now turn to consider what

implications follow from the truth of my thesis.

## **V: Implications**

There are two significant implications that flow from the correctness of my thesis with respect to experimental philosophy. First, the correctness of my thesis provides one possible way to explain the appearance of certain disparate experimental results. Second, it provides a reason to carefully examine whether to use thought experiments or arguments when conducting experimental philosophy.

First, the correctness of my thesis provides a potentially fruitful way to explain contradictory experimental results. Although little experimental philosophy work has been done to date (comparatively speaking), research has already uncovered instances where contradictory results are generated. An experimental philosopher does research involving thought experiments and discovers that a high percentage of his research subjects share a given intuition. But another experimental philosopher gives his research subjects the same thought experiments and discovers that a high percentage of his subjects reject that intuition.

There are, to be sure, many potential explanations for the generation of disparate data (as researchers in the natural sciences are well familiar with). Perhaps the researchers biased the results in some subtle way. Or perhaps some small (unintentional, even) change in the experimental protocol caused the data generated for the followup study to be significantly out of accord with the data generated by the original study.

But another *potential* reason why disparate experimental results can be generated is that, whenever thought experiments are used, research subjects must (explicitly or implicitly) analyze the argument inherent in the thought experiment. But, as already explained, thought experiments can omit or elide over premises or conclusions – and if so, the research subject must (explicitly

or implicitly) supply them for herself. However, if it is incumbent upon each research subject to reconstruct the argument contained in the thought experiment, there is a very real possibility that different experimental subjects are supplying different premises or conclusions – and hence are reconstructing different arguments.

It is unlikely that this is the *only* explanation for disparate experimental results. However, it is at least *prima facie* one possible explanation for divergences in research data – and if the correctness of my thesis is neglected then this possible explanation is likely to be overlooked. Further experimental research itself will be needed to determine whether or not this is, in fact, the best explanation for experimental divergence. To do so, a researcher would need to present a thought experiment and ask the subjects to reconstruct it as a propositional argument. Then, the amount of divergence between reconstructions could be quantified. Obviously if the researcher encounters relatively little divergence among the reconstructions, then this experimental hypothesis would be falsified. But if such divergences are meaningfully encountered, then one explanation would be provided by the correctness of my thesis.

Second, if such experimental divergences are uncovered, philosophers would need to evaluate whether or not to use thought experiments or the arguments they could be reconstructed into when conducting experimental philosophy surveys. That is, if there are in fact various subtypes of arguments (among which are propositional arguments and thought experiments), then researchers will need to carefully evaluate the costs and benefits of using one or another type of arguments in their research.

For example, perhaps the benefits of thought experiments are sufficiently significant to outweigh their costs. Even if thought experiments are less readily reconstructed as propositional arguments, perhaps they are so much more understandable that they are worth using nonetheless.

Perhaps the vivid nature of thought experiments significantly aids the research subjects in comprehending the essential claims of the thought experiment. If so, it might well be the case that the costs of using thought experiments are outweighed by their benefits. Of course, reasoned analysis might also reveal that the costs of using thought experiments in experimental research are greater than the benefits from the available alternatives. If so, then this would help supply a principled reason (deriving from good experimental design) to accept or reject the use of thought experiments in experimental philosophy research. The correctness of my thesis, in short, provides a reason for us to not neglect having an important conversation about the role of thought experiments and their role in experimental philosophy that would otherwise likely be neglected.

## Works Cited

- Brown, James R. "Why Empiricism Won't Work." *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 2 (1992): 271-79. Print.
- Einstein, Albert. *Relativity: The Special and the General Theory*. London: Methuen, 1917. Print.
- Gettier, Edmund L. "Is Justified True Belief Knowledge?" *Analysis* 23.6 (1963): 121-23. Print.
- Norton, John D. "Thought Experiments in Einstein's Work." *Thought Experiments In Science and Philosophy*. Ed. Tamara Horowitz and Gerald Massey. Savage, MD: Rowman and Littlefield, 1991. 129-48. Print.
- Norton, John D. (1996) "Are Thought Experiments Just What You Thought?" *Canadian Journal of Philosophy* 26.3 (1996): 333-66. Print.
- Norton, John D. (2004) "Why Thought Experiments Do Not Transcend Empiricism." *Contemporary Debates in the Philosophy of Science*. By Christopher R. Hitchcock. Oxford: Blackwell, 2004. 44-66. Print.
- Thomson, Judith J. "Killing, Letting Die, and the Trolley Problem." *The Monist* 59.2 (1976): 204-17. Print.
- Williamson, Timothy. *The Philosophy of Philosophy*. Malden, MA: Blackwell Pub., 2007. Print.

## Sorting Through Our Ethical Intuitions

Philosophy, as most (including myself) think, cannot do without intuitions. Certainly the role of intuitions in the history of philosophy has been significant – one need only think of the intuitions elicited by Descartes' Evil Demon scenario or more modern examples such as Dretske's Zebras or Foot's Trolley. But considering particularly the field of ethics, intuitions not only serve a justificatory role but often are pressed into service in the role of ultimate justification. When this happens, we encounter a phenomenon familiar to professional philosophers: The intuitions elicited by some conflict with the intuitions of others.

When this happens, philosophers seemingly face an impasse, for if what ultimately justifies your ethical position is your intuition  $p$  and what ultimately justifies my rejection of it is my intuition  $\sim p$ , forward progress seems impossible. We may appeal to other philosophical considerations (such as the coherence of our intuitions with other known facts) or (for conversational reasons) we may instead talk of baseball, but so long as our intuitions diverge over a matter of ultimate ethical justification it seems that no forward philosophical progress can be made.

My paper's goal is to reject this received view as incorrect, at least for a subset of cases within the philosophical subfield of ethics. When two persons' intuitions conflict I shall argue that, in some limited circumstances, there exists a way to adjudicate this intuitional conflict. I shall not only argue that, in these circumstances, one intuition must be incorrect and another must be correct but also that we can know which intuition must be correct and which one must be incorrect. In order to do so, I shall first sketch out the problem at some greater length and then, second, bring epistemological considerations from Steven Hales' work to bear. My own

contribution will come in the third section where I apply Hales' epistemological position to ethics and in the fourth and fifth sections where I refute some purported objections to my arguments.

## **I: The Problem**

The problem, in brief, is whether or not philosophers can resolve intuitional conflicts between each other. The standard view is that these conflicts between intuitions cannot be resolved – that is, there does not exist some evaluative process such that, using it, we may evenhandedly consider two persons' intuitions and discover that at least one person's intuitions are wrong. Rather than multiplying examples of this view I shall instead pick out its most significant proponent. No less than A.J. Ayer in his famous *Language, Truth and Logic* writes:

...[I]t is notorious that what seems intuitively certain to one person may seem doubtful, or even false, to another. So that unless it is possible to provide some criterion by which one may decide between conflicting intuitions, a mere appeal to intuition is worthless as a test of a proposition's validity. But in the case of moral judgements, no such criterion can be given (Ayer 106).

Here Ayer expresses a common belief and then issues a stinging challenge. The commonplace he expresses is the fact of intuitional divergence – our intuitions, however well trained, educated, and informed can and do conflict. Thus what seems intuitively true to you seems intuitively false to me; what seems intuitively certain to one is doubtful for another.

Ayer then issues a bold challenge: Unless there exists some criterion by which these intuitional divergences may be settled, appeals to intuition (by themselves) are worthless for discovering the truth or falsehood of a proposition. Then, in virtue of his logical positivism, he holds that with regard to moral judgments no such criterion can be given. Although Ayer's challenge is specifically applied to ethics, there is no principled reason why it could not apply to

the rest of philosophy as well – unless some criterion to decide between competing philosophical intuitions exists, the role of intuitions in philosophical practice is significantly undercut.

*Fully* answering Ayer’s challenge is at least a book-length task; it would require giving a series of necessary and sufficient conditions for adjudicating intuitional conflicts within all the various philosophical subfields. While certainly appropriate to a career’s *magnum opus*, my aim here is much more moderate. I shall argue that there exists a method of resolving some kinds of intuitional conflicts within the philosophical subfield of ethics. While thus not a full response to Ayer’s challenge, it shall serve as a first step towards answering Ayer’s objection concerning the role of intuitions in philosophy. In order to demonstrate my position, let us now turn to some work from the philosopher Steven Hales.

## **II: Hales on Foundationalism**

My particular response to Ayer’s challenge approaches the goal obliquely, via Steven Hales’ argument in favor of foundationalism. Hales’ article does not address Ayer’s challenge directly; instead his goal is to solve the “problem of intuitions” – but Hales’ arguments put us in a position to begin responding to Ayer’s challenge. The problem of intuitions, as addressed in his article, concerns a dilemma created by responding to the question “What justifies intuitions?”. In response, one can offer either an empirical (*a posteriori*) or an intuitive (*a priori*) justification of intuitions. For reasons Hales discusses at some length, the possibility of an *a posteriori* justification for intuitions is unlikely. But justifying intuitions on *a priori* grounds will itself involve an appeal to intuitions, and thus is straightforwardly circular. This, then, is the problem faced by every philosophical use of intuitions which purports to serve a justificatory role.

Hales, helpfully, formulates the problem of intuition using formal argumentation.

Among other advantages, this presentation style allows readers to verify the argument's validity – and, subsequently, Hales can focus his efforts on defending the truth value of the propositions.

Hales formulates the problem of intuition, which he wishes to ultimately reject, as follows:

Premise 1: If a proposition is epistemically justified, then it is justified either *a priori* or *a posteriori*.

Premise 2: If a proposition is epistemically justified *a priori*, then its justification depends on the method of intuition justifying some propositions.

Premise 3: If the proposition “the method of intuition justifies some propositions” is epistemically justified, it is not justified *a posteriori*.

Premise 4: “The method of intuition justifies some propositions” is epistemically justified.

Premise 5: Nothing is self-justifying.

(From 1,3) 6: If “the method of intuition justifies some propositions” is epistemically justified, it is justified *a priori*.

(From 2, 6) 7: If “the method of intuition justifies some propositions” is epistemically justified, then its justification depends on the method of intuition justifying some propositions.

(From 4, 7) 8: The justification of “the method of intuition justifies some propositions” depends on the method of intuition justifying some propositions.

(From 5, 8) 9: Thus, “the method of intuition justifies some propositions” is not epistemically justified.

(From 4, 9) 10: “The method of intuition justifies some propositions” is and is not epistemically justified (Hales 139).

Hales then spends the rest of the article analyzing the premises listed for the purpose of avoiding the conclusion found in line ten, since that line contains a contradiction.

Although recreating the full scope of Hales' argumentation is unnecessary, some brief remarks will illustrate the problem's nature. The contradiction is generated from lines four and nine; line four is a premise and line nine follows from the other lines as listed above. Hales (and myself and many other philosophers) believe that line nine is false – that the statement “the method of intuition justifies some propositions” in fact *is* epistemically justified. Since the argument is valid, in order to save line four (and thus reject line nine), at least one of the

previous lines must be false. If so, the argument's validity can be conceded while rejecting its soundness.

All philosophers (save radical empiricists who believe that only *a posteriori* justification is possible) accept premise one, so its being false is implausible. Premise two asserts that *a priori* justifications depend on intuitions. Just *how* this happens is less controversial than *that* it happens (again, assuming one excludes a philosophy of radical empiricism). So denying premise two is unattractive.

Premise three states that, if it is justified at all, the proposition "the method of intuition justifies some propositions" is not justified *a posteriori*. To say otherwise, Hales argues, entails that we must do *a posteriori* appraisals of philosophical propositions. And if, as Kornblith argues, we are (say) concerned with the nature of knowledge (rather than the concept of knowledge), then empirical studies of facts discovered *a posteriori* are not relevant (Kornblith 133).

Premise four holds that "the method of intuition justifies some propositions" is itself epistemically justified. That is, if this premise were false, we would be intellectually unjustified in believing that intuitions can provide propositional justification. But the practice of philosophy itself hangs on premise four's truth, given that philosophy uses and cannot avoid using intuitions. So, if the discipline of philosophy has not been wildly misconceived, premise four is true.

This leaves, Hales argues, only premise five (that no proposition is self-justifying) that can be false. Hales writes: "What does [premise five] mean? The idea is that there are no basic propositions whose justification stems from no source other than themselves that we are justified in accepting" (Hales 145). Hales, though, rejects premise five and instead concludes that "We

must assume the method of intuition delivers justified propositions when we employ this method to show that ‘the method of intuition justifies some propositions’” (*ibid.*).

Hales draws from this a fascinating entailment: If his analysis is correct, then a kind of moderate foundationalism must be correct. “Moderate foundationalism in this sense expresses the thesis that there are justified propositions whose justification depends on nothing other than themselves” (*ibid.*). Since the proposition “the method of intuition justifies some propositions” is itself in need of no justification (and thus serves as a kind of self-justifying justifier), the falsity of premise five entails the correctness of this kind of moderate foundationalism.

### **III: Answering Ayer**

Let us assume that Hales is correct (as I think he is). Hales’ demonstration that a kind of moderate foundationalism is rationally unavoidable (unless one rejects *a priori* justification *in toto*) is philosophically significant. However, how does the truth of moderate foundationalism allow us to (partially) answer Ayer’s challenge? I shall defend the following view: If modest foundationalism is true and at least one self-justifying proposition concerns ethics, then such a self-justifying ethical proposition must be held *a priori*. To answer Ayer then, in cases where one person’s intuitions ultimately derive from an ethical first principle held *a priori* and another’s from an ethical first principle held *a posteriori*, the latter must be wrong. Arguments in favor of this position will be presented and then objections dealt with.

To speak most accurately, Hales has demonstrated the truth of an either/or proposition: *either* modest foundationalism (as he defines it) is true *or* “philosophy grounded in the use of rational intuition is bunk” (Hales 145). Of course, there are no doubt some persons (perhaps

even some philosophers) who would accept the latter claim and agree that the philosophical method (based on rational intuition) has been nothing short of spectacularly false.

What may one say to such a view? Sadly, nothing. Indeed to such persons no reply is possible since one simply cannot do philosophy without *some* reliance on rational intuition, and any philosophical arguments I might make to such persons will thus of necessity rely on some rational intuitions. (The difficulty of refuting such a view is thereby analogously difficult to refuting the views of the radical skeptic about knowledge of the external world.)

Since no reply to such persons can be made, it is more productive to consider Hales' finding that modest foundationalism is true and see what interesting implications follow from his work. Modest foundationalism, recall, is the view that at least one premise is self-justifying. Hales believes that such a self-justifying premise is "the method of intuition justifies some propositions" (Hales 145) – unless, as mentioned before, standard philosophical methodology is grossly erroneous. Now if self-justifying propositions exist, what follows if one of those propositions concerned ethics?

(To be clear, I am considering only a meta-claim about self-justifying propositions: Namely, that one of them concerns the philosophical subfield of ethics. I set to the side for this article whether a self-justifying ethical proposition would entail the truth of utilitarianism, Kantianism, Thomism, etc. Instead we shall explore what follows given the existence of at least one self-justifying ethical proposition.)

If such a self-justifying proposition concerning ethics exists, then we are positioned to answer Ayer's challenge under certain specific conditions. To see how, consider what the nature of a self-justifying ethical proposition must be. Such a proposition, by definition, justifies other

ethical propositions without itself needing justification. In the first chapter of Mill's *Utilitarianism*, he refers to such a proposition as concerning "the foundation of morality [which] has been accounted the main problem in speculative thought" concerning "the criterion of right and wrong" (Mill 1). As Mill recognizes throughout the first chapter, such an ethical "first principle" (Mill 4) is necessary to progress in ethics, for without it one cannot resolve the numerous subsidiary ethical problems humans face. I shall argue that, in order for a proposition to express an ethical first principle, a necessary condition is that it be justified *a priori*.

But what is it for a proposition to be justified *a priori*? Although philosophers have provided various definitions, they standardly assume that a proposition is justified *a priori* just in case its justification derives from pure reason alone (and thus does not depend on experience). All propositions that are justified but are not justified *a priori* are justified *a posteriori*. Thus Hales notes that "the *a priori* and the *a posteriori* are exhaustive and exclusive categories of justification" (Hales 141).

I shall now overview my forthcoming arguments concerning why an ethical first principle must be justified *a priori*. This is the case because it is impossible for an ethical first principle to be justified *a posteriori* – and ethical first principles are justified. So if an ethical first principle is justified but cannot be justified *a posteriori*, it must thus be justified *a priori*. Arguments in favor of each of these statements shall now be presented.

First, I argue that an ethical first principle cannot be justified *a posteriori*. This is so because all propositions justified *a posteriori* require some external inputs; science handily provides these for various empirical propositions. Since *a posteriori* justification of a

proposition must necessarily have external inputs, what ultimately justify *a posteriori* propositions are those external inputs.

But *self*-justifying propositions cannot depend on external inputs for their justification – this would be a contradiction inasmuch as self-justifying propositions cannot rely on something external to themselves in order to be justified. Thus it is impossible for an ethical first principle to be justified *a posteriori*.

But an ethical first principle *is* justified. Most properly speaking, an ethical first principle is a self-justifying justifier: It justifies other propositions without itself requiring any other justification. Since a thing cannot give what it itself lacks, an ethical first proposition could not justify other propositions if it were not itself justified. This is, then, the sense in which an ethical *first* principle is said to be first: It provides the justification (which it possesses) to other secondary and tertiary ethical principles.

So since an ethical first principle is justified, but cannot be justified *a posteriori*, it must be justified *a priori*. While this result might seem straightforward, it carries a powerful implication which will enable us to partially meet Ayer's challenge. Let us now see how this is true.

Consider two philosophers having an ethical disagreement (that is, a disagreement not over pragmatic or prudential ethical considerations but, properly speaking, a genuine *philosophical* disagreement concerning an ethical matter). They discover that their disagreement stems from differing rational intuitions, and that these differing intuitions derive from deeper ethical commitments. Eventually, assuming this is a genuine *philosophical* disagreement, they will discover the ethical first principle about which their intuitions disagree. This is precisely a

kind of situation Ayer alludes to when he discusses “what seems intuitively certain to one person” seeming “false to another” (Ayer 106). Now since both philosophers have an intuitional conflict over which ethical first principle is correct, in a certain case we are able to offer what Ayer thought impossible: a “criterion by which one may decide between conflicting intuitions” (Ayer 106). How may this be done?

We may presume that both philosophers take their respective ethical first principles to be justified (if one does not, or neither do, then the reason for disputing evaporates). On the assumption that both philosophers take their respective ethical first principles to be justified, we may then ask, of each, whether her ethical first principle is justified *a priori* or *a posteriori*. Since I have shown that an ethical first principle must be justified *a priori* if it exists at all, we then have a criterion to decide between conflicting intuitions in the following case: If one ethical first principle is justified *a priori* and the other is justified *a posteriori*, and if an ethical first principle exists, we must reject the intuitions behind the *a posteriori* ethical first principle.

As an example of my strategy in action, let us briefly consider Sabine Roeser’s affectional intuitionism. According to Roeser, intuitions should be comprised of both rational and emotional components (Roeser 110). Further, she holds that these intuitions can serve as foundational beliefs in terms of ethical justification (Roeser 152).

Significantly, and unlike many philosophers, Roeser explicitly considers Ayer’s challenge and its severe difficulties for ethical intuitionism. Ultimately, she grants that if “intuitionism cannot help us decide between conflicting intuitions, [this] could still be a reason to reject intuitionism as a useless theory” (Roeser 102). Since Roeser correctly recognizes the potential for Ayer’s challenge to refute her entire ethical paradigm, she offers two responses.

Her first response, while important for Roeser's purposes, does not concern my own arguments. (Roeser argues that Ayer's challenge is not a unique difficulty for intuitionism but applies equally to any other fallibilist foundationalist account of belief [*ibid.*]) Her second response, however, evinces a position that likely falls afoul of the arguments I have given – namely, a case where an ethical first principle is plausibly understood as being justified *a posteriori*.

Roeser's second response holds that, in cases of intuitional divergence, a time of review is in order. She writes that "...through discussion, evaluation, and reflection we can reconsider our beliefs. This is even more so with moral judgments which are formed not just by a simple impression on our senses, but by evaluating many aspects relevant to the situation" (Roeser 102-103).

Roeser's discussion of Ayer's challenge is sufficiently brief that some aspects of her reply are opaque. But her appeal to discussion, evaluation, and reflection in the face of intuitional divergence is plausibly understood as justifying an ethical first principle on *a posteriori* grounds if it responds to Ayer at all. Let us now see why this is the case.

Consider a distinction familiar to all philosophers of science – namely, the distinction between the contexts of discovery and justification. With regard to the philosophy of science, Arabatzis defines the difference between these two contexts in the following manner: The context of discovery "consists in the processes of generation of scientific hypotheses and theories" whereas the context of justification concerns the "testing and validation" of those hypotheses and theories (Arabatzis 1). For this paper's purposes, however, we need not construe the difference between these contexts in strict scientific terms; the contexts of discovery and

justification are also present when one engages in reflection concerning an ethical first principle. Roeser's suggestion that discussion, evaluation, and reflection should occur following a divergence of ethical intuitions can be interpreted in one of two senses, then: The discussion, evaluation, and reflection she refers to may either occur within the context of discovery or occur within the context of justification.

If Roeser is interpreted as suggesting that discussion, evaluation, and reflection should occur within the context of discovery, then I believe that the advice she offers her readers is good. It is often the case that discovering an ethical first principle can require a great deal of discussion, thoughtful evaluation, and significant reflection. Indeed discussion, evaluation, and reflection which occurs within this context is not incompatible with my own views since my arguments concern the impossibility of an ethical first principle being *justified a posteriori* (rather than its being *discovered* that way).

However, it must be recognized that, so interpreted, Roeser's position does not respond to Ayer's challenge. For as was noted earlier, Ayer's challenge requires a criterion to be given which may be used to decide between conflicting intuitions. Insofar as discussion, evaluation, and reflection occur within the context of discovery, however, they cannot justify an intuition and likewise cannot justify a choice between two competing intuitions. But since Ayer's challenge requires that some criterion be given which justifies our choice of one intuition over another, this interpretation of Roeser's argument cannot respond to Ayer's challenge.

However Roeser's argument can respond to Ayer's challenge if the discussion, evaluation, and reflection she suggests is interpreted as occurring within the context of justification. So understood, after discussing, evaluating, and reflecting in the face of intuitional

divergence, one would choose between competing intuitions based on which one has been the most thoroughly discussed, the most deeply evaluated, or the subject of the most reflection (or perhaps some combination of all three). This would successfully respond to Ayer's challenge by providing a criterion for choosing between conflicting ethical intuitions.

But if interpreted in this manner, Roeser's position clearly falls afoul of the arguments I have given. As I have already shown, it is impossible for an ethical *first* principle to be justified *a posteriori*, and if we justify our choice among ethical intuitions on grounds such as having been the most deeply evaluated, then the ethical first principle is being justified *a posteriori*.

So if Roeser's second argument is interpreted such that it can respond to Ayer's challenge, it does so in a manner which falls afoul of my arguments. But it is certainly worthwhile to note that this interpretation of Roeser's suggested strategy is severable from her wider theory of affective intuitionism. That is, nothing in the theory of affective intuitionism itself entails that one should justify the choice between competing ethical intuitions on *a posteriori* grounds. So the problem I identify with one interpretation of Roeser's remarks could be amended while still maintaining the overall theory of affective intuitionism.

Although I offer a partial resolution to Ayer's challenge, it is worth noting that, at best, what I offer is *partial*. There will still remain many scenarios in which the strategy I offer does not answer Ayer's challenge. (I hope to address some of these remaining scenarios in my own future philosophical works.) But however limited my response is, it is significant inasmuch as most philosophers have not dealt with Ayer's challenge at all. Thus, while a full resolution to Ayer's challenge does not exist, my arguments have made progress towards a full response to Ayer. Let us consider some cases to discover what, precisely, my arguments entail.

First, let us imagine a scenario where two persons have divergent intuitions concerning some ethical matter, but both persons neither affirm nor deny the existence of an ethical first principle. My solution does not bear on this situation since both persons are agnostic about the existence of any ethical first principles. So in a scenario such as this, my solution is not applicable and Ayer's challenge still stands.

Second, consider a case where two persons have divergent ethical intuitions because both accept the existence of differing first ethical principles. For example, perhaps one's intuitions derive from Kantianism and the other's derive from utilitarianism. In this scenario, too, my solution is not applicable and Ayer's challenge remains unanswered.

Third, consider a case where two persons have divergent ethical intuitions stemming from the fact that both accept the existence of an ethical first principle justified *a posteriori*. To meet Ayer's challenge, we would have to possess a criterion to adjudicate between their conflicting intuitions. Can we do so? Interestingly, their intuitions cannot be adjudicated between because *both* of their intuitions are incorrect, and there can be no adjudication between two sets of beliefs when both are false.

To be more precise, Ayer's challenge requires that we "...provide some criterion by which one may decide *between* conflicting intuitions" (Ayer 106, emphasis mine). If taken in its strictest sense, this seems to require that we have some criterion by which at least one of the two competing intuitions is accepted. But my arguments, if successful, have shown that a necessary condition of a proposition's being an ethical first principle is that it must be justified *a priori*. These two interlocutors, then, are in a unique and somewhat bizarre case where we can know

that *both* are wrong – but since both are wrong, we cannot adjudicate *between* them (if betweenness is understood to require choosing at least one of the proffered intuitions).

However, if Ayer’s challenge is understood in a more natural sense (such that the criterion by which we decide may licitly reject *both* intuitions), then this scenario is one where Ayer’s challenge can be met: My arguments justify the rejection of both intuitions.

I shall close by addressing a more general scenario: Imagine one ethical first principle is justified *a priori* and another is justified *a posteriori*. Should we automatically think that the former is more plausible than the latter?

I answer that the former should be considered more plausible than the latter when considered *as a first principle*. That is to say, a principle justified *a posteriori* may be perfectly plausible (many beliefs justified *a posteriori* are), but a belief justified *a posteriori* cannot serve as an ethical first principle. A belief justified *a priori* may be implausible (since many are), but beliefs justified *a priori* fulfill a necessary condition of being an ethical first principle. Thus, *when considered as an ethical first principle*, I affirm the view that beliefs justified *a priori* are automatically more plausible by virtue of fulfilling a necessary condition of being an ethical first principle.

But an important limitation on this answer to Ayer’s challenge should be highlighted: It is a necessary condition that an ethical first principle be justified *a priori*, but this does not suffice to make such an ethical first principle *true*. After all, some ethical first principles justified *a priori* are false. For example, consider that utilitarianism and Thomism cannot both be true but both nonetheless offer an ethical first principle justified *a priori*. Consider also any number of hypothetical ethical systems based on false ethical first principles – but whose first ethical

principles are justified *a priori*. (So an ethical system whose first principle involved the rightness of sexism would be a false ethical system even if its first principle were justified *a priori*.)

There are two main objections which could be lodged against my view. The first concerns resolutions of disputes where one interlocutor does not accept the existence of an ethical first principle. The second concerns a purported implication of my view: That it entails the rejection of reflective equilibrium. These shall be examined in turn.

#### **IV: Objection Concerning First Principles**

Of all objections which could be made against this argument, the most serious one concerns the rejection of an ethical first principle's existence. My analysis may succeed if two philosophers disagree concerning an ethical first principle, but what if one or both reject the existence of an ethical first principle in the first place?

There are two ways one might reject the existence of an ethical first principle. First, one may do so in virtue of rejecting the existence of *any* first principles (and thus, by implication, reject the existence of an *ethical* first principle). Second, one may accept the existence of some first principles but argue that none of those first principles concern ethics. These two approaches shall be addressed in turn.

If one rejects the existence of ethical first principles *via* his antecedent rejection of *any* first principle's existence, then ultimately this disagreement concerns not my position but Hales'. My own work here is, properly speaking, an extension of Hales' argument that self-justifying propositions exist (or, if not, that rational intuition is bunk). To properly follow this line of attack, one would need to either argue that Hales' conclusion does not follow from his premises

or argue that one or more of his premises are false. It is then necessary to either accept the (I think) disastrous consequences Hales identifies or show how they do not follow. But however followed, this line of attack ultimately concerns a rejection of Hales' position (and only implicates my arguments by extension).

But if one agrees with Hales that at least one self-justifying proposition exists, then rejecting my thesis entails that one maintain the following position: While at least one self-justifying proposition exists, no self-justifying *ethical* propositions exist. If someone were to maintain this position, how would I respond?

I would begin by noting that the burden of proof cannot rest upon my position but rather upon the one making such an assertion, inasmuch as *many* divergent ethical theories assert the existence of an ethical first principle. Of course, it is possible that Stoicism, Aristotelianism, Theravāda Buddhism, Thomism, Utilitarianism, and Kantianism are *all* wrong (since they all rely on the existence of some ethical first principle). However, given the wide sweep of cultures, time periods, and significant thinkers encompassed by these schools of ethics, the burden of proof must lie with whoever rejects the foundational principle on which all these schools rest.

If the burden of proof lies with my opponents, then, it will necessitate their giving some argument demonstrating that while at least one self-justifying first principle exists, self-justifying *ethical* principles do not. Whatever argument given to that effect is likely to be highly controversial. As an example, one could appeal to widespread cultural variations as grounds for rejecting the existence of any ethical first principle while accepting the existence of at least one first principle. But, obviously, ethical relativism is hardly unproblematic in its own right (and

whether it successfully avoids all reliance upon an ethical first principle itself is an open question).

### **V: Objection Concerning Reflective Equilibrium**

The second major objection to the position I've argued for concerns reflective equilibrium. It might seem that since my position embraces moderate foundationalism it must thus reject reflective equilibrium (which is typically understood to be coherentist in nature). While foundationalism (as an epistemic position) and reflective equilibrium (as a philosophical methodology) might seem to be separate topics, they are not wholly disparate. Rather, as I shall explore, they are related. But given that they are related, one then might easily wonder how far a foundationalist epistemology can go in adopting a seemingly coherentist methodology like reflective equilibrium. I shall argue that my position does not entail a rejection of either narrow or wide reflective equilibrium. I shall begin by briefly examining what reflective equilibrium is in both its narrow and wide variants. This will set the stage for, second, exploring two different reasons why a wholesale rejection of reflective equilibrium might seem to be necessitated by my views. Finally, I shall argue that, in each case, no such wholesale entailment follows.

What is reflective equilibrium? Daniels provides a helpful initial understanding of the concept of reflective equilibrium when he writes:

The method of reflective equilibrium consists in working back and forth among our considered judgments (some say our "intuitions") about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them. The method succeeds and we achieve reflective equilibrium when we arrive at an acceptable coherence among these beliefs (Daniels 1).

Reflective equilibrium, thus, is a method of seeking out and procuring an acceptable degree of coherence among our many judgments (particularly, for present purposes, our ethical judgments). Note that the paradigm of reflective equilibrium is not merely limited to considering particular cases to form judgments about them but that it also entails examining the rules behind our particular case judgments to, once again, obtain an acceptable degree of coherence.

Daniels' broad summary, however, makes no mention of either wide or narrow reflective equilibrium. The distinction was initially made by John Rawls. To summarize him, narrow reflective equilibrium requires that we construct moral theories by screening moral judgments in order to eliminate inferior views (views exemplified by, for example, those based on incomplete information or bias, etc.). Wide reflective equilibrium, however, potentially requires a disruption of narrow reflective equilibrium by deliberately considering alternative moral theories and attempting to construct arguments which will adjudicate between them (Rawls 19-21, 48-51).

There are two main reasons why reflective equilibrium might seem to conflict with the views I hold. First, it might seem as if reflective equilibrium fits better within a coherentist account of justification rather than one of moderate foundationalism. Second, since reflective equilibrium allows beliefs held *a posteriori* to affect a belief held *a priori* about a first ethical proposition, it might seem as if moderate foundationalism is consequently undermined. These objections shall be examined and refuted in turn.

Broadly speaking, both wide and narrow reflective equilibrium as methodologies may seem to fit within a coherentist account of justification better than within a foundationalist one. If two separate persons engage in the process of obtaining narrow reflective equilibrium, there is no presumption that one person's moral judgments will be the same as the other person's. But

the opposite would be expected if narrow reflective equilibrium were foundationalist, and thus narrow reflective equilibrium might well seem coherentist in nature.

Similar reasoning will serve to show why wide reflective equilibrium might also be thought of as supporting a coherentist rather than foundationalist understanding of justification. When deciding between competing moral theories (as wide reflective equilibrium necessitates), as DePaul notes, one does so by making use of what he refers to as “background theories.” These background theories are our own views which are prior to and, consequently, inform the moral theories we accept. Thus, one might have a background theory about human nature or the role of moral theories in society. Then when attempting to establish wide reflective equilibrium these background theories are consulted to help adjudicate between competing moral theories. But conflict between background theories and moral theories are not necessarily resolved in favor of the former (or the latter, for that matter). Consequently, “A point of wide reflective equilibrium is reached when the considered moral judgments, and moral and background theories one accepts are coherent and seem more likely to be correct to one than any alternatives one has considered” (DePaul 59).

Given the seemingly coherentistic nature of both narrow and wide reflective equilibrium, as well as the obvious importance of the methodology itself, it might seem as if moderate foundationalism entails the wholesale rejection of reflective equilibrium. However, this is not so. Certainly narrow reflective and wide reflective equilibrium can be obtained in the fashion outlined above – and if done *in that manner*, then narrow and wide reflective equilibrium will be coherentist in nature. However, one can also imagine other processes of narrow and wide reflective equilibrium which are consonant with a foundationalist account of justification.

For example, one can imagine holding an ethical first proposition and then engaging in a kind of narrow reflective equilibrium to discover which other views cohere with it – and rejecting those which do not cohere with the ethical first proposition. Alternatively, one could engage in narrow reflective equilibrium and make the resulting ethical views one’s ethical first propositions. So while narrow reflective equilibrium can be consonant with a coherentist account of justification, it need not only be so. Consequently, these foundationalist methodological considerations demonstrate how narrow reflective equilibrium can easily coexist with moderate foundationalism.

Wide reflective equilibrium with respect to ethics is similar. Within it, one’s entire moral paradigm is to be weighed against another moral paradigm based on various background theories. One can proceed as described above and hold neither the background theories nor the moral paradigms as constants. The end result of this methodology will be a fundamentally coherentist system of justification. If, however, the ethical first proposition is regarded as foundational (and consequently that any background theories or subsequent ethical propositions must be brought into conformity with it) then the resulting moral paradigm will show the compatibility of wide reflective equilibrium and foundationalism.

Having concluded this matter, we shall now turn our attention to the second reason why moderate foundationalism might seem to require the rejection of reflective equilibrium. As I have argued, any ethical first proposition ought to be a belief held *a priori*. But the reflective equilibrium deliberative process allows beliefs held both *a priori* and held *a posteriori* to enter into evaluation. If one holds an *a posteriori* belief and this belief consequently affects the adoption of an ethical first proposition, then it seems as if the foundational ethical first proposition is no longer held *a priori* but is instead held *a posteriori*.

As before, such an argument only serves to demonstrate that a certain understanding of reflective equilibrium is incompatible with moderate foundationalism, not that the entire methodology of reflective equilibrium is incompatible with moderate foundationalism. Moderate foundationalism, as discussed before, is the belief that “There are justified propositions whose justification depends on nothing other than themselves” (Hales 145). If a belief held *a posteriori* serves a *justificatory* role for the ethical first proposition, then it is correct to say that the ethical first proposition is itself held *a posteriori*. This particular kind of reflective equilibrium, then, is not supported by my argumentation.

However, of course, reflective equilibrium need not be understood as operating in this fashion. There are other relationships which two or more propositions can have other than that of justification, and if a belief held *a posteriori* relates to the ethical first proposition in a non-justificatory sense, then the ethical first proposition is still held *a priori* and moderate foundationalism is not undermined. One could use reflective equilibrium, for example, to discover that a relationship of consonance (or dissonance) exists between a given belief held *a posteriori* and the ethical first proposition held *a priori*. This should demonstrate that, as before, a certain type of reflective equilibrium is not compatible with moderate foundationalism – but the entire methodological process need not be rejected.

Consequently, while reflective equilibrium alone cannot generate a first ethical proposition, a foundationalist epistemology can indeed make use of the method of reflective equilibrium. Indeed, once the ethical first proposition is established, narrow reflective equilibrium will play a crucial role in ensuring that all our various judgments cohere with that ethical first proposition. Consequently, rather than the methodology of reflective equilibrium

being at odds with foundationalism, it is rather the case that reflective equilibrium will continue to be significant in the practice of philosophy.

In conclusion, I have shown that a partial answer to Ayer's challenge is possible, specifically within a subset of the philosophical subfield of ethics. Inasmuch as first principles exist, and inasmuch as one ethical first principle exists, I have identified a necessary condition in order for one's intuitions concerning that ethical first principle to be true. Although this does not fully answer Ayer's challenge, I have offered a valuable step forward towards resolving an overarching challenge presented against the role of intuitions in the practice of philosophy.

## Works Cited

Arabatzis, Theodore. "On the Inextricability of the Context of Discovery and the Context of Justification." *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*. Ed. Jutta Schickore and Friedrich Steinle. Dordrecht: Springer, 2006. 215-30. Print.

Ayer, A. J. *Language, Truth, and Logic*. New York: Dover Publications, 1952. Print.

Daniels, Norman, "Reflective Equilibrium", *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), Online.

DePaul, Michael R. "Reflective Equilibrium and Foundationalism." *American Philosophical Quarterly* 23.1 (1986): 59-69. Print.

Hales, Steven D. "The Problem of Intuition." *American Philosophical Quarterly* 37.2 (2000): 135-47. Print.

Kornblith, Hilary. "The Role of Intuition in Philosophical Inquiry: An Account with No Unnatural Ingredients." *Rethinking Intuition*. Ed. Michael R. DePaul and William M. Ramsey. Lanham, MD: Rowman & Littlefield, 1998. 129-42. Print.

Mill, John S. *Utilitarianism*. London: Longman, Green, Reader, and Dyer, 1867. Print.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap of Harvard UP, 1971. Print.

Roeser, Sabine. *Moral Emotions and Intuitions*. Basingstoke: Palgrave Macmillan, 2011. Print.