

## The Audacity of Hopeful Philosophical Appeals to Intuition

Jonathan Ichikawa  
Arché Philosophical Research Centre  
The University of St. Andrews

ichikawa@gmail.com  
Draft: 8 July, 2009

DRAFT IN PROGRESS—DO NOT CITE UNPERMITTEDLY  
COMMENTS VERY WELCOME—FURTHER EXAMPLES PARTICULARLY SOLICITED

**Abstract.** According to some critics, traditional armchair philosophical methodology relies in an illicit way on intuitions. But the particular structure of the critique is not often carefully articulated—a significant omission, since some of the critics arguments for skepticism about philosophy threaten to generalize to skepticism in general. In a recent paper, Jonathan Weinberg has articulated a version of the critique that is sensitive to this worry: traditional philosophical methods are importantly flawed, Weinberg says, in that they are in a technical sense *hopeless*. I argue that Weinberg’s critique of traditional methods is insufficiently attentive to the practices in which philosophers actually engage. I will conclude, with reference to a number of philosophical case studies, that contrary to Weinberg’s claim, philosophical practices are rather hopeful indeed.

### §1. Against Armchair Philosophy

Some philosophical stormy petrels will tell you that something is deeply wrong with traditional armchair philosophical methodology. Traditional philosophy relies, it is thought, on armchair intuitions in an illicit way. Different flavors of the critique have presented different challenges to traditional methodology; some arguments against the use of intuitions were more or less *a priori*: Cummins (1998) argued that intuitions were poor candidates for philosophical evidence, since, unlike more respectable scientific sources of evidence, they are not susceptible to external calibration. Stich (1990) gave an *a priori* argument against the use to which intuitions are put, based on thought experiments involving possible cultures with different intuitions and norms than ours. More recent work in a similar vein has presented empirical evidence that such cultural divergence in intuitions is indeed actual,<sup>1</sup> and also that intuitions seem to be manipulable in surprising ways.<sup>2</sup>

It is not always obvious in the presentations of these critiques, however, in what exactly the argument against armchair methodology is meant to consist. In a prototypical argument form, a critic suggests a psychological or sociological fact about intuitions, motivated by surveys of undergraduates, and we’re told that this fact presents a problem for traditional philosophy, but we’re not told how. For example, Machery, Mallon, Nichols and Stich (2004) tell us that they “suspect that most philosophers ... assume the Kripkean intuitions [about reference in the Gödel-Schmidt case] are universal,” and that evidence to the contrary shows that this assumption is “spectacularly misguided,” and that as a consequence, “philosophers must radically revise their methodology.” But they do not argue for either the truth or the

---

<sup>1</sup> Weinberg, Nichols, and Stich (2001); Nichols, Stich, and Weinberg (2003); Machery, Mallon, Nichols, and Stich (2004).

<sup>2</sup> Swain, Alexander, and Weinberg (2007).

importance of this claim about philosophers' expectations, and they do not explain why these facts about cultural disagreement force philosophers radically to revise their methodology.

And indeed, as defenders of armchair methodology have emphasized, the skeptical challenge pressed by these experimentalist critics, if it is not presented in a suitably subtle way, runs the risk of providing arguments that result in very general forms of radical skepticism.<sup>3</sup> Although the refutation of radical skepticism represents an important project in epistemology, few philosophers today—and certainly not the science-loving experimentalist critics in question—think that radical skeptical arguments shows that there is something deeply wrong with our ordinary belief-forming practices. The challenge for the experimentalist critique is to articulate a problem for the target philosophical methodology that does not generalize to their preferred practices, including, at least, a great deal of science, as well as the use they themselves put to philosophical considerations in their own arguments against armchair philosophy. Jonathan Weinberg (2007) addresses this challenge in an admirably direct way. He sets himself the task of identifying an “epistemically deleterious characteristic” that applies to the target philosophical practice, but does not “apply to too large and important a set of our evidential sources”, or to the considerations that he himself brings to bear in his paper.<sup>4</sup> If he succeeds in identifying such a deleterious characteristic, then he has the resources to reject traditional philosophy without providing arguments that generalize to widespread skepticism.

I contend that Weinberg does not so succeed. The case against traditional philosophical armchair practices is not stronger than that against many obviously-respectable practices in obviously-respectable fields. Earlier, more naïve formulations of the experimentalist critique raised challenges that many respectable forms of inquiry cannot meet; Weinberg's newer, more sophisticated charge can be met by those uncontroversially respectable disciplines; so too can it be met by philosophy.

## §2. Hope

Weinberg identifies the ‘epistemically deleterious property’, which he hopes to apply to his target instances of philosophical methodology and little else, as ‘hopelessness’. Any use of a fallible source of evidence is assessable not only for reliability—how likely it is to issue into truths?—but also for hopefulness—in the inevitable event that the use of the evidence issues into an error, how well able are we to recognize and correct for the error? Weinberg gives us a definition and a principle:

A source of evidence is hopeful to the extent that we have the capacity to detect and correct its errors. ... [W]e similarly stipulate the term *hopeless* for devices for which our practices lack an appropriate sensitivity to their errors, and capacity for correction when such errors are found. In general, a source of evidence will be hopeful to a greater or lesser degree, but I will reserve “hopeless” for those that are pretty far down at the low end of the spectrum.

And:

Any putative source of evidence that is hopeless ought not to be trusted. (327)

---

<sup>3</sup> For arguments to this effect, see e.g. Bealer (1992), Pust (2000), Sosa (forthcoming).

<sup>4</sup> p. 323.

Weinberg goes on to argue that philosophical practices are hopeless, but that many other, more respectable practices—such as science, and the epistemology being executed in his paper—are not. I have no wish to dispute the highly plausible observation that hope, in Weinberg’s sense, is an epistemic virtue. It is better to be hopeful than to be hopeless. We may notice, however, that Weinberg’s maxim doesn’t quite follow from this truism; it’s consistent with hopefulness being a good property, and hopelessness being a bad property, that sometimes, it’s nevertheless appropriate to trust hopeless sources of evidence. This would probably be an appropriate response if, for instance, it turned out that all evidence is hopeless. Nevertheless, I am prepared to grant to Weinberg that the demonstration that philosophical practices are on the whole hopeless would be a series blow to traditional philosophical methodology. I am accepting the challenge to identify hope for armchair methods.

### §3. Philosophers’ Appeals to Intuitions and ‘PAI’

A naïve version of the experimentalist challenge might run thus: philosophers use intuitions as evidence, but intuitions are hopeless, therefore, philosophers are mistaken. Weinberg makes it clear that his position is much more subtle than this heavy-handed one; he emphasizes, for instance, that not all intuitions are put into doubt, and he describes his target as a particular philosophical practice, rather than on the alleged evidence itself. As he puts it: “Instead of thinking in terms of a problem with something philosophers *have*—the intuitions themselves—I would suggest that we turn our attention to something philosophers *do*: the current analytic philosophical practice of appealing to intuitions as evidence for philosophical claims.” (320) Both distinctions are good ones in principle, but neither, I think, is as clear as might be hoped. It is worth examining them with some care.

Weinberg is happy to admit that not all intuitions are inappropriate to serve the uses to which philosophers and others put them. He does not understand his own view to be in substantial disagreement with those who claim that intuition in general is *prima facie* trustworthy, perhaps because they’re partly constitutive of our concepts, or because to reject intuition *tout court* would be to capitulate to radical skepticism. These *generalist* defenses of intuition, Weinberg claims, are irrelevant to the best formulation of the experimentalist critique, which has “raised specific charges against a *particular* flavor of philosophical intuition.” (319) What is that flavor? Here, Weinberg is less specific than one might like. Here is what Weinberg writes:

[The] challenge arises from the concern that philosophers are perhaps a bit too liberal in the class of seemings that they deem acceptable candidates for this dialectical move. Intuitions may be fine as a class, taken on the whole, and the opponent has neither the need nor the desire to attack that whole class. But philosophers do not invoke a vast undifferentiated mass of intuitions in defense of their claims—rather, we cite particular intuitions about particular hypothetical cases. And the opponent is concerned that some significant number of these cases may be far less than ideal for this sort of appeal. For the practice appears to set no constraints on how esoteric, unusual, far-fetched, or generally outlandish any given case may be. Everyone is familiar with the likes of Davidson’s Swampman and Seale’s Chinese room, but one can look at the very recent literature and find the likes of double-lesioned testifiers, new evil demons, and fissioning/fusioning/teleporting pairs (or are they?) of persons. So this anything-goes aspect of the practice is what makes it particularly ripe for the opponents’ challenge. ... I hope it makes clear that it may turn out to be a practice of appeal to a particular sort of intuition that gets challenged, and not a broad swipe at all of intuitionkind. I will refer to this practice as the “philosophers’ appeals to intuitions,” or just PAI for short. (321)

This passage warrants some scrutiny. Weinberg's name for the target practice, "philosophers' appeal to intuitions", will, if one treats it carelessly, mislead. For as Weinberg emphasizes, not all philosophical appeals to intuitions are part of PAI. PAI comprises some "particular flavor" of invocations of intuitions by philosophers, and it is not yet obvious, I think, just what is included. Philosophers sometimes rely on esoteric cases, and sometimes rely on quotidian ones. Which are part of PAI? Clearly, the esoteric intuitions, such as the ones given in his list, are meant to be included. What of the quotidian ones? Here, Weinberg's remarks generate some tension. On the one hand, Weinberg's treatment of quotidian intuition throughout the paper seems generally positive, and he considers, apparently favorably, a revised methodology that strips traditional practice only of the esoteric cases. (336) On the other hand, if the target is, as he says, a *practice*, and the practice is, as he says, an "anything-goes" one, then it should be, one would think, that the practice also includes appeal to quotidian intuitions. If *anything* goes in PAI, well, aren't the everyday judgments part of 'anything'? Some of Weinberg's statements suggest this broad view, as when he says that the target is the "practice of appealing to intuitions as evidence for philosophical claims." At other points, however, it seems clear that he has a much more restricted set of intuitions in mind. The following intuitions, Weinberg explicitly claims to be exempt from his critique:

- The 'ordinary application of concepts to particulars' (320, 335)
- The 'claim that no object can be red all over and green all over' (320)
- Elementary mathematics (320)
- At least some philosophical intuitions (320)
- The epistemic intuitions that justify the importance of hopefulness
- Trained expert judgments, like those of chess grand masters or medical diagnosticians (335)
- Logic (339)
- Folk psychology (339)
- That truth is epistemically valuable as a guiding norm (340)
- That if a norm is well established, then the burden is on someone who wants to reject it. (340)

One worries whether there is any principled distinction between PAI and the use of the intuitions listed above, which Weinberg does not mean to challenge. The question is important, because the restricted nature of PAI plays a central role in Weinberg's attempts to prevent the potential overgeneralization of his skeptical arguments, as for example when he writes: "Bealer's claim [that intuitions are necessary for anti-skepticism] could be true and still PAI could be legitimately challenged if one can appeal to the intuitions that *ex hypothesi* are needed to justify the principle without thereby participating in PAI." (323, referencing Bealer (1992).)

Why think that PAI is hopeless? Weinberg will identify four criteria that contribute towards hopefulness of an evidential source—I'll consider these in some detail below. No one, he thinks, is a necessary condition for hopefulness, but he thinks that his criteria exhaust the ways in which an evidential source can be hopeful, and therefore that any alleged source of evidence with none of these four features would have to be hopeless. This, he thinks, is the unhappy state of PAI. As indicated above, one worries about the intended scope of PAI; has Weinberg gerrymandered an artificial category to serve as the target of his critique? If so, it is not clear what force the critique has against armchair philosophy, as practiced. Weinberg recognizes the worry:

[B]efore arguing that PAI is currently hopeless, it is worth considering the question of

whether my restriction to that target is legitimate—of if I run the risk of running afoul of [the condition that] ... we not impeach too significant a fraction of our cognitive resources. With a general framework for the evaluation of hopefulness on the table, we can see how it could make sense to attack PAI without thereby catching everything that might reasonably be called “appeal to intuition” in the fire. Trained expert judgments, for example, like those of chess grand masters or medical diagnosticians, lie outside the range of my critique; so too do our judgments in most ordinary cases that some particular object or event falls under a particular concept. I am not attacking such intuitions as those not because they are immune to worries about hopefulness, but because by and large they are in fact hopeful. Both expert judgments and ordinary categorizations usually possess a great deal of external corroboration and internal coherence, for example. The practices involving appeal to intuitions such as those have resources of hope that (I will shortly argue) are lacking in PAI. (334-35)

Weinberg here unambiguously restricts PAI to a certain subset of traditional appeals to intuition, but he says little more than before about what exactly that subset comprises, or why it in particular is a legitimate target. Where one draws the lines matters. Hopefulness is a feature of a practice, not a particular judgment. So when we’re evaluating whether a judgment is the product of a hopeful practice or not, the individuation of practices makes all the difference in the world. (Compare the ‘generality problem’ for the individuation of methods for process reliabilism.) Take a particular philosophical judgment on the basis of some putative evidence: for concreteness, take the judgment I made in 1999, in response to a thought experiment about Wimbledon on television, that justified true belief is insufficient for knowledge. Whether that judgment was made as part of a hopeful practice or not depends critically on how one draws the boundaries of the practice. I was participating in a number of practices: I was applying an ordinary concept to a particular case; I was trusting my unreflective judgments about improbable hypotheticals; I was accepting the testimony of my professor; I was accepting known testimony; I was engaging in a traditional argument in armchair philosophy; I was thinking carefully and using my best judgment; I was judging on a Tuesday; I was trusting my judgments and also accepting any insights that seemed to be spoken to me directly from God; I was thinking about philosophy in the morning; etc. Maybe I was also engaging in PAI—Weinberg doesn’t tell us whether he thinks Gettier judgments are included in PAI, but one suspects from his other work, and from his citation here of that work, that he thinks so. Plausibly, some of the practices that my judgment can be seen as a part of are more hopeful than others. Without some privileged classification of the relevant practice, then, identifying one hopeless practice that a target judgment partially comprises does little to impugn the judgment, or the actual practices, taken on the whole, that constitute traditional analytic philosophy. The worry is especially acute if, as in this case, we aren’t even given a positive characterization of the target practice. Weinberg’s PAI seems to be picked out only negatively: “appeals to intuition, except for things like this list of intuitions, which I don’t want to dispute.”

I believe, however, that even if we let Weinberg focus only on the pieces of philosophical methodology he intends—even if we concede to him the right to gerrymander whatever practice he likes—he still dramatically underestimates the extent to which PAI has the resources to claim hopefulness. Let us consider Weinberg’s four sources of hopefulness in turn, before considering whether they can be thought to apply even to a gerrymandered practice of PAI.

#### **§4. Sources of Hope**

Weinberg’s proposed four sources are:

#### 4.1. External Corroboration

Hopeful sources of evidence are those whose results can be compared against results obtained independently from those sources. Ideally, the target evidence shouldn't be the only source one has for the purported knowledge. I see my colleague across the room; I can also hear her, and I can also explain and predict her presence via independent background knowledge. So the evidence my eyes provide can, in many cases, be corroborated by extra-visual evidence. This confers some degree of hopefulness to my use of visual evidence, since very often, if my eyes deceive me, I will discover extra-visual reason to suspect their deliverances. If I seem to see someone, for example, whose presence I cannot verify with my other senses, and whose presence my background knowledge gives me strong reason to doubt—I have a hallucination of Abraham Lincoln, say—I know better than blindly to trust the visual inputs.

#### 4.2. Internal Coherence

The deliverances of an evidential source ought to connect with one another in a coherent way. This provides a means of internal corroborative verisimilitude: if my evidential source tells me something inconsistent with other things that it has told me, this provides me reason to be suspicious of the outlier, if it is one, or of the source on the whole, if all I have is a mutually inconsistent set with no means of distinguishing the proper verdict. Weinberg, it is worth noting, is using 'internal' in this context in a sense that includes the deliverances of a method among different members of a shared intellectual practice. He cites, as an example of the internal coherence of the scientific use of electron microscopes, agreement both "within and across laboratories". So potential corroboration by internal coherence may obtain in either intersubjective or intrasubjective forms. If there is general internal coherence, it will be easier to detect the outliers that represent errors. Plausibly, then, internal coherence does contribute to hopefulness.

#### 4.3. Detectability of Margins

The most hopeful sources of evidence provide outputs of varying strengths, with some appreciable—and indeed appreciated—correlation between strength of signal and reliability. Weinberg offers two helpful examples: "visual perception in bad conditions very often yields percepts that indicate that badness to us. We can find the value of a richness of output in electron microscopy as well, in that a poorly done scan will more likely look like nothing at all than it is to look like something clear but wrong; or the other way around, when the procedure gives clear results that clarity is itself evidence in support that the procedure's results are correct." (330) When practices encode sensitivity to margins of reliable signals of evidential sources, that contributes to hopefulness in an obvious way.

#### 4.4. Theoretical Illumination.

The better we understand how an evidential source works, the better able we will be to recognize its failures. An explanation for how and why a source makes correct deliverances when it does might well have relatively direct implications as to in which circumstances it is particularly likely to be correct, and in which circumstances it is likely to err. Here is a toy example: suppose I have a smoke detector and am entirely ignorant as to the mechanism upon which it operates, and know only that it is a reliable indicator of fire. My practice—treating the detector's activation or silence as evidence for or against fire, respectively—is improved if I learn that it operates by detecting small particles of smoke in the air. This knowledge will help me to avoid false positives, for example, when I burn my dinner, creating smoke with no fire.

None of these four criteria, Weinberg admits, are essential to hopefulness. But collectively, he suggests, they exhaust the ways in which an evidential practice can be hopeful: “certainly any practice for appealing to a source of evidence that lacked all four would be well night hopeless.” And, he claims, PAI fails in all four dimensions. In my view, even if the lines are drawn around PAI where he intends them, the resources that traditional philosophy has and makes use of include all four hope-making features. Weinberg’s arguments notwithstanding, philosophical methodology, including PAI, is rather hopeful indeed. Let us examine Weinberg’s arguments to the contrary.

## §5. External Corroboration and PAI

With respect to external corroboration, Weinberg writes:

As for external corroboration, I take it to be clear that philosophical intuition has not had very much of this. By and large, we have not found much other access to the particular propositions in question than the intuitions themselves. For many of the domains in question (e.g., the metaphysics of modality) there simply may be no area outside of philosophy that really can speak to it, and for many of these domains we simply have not found any other way of investigating them other than intuition-mongering. (338-39)

The argument, compelling on its face, is viciously equivocal. Recall that Weinberg’s target is PAI—a limited subclass of philosophical invocations of intuition. Many philosophical appeals to intuition are not a part of PAI. But this passage is only compelling against the externally corroborative resources available to PAI if one equivocates between the two. Even if Weinberg is right that there is no access to the relevant facts outside of philosophical intuition, there might well be access to them outside of the more restrictive set of practices that is PAI. And the deliverances of this extra-PAI evidential use can, and plausibly does, serve as external corroboration for the deliverances of PAI.

### 5.1. Extra-PAI corroboration from Philosophers’ Appeals to Intuition

Constructing an example is less straightforward than one might like, since the boundaries of PAI are not offered in any precise way; however, we may begin, at least, on firm ground, by considering practices that Weinberg has explicitly ruled out of PAI. For example, as indicated above, PAI does not include the ‘ordinary application of concepts to particulars’. We can know, then, without any invocation of PAI, of many particular cases, whether they are cases that fall under some concept or not. It is no part of PAI to judge that this shooting caused that death, or that this subject knows this proposition, and is justified in believing that one, or that this word refers to that person, etc. So we do have some access, *external to PAI*, to these prototypically philosophical subject matters. We also, presumably, have some access to theory-building methods, independent of PAI. Weinberg never explicitly says so much, but since theory construction is such an important part of scientific methodology, which Weinberg is careful not to impugn, it is difficult to imagine these transitions from particular data points to theories as the sort of thing that is included in the allegedly hopeless PAI.

So now, before ever invoking PAI, we have as evidence facts about particular cases, and abilities to abstract from them into general principles. Now we have, external to PAI, the resources to corroborate judgments made within PAI. If consideration of mundane cases, along with by theory-building capacities, leads me, independent of PAI, to conclude that knowledge is factive, then I have an independent corroboration for a PAI-type judgment that, say, a subject in an evil demon world where the demon re-arranges the subject’s memories every Tuesday morning, who believes himself not to be in an evil demon world, doesn’t know that he’s not in an evil demon world. I’ve made the case baroquely fantastic in order to

try to ensure its inclusion in PAI; but since Weinberg doesn't give either sufficient conditions for PAI membership or a rich set of examples, I cannot be sure whether I've succeeded. However, it should be clear how to generalize the point for any judgment about knowledge, or anything else that we have some non-PAI access to.

So, philosophers very often have resources available to achieve hopefulness for PAI via corroborate PAI-external corroborative resources. Perhaps Weinberg thinks that philosophers only very rarely exploit such potential for hopefulness. I believe, however, that it much more the rule than the exception. The following case studies do not seem to be at all anomalous.

## 5.2. Case Study: Goldman on Reliabilism

Here is a historical example of the interplay, in traditional armchair analytic philosophy, between non-PAI judgments and (plausibly) PAI ones. The point is not to judge in favor of or against the proposed non-PAI theory or PAI judgment, but merely to recognize that they are not treated independently; that the extra-PAI theory, in practice, bears on the PAI judgment, thus refuting the general claim that PAI is not susceptible to external corroboration. I am confident that many more examples like this one can be found.

Alvin Goldman (1979) considered quotidian examples of justified beliefs and unjustified beliefs, with an eye toward theorizing about justification. Goldman writes:

Granted that principles of justified belief must make reference to causes of belief, what kinds of causes confer justifiedness? We can gain some insight into this problem by reviewing some faulty processes of belief-formation, *i.e.*, processes whose belief-outputs would be classed as unjustified. Here are some examples: confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch or guesswork, and hasty generalization. What do these faulty processes have in common? They share the feature of *unreliability*: they tend to produce *error* a large proportion of the time. By contrast, which species of belief-forming (or belief-sustaining) processes are intuitively justification-conferring? They include standard perceptual processes, remembering, good reasoning, and introspection. What these processes seem to have in common is *reliability*: the beliefs they produce are generally true. My positive proposal, then, is this. The justificational status of a belief is a function of the reliability of the process or processes that cause it, where (as a first approximation) reliability consists in the tendency of a process to produce beliefs that are true rather than false. (338)

Goldman here seems to be relying only on the general application of concepts (justifiedness) to particular quotidian sorts of beliefs, and a general ability to recognize underlying similarities between alike beliefs, and to generate theories that cite those underlying features. There is here nothing that falls under PAI. Extra-PAI methods have given Goldman a general principle about justification of beliefs. This is not to take a stand as to whether or not Goldman's process reliabilism is correct. The point here is merely that it is a not-obviously-bad philosophical product of an extra-PAI methodology; it gains some support in a way independent of PAI. If this fact plays a moderating influence on intuitions that are invoked as part of PAI, then PAI is to some degree moderated via external corroboration. And indeed, it does appear to do so. In Goldman's own continuation of the discussion of reliabilism, he considers a less quotidian intuitive counterexample: he imagines "a possible world where a benevolent demon so arranges things that beliefs formed by wishful thinking usually come true." (342) Let's assume that intuitions about such worlds are part of PAI. Intuitively, Goldman concedes, beliefs formed by wishful thinking in that world are not examples of justified belief. He does not, however, reject the theory on this basis; the extra-PAI theory moderates the intuition, and Goldman is prepared to take seriously the idea that the PAI intuition is simply mistaken; that in such a strange world, wishful thinking can be a justified

belief-forming method. The PAI intuition, in coming into conflict with the non-PAI theory, is discounted. PAI is susceptible to external corroboration.

Of course, the moderating influence of PAI intuition and extra-PAI theory runs both ways; theories are sometimes modified on the basis of PAI intuitions, too. Goldman also takes suggestions along these lines seriously: he considers indexing the relevant reliability as reliability in 'normal' worlds, or in the actual world, or in what we believe to be the actual world. This is, of course, entirely consistent with the corroboration and moderation of PAI by external theorizing; external theorizing is also corroborated and moderated by PAI. This is just what one should expect for a well-integrated, hopeful practice.

### 5.3. Case Study: Responses to Goldman: The New Evil Demon Problem

A similar pattern emerged in responses to Goldman's work. Consider, for example, Lehrer and Cohen's (1983) 'New Evil Demon Problem'. (This is a particularly fine example since it is one of the few bits of philosophy explicitly mentioned by Weinberg as part of PAI.) Their argument against Goldman's reliabilism is related to the one of Goldman's own that I've just mentioned. Lehrer and Cohen invite us to imagine a subject in a traditional evil demon world, whose environment is such that the subject's ways of forming beliefs are grossly unreliable. However, the subject in question is forming beliefs in, internally speaking, the same way that we do in our demon-free world. Intuitively, if my beliefs are justified, then so are those of my counterpart. But according to reliabilism, my beliefs are justified, and my counterpart's are not. Similarly, Cohen (1984) invites us to suppose that the actual world is an evil demon world, and to reject, on the basis of a PAI-judgment, the consequence that none of our beliefs are justified, or that there is no difference in justificatory status between someone who reasons clearly in the demon world, and someone who reasons in a confused manner. Again, we have a PAI intuition that is at odds with a non-PAI theory. And again, they moderate one another. The force of the PAI intuition is informed by the strength of the extra-PAI theory it contradicts. Although Cohen and Lehrer argue that the PAI judgment is decisive against the theory, philosophical practice has, in effect, weighed the competing considerations against one another. That is to say, the fact that the New Evil Demon intuition is at odds with an external-to-PAI theory has limited the acceptance of the intuition. Engel (1992) rejects the PAI intuition on non-PAI grounds, citing the independent plausibility and attractiveness of reliabilism (defended in non-PAI ways as articulated above) and the importance of truth as a guiding norm for belief (one of the items mentioned on Weinberg's list of non-PAI intuitions). Since a belief held in a demon world, Engel says, is not at all likely to be true, it is not epistemically valuable, and therefore unjustified. We therefore ought not to reject the independently plausible reliabilist theory.

We needn't adjudicate Cohen and Lehrer versus Goldman and Engel. The point is descriptive and methodological: in evaluating whether to accept the deliverances of a PAI intuition, external considerations can be, and are, taken seriously, even in traditional armchair practices. Therefore, *contra* Weinberg's claim, PAI is susceptible to external corroboration. Since there is every reason to expect the false deliverances of PAI to be more likely than the true ones to conflict with what we have PAI-external reasons to believe, this feature of PAI confers on it some degree of hopefulness.

It should, I trust, be easy to generalize the treatment to additional examples. Any time general theories can be posited to explain quotidian phenomena, those theories can corroborate PAI judgments about related phenomena in hypothetical cases, thus conferring hopefulness to PAI via external corroboration.

## **§6. Internal Coherence and PAI**

Internal coherence confers hopefulness by helping practitioners to recognize conflicting data points. Weinberg discusses two forms of practice-internal coherence: intersubjective coherence and intrasubjective coherence. Neither, Weinberg thinks, confers hopefulness on philosophy.

### **6.1. PAI's Boundaries**

Before evaluating Weinberg's particular arguments, it's worth emphasizing that whether a practice enjoys a significant degree of corroboration by internal coherence or not depends, like whether it enjoys external corroboration or not, in significant part on the precise delineation of the practice. Roughly speaking, the broader the practice, the greater its resources for internal corroboration. The narrower it is, the greater is its ability to corroborate externally. I argued in the previous section that Weinberg's target practice, being rather narrow, and excluding much prototypical philosophy, is susceptible to a great degree of external corroboration. If a practice is defined more broadly, to include things like judgments that ordinary cases fall under particular concepts, and theory-building inferences to the best explanation, then the corroboration I focused on in the previous section will be counted as internal instead of external. Since Weinberg has not claimed that either internal or external corroboration is by itself necessary for hopefulness, this classification doesn't much matter; wherever one draws the line, philosophical practices will attain the hope articulated in the previous section; so likewise for the hope I will articulate in this section.

Since the boundaries of PAI are both narrow and imprecisely defined, then, it is difficult to point with confidence to examples of PAI-internal coherence that confer hopefulness. In what follows, I'll consider examples of traditional philosophical methodology that at least might plausibly be thought to be PAI-internal, with the caveat that I am prepared, if Weinberg insists, to classify the relevant judgments as extra-PAI. In this event, the corroboration-based hope that I am pointing to can be classified as external.

### **6.2. Intersubjective Corroboration**

Let's begin with intersubjective internal corroboration, before turning to ways in which philosophers achieve hope via intrasubjective coherence. Weinberg argues that “[a]t present we have little reason to expect to mitigate the fallibility of intuition by appeal to intersubjective agreement,” citing cross-cultural data and anecdotes about anonymous philosophers who demur from standard intuitions about some philosophical cases. (337) Two responses come to mind.

#### **Sameness of Practice**

First, for the cross-cultural data Weinberg cites to bear on the question of intersubjective PAI-internal corroboration, it must be the case that the subjects tested are employing PAI in their generation of dissenting judgments. As a number of authors have remarked, however, it is far from clear that the subjects in the studies in question are engaging in the same practice as the target philosophical one. Weinberg, Nichols and Stich (2001), for instance, take themselves to target at least “just about all of the vast literature that arose in response to Gettier's classic paper”. But the survey responses that are taken to establish the internal incoherence of PAI involve students who consider the cases in a vacuum, with relatively little reflection; by contrast, PAI—if it is to be a description of how the post-Gettier literature is actually performed—involves careful consideration of the relevant cases by philosophers who have considered the consequences of various verdicts about similar cases, and who are

informed by greater theoretical background knowledge. (See Kauppinen's (2007) 'surface' versus 'robust' intuitions.) For example, it might not be at all clear how to categorize a certain case, until one recognizes its structural similarity with another, clearer case. This is a consideration philosophers will be sensitive to—presumably, including philosophers who are engaging in PAI—but students engaging with a short questionnaire, who are not attempting to theorize, will not. So I see no particular reason to think that the kinds of intuitions probed in the studies cited are generated in anything very much like PAI.

The other case that Weinberg cites is Machery, Mallon, Nichols, and Stich's (2004) critique of Kripke's arguments against descriptivism. (Let us assume, then, that Kripke's refutation of descriptivism is part of PAI; otherwise, the data is irrelevant for Weinberg's purposes.) As Michael Devitt (forthcoming) has emphasized, the target intuitions comprise only a very minor part of Kripke's argument. A philosopher engaging with Kripke's text, considering whether names are synonymous with, or have their references fixed by, descriptions, will follow the arguments that *lead up* to the thought experiments that Machery, Mallon, Nichols, and Stich tested on cross-cultural samples of students. We often use names to refer even in ignorance of definite descriptions, so names are not synonymous with definite descriptions. The properties most closely associated with names are nevertheless contingent properties of the individuals named, so names are not necessarily co-extensive with definite descriptions. *Once these facts are appreciated*, it is obvious that widespread error about the properties closely associated with names is possible; not only *need* Gödel not have discovered the incompleteness of arithmetic, it might, for all we know *a priori*, turn out that *in fact* he did not do so; this is consistent with the actual use of the name 'Gödel'. Kripke drove this last fact home in a particularly vivid way, with his diverting yarn about the hapless Schmidt and the treacherous Gödel. The story is more illustration than argument. The use to which philosophers who follow Kripke in rejecting descriptivism put the story is one that centrally involves the arguments and claims and examples leading up to it; that consideration of the story in a vacuum generates deviant intuitions in some groups of people does not show that PAI doesn't enjoy extremely strong intersubjective agreement. Consideration of the story in a vacuum is no part of PAI.

### **Sensitivity to Agreement**

Second, that there are in some cases surprising examples of disagreement does not show that intersubjective agreement among peers does not play an important corroborative role in the methodology of philosophy, in a way that confers hopefulness on the practice on the whole. Recall that Weinberg's project is to argue that PAI is hopeless, and that in particular, it does not derive hope from corroboration of intuitions with colleagues. Even if, *contra* the previous argument, surveys demonstrated that there is some surprising disagreement among practitioners of PAI, this would not show that intersubjective PAI agreement does not confer hopefulness on the practice. At best, it would show that there is some potential source of hopefulness that has been less than fully exploited. It is consistent with there being some disagreement about which we're ignorant or insensitive that there is much agreement and disagreement about which we are aware and sensitive, and that our practice makes use of this data in a way that confers hopefulness. And indeed, it is overwhelmingly plausible that this is so. Not all intuitions are generally agreed upon, even within mainstream analytic philosophy circles. When a philosopher relies on an idiosyncratic intuition, his claim is less likely to take hold, and to be widely accepted in the practice at large. The greater the extent to which there is general agreement that an intuition is correct, the greater its chances of general acceptance. It is, for example, more likely to figure in a peer-reviewed publication, or to be taken as a starting point in future investigations. So at the level of the discipline on the whole,

intersubjective PAI-internal corroboration can play a significant role in selecting judgments about which there is consensus; controversial intuitions are more likely to be singled out and rejected.

At the individual level, too, it is plausible that a philosopher's use of her own PAI intuitions is calibrated by intersubjective corroboration with those of her peers. Some of the mechanisms just mentioned as conferring hopefulness on the discipline on the whole can also confer hopefulness on an individual's use of intuitions: when a philosopher makes an argument based on an intuition, and her colleagues reject the argument because they do not share the intuition, they give her reason to attend to that intuition with greater scrutiny; the intuition's controversial reception leads her to treat it with less force; so likewise when her articles are rejected by journal referees for relying on idiosyncratic intuitions. Since this is more likely to occur when her intuition is erroneous, these elements of philosophical practice confer hopefulness. Such corroboration can also happen more informally, as when a philosopher tests her colleagues intuitions about new cases she is considering using in her writing: Do you think this sounds like a felicitous utterance? Do you think this is a counterexample to that safety as a constraint on knowledge? Do you have a clear intuition about the moral permissibility of this action? If a philosopher makes use of her professional colleagues in this way, treating her own intuitions differentially according to the reactions of her peers, then her practice likewise gains a degree of hopefulness. This, even when her practice, and those of her peers, all count as PAI.

It is, of course, an empirical question to what extent philosophers *do* rely on the intuitions of their colleagues in the ways outlined. If it were established that, on the whole, philosophers almost never engage in PAI in a way exhibiting sensitivity to their colleagues intuitions in the ways suggested above, then this particular defense of the hopefulness of PAI would then fail. But neither Weinberg nor anyone else has, to my knowledge, produced any evidence to this effect. And indeed, given my own experiences with philosophers in the discipline, the sorts of sensitivities discussed above seem extremely common. (And, when I asked my philosophical colleagues whether they shared this impression about the frequency of this activity, most of them corroborated my opinion.<sup>5</sup>) Many hallway conversations begin: "Can I ask your intuition about a case?" That is to say, my colleagues' practice includes a mechanism for detecting and correcting intuitive errors.

Philosophy, even when performed in armchairs, is performed in armchairs in philosophy departments and research centers. It is therefore much more hopeful than it would be in armchairs on desert islands.

### 6.3. Intrasubjective Agreement

Against the possibility of PAI's hopefulness from intrasubjective agreement, Weinberg suggests (337-38) that philosophers are likely to overestimate the internal coherence of their own judgments: past judgments can influence future concept application; categorization responds to different features under different circumstances; we're susceptible to order effects and bias towards our own theories. There is, Weinberg argues, probably less intrapersonal convergence than philosophers typically assume. He rejects as speculative and

---

<sup>5</sup> This corroboration was, however, not entirely universal. For example, Jonathan Weinberg has told me that he thinks that such observations are idiosyncratic, and that philosophical practices on the whole do not take advantage of this potential source of hope.

unmotivated the suggestion that “greater reflection will yield greater convergence” of intuition.

Here, Weinberg seems to have lost the thread of his own argument. The suggestion that PAI judgments do not converge internally does not bear against the suggestion that internal coherence can be a source of hope for PAI. Quite the contrary. The reason that tests for internal coherence confer hopefulness to a practice is that when an evidential source issues into an error, it will be *incoherent*; this is valuable because incoherences are recognizable. As Weinberg says in his introduction of the value of considerations for internal coherence: “This allows us generally to check our results against further results in the expectation that the odd-man-out deliverances are mistaken.” (330) Insofar as we’re looking to detect and correct mistaken PAI judgments, we *want* there to be intrasubjective divergence. The fact that our intuitions are jointly incoherent is of critical importance in traditional methodology: it provides us with an important means for recognizing cases where we might be going wrong. If our fallibility manifested itself only in coherent packages of intuitions, it would be much more difficult to detect errors, and our practices would be much less hopeless. This feature of philosophy has been widely recognized. Bertrand Russell (1986) famously quipped, in his lectures on logical atomism:

I am trying as far as possible ... to start with perfectly plain truisms. My desire and wish is that the things I start with should be so obvious that you wonder why I spend my time stating them. That is what I aim at, because the point of philosophy is to start with something so simple as not to seem worth stating, and to end with something so paradoxical that no one will believe it. (172)

This is not, as one might suppose without context, a cynical ode to the power of sophistry; it is a recognition of, and appreciation for, the fact that the set of the apparently obvious is inconsistent. Some things, apparently so obvious as to be unworth stating, with negations so paradoxical as to deny belief, are false. A prototypical philosophical activity is to identify these surprising truths. One way to discover them is by exploiting the internal incoherences of the deliverances of intuition. To the extent that PAI intuitions are among those examined in this way, PAI enjoys a degree of hopefulness from intrasubjective internal corroboration.

Notice also that sometimes, consideration of distant, unusual, or esoteric cases is the only way to discover such inconsistencies; if we restricted our attention to the mundane, we would overlook the possibilities about which our implicit and explicit commitments yield contradictory verdicts. This would make it much more difficult, in such cases, to detect and correct the errors delivered by our quotidian judgments. Our practice is more hopeful for its incorporation of judgments about far-fetched cases.

George Bealer (1992) aptly describes some examples of rejection of intuitions in this way; these are all cases in which recognized intrasubjective incoherences play a central role in correcting false intuitions.<sup>6</sup>

(1) In the Galileo paradox of infinity, I have an intuition that there are fewer odd numbers than there are natural numbers (odd numbers plus even numbers). But I also have an intuition that the odd numbers are in one-to-one correspondence with the natural numbers and a collection

---

<sup>6</sup> Bealer himself describes things a bit differently; he thinks the project is to explain how all the intuitions are in fact true, by refining our understanding of the content of the relevant intuition. Whether it is the intuition itself that is in error, or our conception of it, is immaterial for present purposes. Whichever way one thinks about intuitions, Bealer is describing a way in which we can correct the erroneous products of our appeals to intuition.

that is in one-to-one correspondence with another does not have fewer things in it. These two intuitions can be reconciled by invoking a distinction between fewer-than in the proper-subset sense and fewer-than in the no-one-one-correspondence sense. My first intuition was that the odds are fewer than the naturals in the former sense, and my second intuition was that they are not fewer in the latter sense. ...

(2) The scientific-essentialist literature (Kripke *et al.*) provides a second illustration of how redescription can be used to reconcile intuitions that initially appear to be in conflict. Initially, there appears to be a conflict between old-fashion anti-scientific-essentialist intuitions (e.g., the intuition that there might be some water with no hydrogen in it) and the new pro-scientific-essentialist intuitions. Such conflicts would result in a mere stalemate between the old view and the new view. However, by redescribing these intuitions in terms of the distinction between epistemic possibility and metaphysical possibility, scientific essentialists are able to resolve the apparent conflict in favour of their view. Evidently, anti-scientific-essentialists are unable to do the analogous thing for their view. So the stalemate is evidently broken in favour of scientific essentialism.

(3) The Gricean distinction between genuine semantical implication and mere conversational implicature also yields redescriptions that serve to reconcile a great many intuitions that initially appear to be in conflict. (112-13)

They are all examples of hopefulness in action in traditional philosophy.<sup>7</sup> One particularly impressive example of the third kind Bealer mentions is Robert Stalnaker's rejection of the truth-functional theory of indicative conditionals, in the face of a powerful intuitive argument in favor of it. I include a discussion of this case study in some depth in §8.2 below.

## §7. Margins and PAI

About the hopelessness of PAI with respect to the detectability of margins, Weinberg writes that the output of the relevant intuitions is "basically a 1-bit signal: is p possible, yes or no? Or: Does the hypothetical situation fall under the concept or not?" Any difference in degree or strength, Weinberg says, is "largely unexplored—and unexploited—by current philosophical practice." (335) It is not at all clear why Weinberg takes this to be the case. There's not an argument to this effect; Weinberg appears simply to be reporting what seems to him obvious. But armchair descriptive philosophical methodology is a dangerous game.

### 7.1. Distant Cases

The closest Weinberg gives to an argument is this:

One candidate way of trying to detect intuitions' margins would be to think about what sorts of cases we might be more likely or less likely to be able to reason well about. For example, we could, as a profession, decide to be particularly cautious about using intuitions under circumstances far removed from ordinary conditions—such as cases involving wildly unusual or even nomologically impossible situations, or that can be described only using fairly highfalutin lingo. ...

But clearly we have not as of yet made such a decision, as indicated by the promulgation of cases (just to name an egregious two of a rather large set) like the phenomenal zombies in Chalmers (1996), or Byrne and Hilbert's (1997) poor schnook in their "Color and Reflectances," one of whose eyes has our spectrum but the other has an inverted spectrum, but he only uses one eye on any given day, and his environment just happens to switch colors back and forth in alignment with which eye he's using on any given day. (336)

---

<sup>7</sup> Are the cases in question PAI? As before, it is difficult to be certain, but there is some reason to think so. The relevant thought experiments engaging with scientific essentialism are of the twin-earth variety.

Two immediate remarks are in order. First, it is not clear that “circumstances far removed from ordinary conditions” are always circumstances about which we’re less likely to reason well. Some cases extremely far from ordinary experience are rather easy to think about. Consider an otherwise empty world with two distinct qualitatively intrinsic objects in it; consideration of this possibility leads one by rather ordinary reasoning to the conclusion that either there are haecceity properties or the identity of indiscernibles is false. And actual cases may be highly complicated and difficult to think about. Sometimes it is important to abstract away from real-world complications, and sometimes the best way to do that is to stipulate a fantastically simple set of circumstances.

Second, and more centrally, it is far from clear that philosophers are not “particularly cautious” about intuitions about very strange cases. Reflections on methodology very frequently invoke some kinds of intuitions as stronger than others—both in terms of phenomenological strength and evidential strength. Intuitions are sometimes weakly felt, and discounted on such grounds. And even when they are strongly felt, they are sometimes discounted because their subject matter is something they’re expected to be less reliable about.<sup>8</sup> And the particular examples Weinberg cites—phenomenal zombies and unusual spectrum inversion cases—have hardly been decisive thought experiments in their relative fields. Part of the reason why they haven’t been universally accepted, and why we’re not all dualists about consciousness and physicalists about color, is that there *is* uneasiness in philosophy, as currently practiced, about relying too much on such esoteric intuitions. Philosophers are, when engaging in PAI, sensitive to the kinds of intuitions produced.

## 7.2. Example: Bealer

To take another rather simple example, George Bealer (1992) writes that “[T]he standard justificatory procedure directs us to give greatest evidential weight to intuitions about specific concrete cases. By comparison, ‘theoretical’ intuitions have relatively less evidential weight.” (104) Whether this general principle is right or not—and I have my doubts—it is not at all difficult to find examples of philosophers treating some intuitions with more weight than others, because of a felt difference in strength, or epistemic significance.

## 7.3. Example: Goldman

We may take another example from a case I’ve discussed already: Goldman’s argument for reliabilism. About the intuition that someone in a friendly-demon world shouldn’t be able to have justified beliefs formed on the basis of wishful thinking, Goldman demurs:

There are several possible ways to respond to this case, and I am unsure which response is best, partly because my own intuitions (and those of other people I have consulted) are not entirely clear.<sup>9</sup> (342)

Perhaps Weinberg thinks that treatments such as this one, although they do obviously attempt to recognize such differences in strength of signal, are in fact insensitive to the appropriate

---

<sup>8</sup> As Timothy Williamson writes: “We may be correspondingly more reliable in evaluating the possibility of everyday scenarios than of “far-out” ones, and extra caution may be called for in the latter case. ... But current philosophical practice already shows some sensitivity to such considerations. Many philosophers are more confident in their judgments about more or less realistic thought experiments in epistemology and moral philosophy than about more radically strange ones in metaphysics.” (2007, p. 164)

<sup>9</sup> Goldman here also indicates his sensitivity to the intuitions of others, thus conferring a degree of internal coherence to his invocation of intuition.

degrees. If so, he's provided no argument to that effect; we have only the bald claim that divergences in strength of intuition, if they exist, are ignored by standard practice. The more examples of invocation of strength of intuition we find, the less motivated Weinberg's assertion appears.

#### 7.4. Example: Devitt

To that end, consider another example. Michael Devitt (forthcoming) criticizes Machery, Mallon, Nichols, and Stich's treatment of Kripke on grounds related to the ones I discussed in §6.2 above. He thinks that the intuition they focus on—about a relatively esoteric case—is much less important in Kripke's argument, which relies primarily on ordinary cases. These, Devitt says, carry greater evidential weight:

Kripke showed, with a range of actual cases, that most users of a lot of names are not able to supply descriptions that are adequate to identify what are intuitively the referents. ... Yet Kripke's intuitions—indeed, the intuitions of just about all philosophers—are that these users nonetheless succeed in designating these people with their names.

It is important to note that these cases are humdrum: they are not counterfactual, hypothetical, fictional, or in the least bit fanciful. Inspired by them it is very easy to come up with countless cases, each just as humdrum and each yielding the intuition that a speaker's use of a name designates an object despite the speaker's ignorance or error about the object. These intuitions are very powerful because to reject them is to rule that many names out of just about every mouth fail to designate what they should.

...The further one gets away from actual error toward hypothetical cases that may seem a bit fanciful, the less confident we should be about our own intuitions and, foreshadowing a bit, *especially, those of the folk*. Thinking about fanciful hypothetical cases is hard, particularly for the folk. (Emphasis in original)

Devitt attributes the power of the humdrum cases to their ordinariness; intuitions about usual cases are treated more strongly than are esoteric ones. If he is right—and it appears very plausible to me that he is—then philosophers are sensitive to differential strengths of intuitions.

#### 7.5. Example: Lewis

In a famous passage in 'Causation', David Lewis writes:

When common sense delivers a firm and uncontroversial answer about a not-too-far-fetched case, theory had better agree. If an analysis of causation does not deliver the common-sense answer, that is bad trouble. But when common sense falls into indecision or controversy, or when it is reasonable to suspect that far-fetched cases are being judged by false analogy to commonplace ones, then theory may safely say what it likes. Such cases can be left as spoils to the victor, in D. M. Armstrong's phrase. We can reasonably accept as true whatever answer comes from the analysis that does best on the clearer cases. ...

In my paper, I distinguished one kind of case—preemption with chains of dependence—in which common sense delivers clear positive and negative answers, and my counterfactual analysis succeeds in agreeing. I left all other cases of redundant causation as spoils to the victor, doubting that common-sense opinions about them would be firm and uncontroversial enough to afford useful tests of the analysis. (194)

Indeed, Lewis's 'spoils to the victor' language is now in common philosophical parlance; to describe a case as 'spoils to the victor' is to admit that one's intuitions about the case are weak, and that it ought to be classified according to the best theory, adjudicated on other

grounds.<sup>10</sup> That Lewis's term<sup>11</sup> for this move has become so widespread—it is often used without citation—is strong evidence that the move is part of a well-entrenched practice.

## §8. Theoretical Illumination and PAI

The final source of hopefulness Weinberg considers is theoretical illumination. The more we understand about the mechanisms of an evidential source, the better situated we are with respect to identifying and correcting its errors. Since philosophers have not been very interested, Weinberg says, in how intuitions work or in where they come from, PAI does not acquire hopefulness from theoretical illumination.

### 8.1. Explaining Away

Weinberg here ignores one of the favorite items in the philosopher's toolkit: the explaining away. Very often, the evidential force of an intuition is undercut by an attempt to *explain it away*. An explaining-away is an explanation for the intuition that does not support its truth.<sup>12</sup> For example, some linguists and philosophers used to think that 'and' in English was not truth-functional, because of the intuition that the truth in my mouth of "I gave my talk and went to dinner" entails that I *first* gave my talk and *then* went to dinner. We now know that this intuition is false; we explain it away with a pragmatic story: asserting the conjunction *implicates* a temporal order, but does not entail it. When we intuit that the one must be true if the other is, it is because we are not being sensitive to the difference between entailment and conversational implicature. (A helpful test suggests itself, once the distinction is in hand: implicatures are cancelable in a way entailments are not; the acceptability of "I gave my talk and went to dinner, but I actually gave my talk after dinner" corroborates the pragmatic treatment.)

I'll conclude with a more detailed case study of an influential explaining-away: Robert Stalnaker's (1968) treatment of an argument for truth-functional indicative conditionals. It demonstrates, I believe, the way in which traditional philosophical practices, including plausibly, PAI, derive hopefulness from consideration of the sources of intuitions. It is also an excellent example of internal corroboration at work in traditional philosophical methods.

---

<sup>10</sup> Some examples of 'spoils to the victor' language in philosophy: Frances Howard-Snyder (2007) suggests that intuitions regarding whether safety-net-removing comprises killing or letting die, being "by their nature troublesome and controversial", may be "cases of 'spoils to the victor'—cases we should classify in whatever way the otherwise best theory does." Gregory Currie writes, "[w]hether the work of the artist is to be described as creation or discovery can be counted as spoils to the victor in this debate." (482) Tim Lewens describes the intuition that a swamp-man-like "instant lion" has an organ whose function is to pump blood as one for which "the rule should be 'spoils to the victor', as David Lewis sometimes argues." (535) Jonathan Schaffer responds to a hypothetical objection to his claims about causation that the case on which he relies should be left as spoils to the victor by affirming that the judgment in question is "intuitively clear", and sufficiently central to causal concepts that such a treatment is inappropriate. (166) Tim Maudlin (2007) rejects intuitions about impossibility or possibility of time travel as ungrounded, calling for a spoils-to-the-victor approach. (190)

<sup>11</sup> Lewis credits the term to D. M. Armstrong, but I have not found Armstrong using it in print. Perhaps Armstrong spoke, but did not write, with that vivid term.

<sup>12</sup> See Ichikawa (forthcoming).

## 8.2. Case Study: Stalnaker on Or-to-If Validity for Indicative Conditionals

Here is an example of hopefulness in action in traditional philosophy. Begin with this powerful intuitive argument for the truth-functional semantics for the indicative conditional:

Suppose *if A, B*. Then it is not the case that  $A \ \& \ \sim B$ . So it is the case that  $\sim A \vee B$ . So it is the case that  $A \supset B$ . So the indicative conditional *if A, B* entails the material conditional  $A \supset B$ .

Suppose  $A \supset B$ . Then  $\sim A \vee B$ . Then *if not  $\sim A$ , B*. So *if A, B*. So the horseshoe  $A \supset B$  entails the conditional *if A, B*.

So *if A, B* is equivalent to  $A \supset B$ .

This is not a result to be accepted lightly; it has surprising implications like this one:

If Obama is in Cairo, then Obama is in Moscow.

But the argument is valid and has highly intuitive premises. Grice tried to embrace the surprising implication with a pragmatic story. Stalnaker did better: he solved the puzzle by expertly disarming the intuitive second premise.

Stalnaker distinguishes two distinct notions that are easily conflated in informal thought: semantic entailment versus reasonable inference. He writes:

Reasonable inference, as I shall define it, is a pragmatic relation: it relates speech acts rather than the propositions which are the contents of speech acts. Thus it contrasts with entailment which is a purely semantic relation. Here are rough informal definitions of the two notions: first, reasonable inference: an inference from a sequence of assertions or suppositions (the premises) to an assertion or hypothetical assertion (the conclusion) is reasonable just in case, in every context in which the premises could appropriately be asserted or supposed, it is impossible for anyone to accept the premises without committing himself to the conclusion; second, entailment: a set of propositions (the premises) entails a proposition (the conclusion) just in case it is impossible for the premises to be true without the conclusion being true as well. (194)

With this distinction in hand, Stalnaker will argue that the key move in the second premise of the target argument—that from from the ‘or’ to the ‘if’—represents a reasonable inference, but not a valid one representing semantic entailment. Our intuition that the inference is a good one is explained in a way consistent with its invalidity.

On Stalnaker’s view, conditionals are context-sensitive: the propositional content of the conditional is fixed by the antecedent, the consequent, the world of evaluation, and a context-sensitive selection function from proposition-world pairs to worlds. The function takes the antecedent and the world of evaluation and delivers a target world. (It plays more or less the role of Lewis’s similarity relation, except that it picks out a single world instead of a sphere: think of the selection function as picking out the antecedent-world most similar to the world of evaluation.) The conditional is true just in case the consequent is true in the world picked out by the selection function.

The selection-function is context-sensitive; the circumstances of use play a role in determining the relevant world for evaluation of the consequent. (Again, this is parallel to Lewis, who has a context-sensitive similarity relation; see his remarks on backtracking and Caesar-in-Korea counterfactuals.) But one pragmatic constraint on selection-functions, Stalnaker says, is this:

...if the conditional is being evaluated at a world in the context set, then the world selected must, if possible, be within the context set as well (where C is the context set, if  $i \in C$ , then

$f(A,i) \in C$ ). In other words, all worlds within the context set are closer to each other than any worlds outside it. (199)

Stalnaker goes on to make a plausible case that this constraint is motivated by general conversational principles. From this constraint it follows (and it is independently plausible that) that indicative conditionals are appropriately asserted only in contexts consistent with their antecedents; *if A, B* is assertable only in contexts that include some worlds in which A. Similarly, and on parallel grounds, Stalnaker says that disjunctions are appropriately asserted only when either disjunct may be true without the other; don't say *A or B* if there's no way it's  $A \ \& \ \sim B$ , or if there's no way it's  $B \ \& \ \sim A$ . If there's no way it's  $A \ \& \ \sim B$ , instead of saying *A or B*, you should just say *B*.

Now we have enough to show that the or-to-if inference is always a reasonable one:

Suppose *A or B* is assertable. Then  $\sim A \ \& \ B$  is in the context set. When *A or B* is accepted, all  $\sim A \ \& \ \sim B$  worlds are eliminated. So all remaining  $\sim A$  worlds in the context set are *B* worlds. Since there are some  $\sim A$  worlds in the context set, the selection function used in evaluating *if  $\sim A$ , B* returns a  $\sim A$  world in the context set. Since all such worlds are *B* worlds, the selection function returns a *B* world. So the conditional is true. So any time *A or B* is true, assertable, and asserted, *if  $\sim A$ , B* becomes knowable. So it is reasonable to infer the conditional from the disjunction.

This, even though there are worlds where the disjunction is true but the conditional is false. These are worlds outside the context set. The suggestion is that, when we're tempted to call the inference valid, we're really reacting to the practically similar, but semantically distinct, fact that the inference is a good one for us to use.

And indeed, once we see the suggestion, we can test our intuitions on cases in which the relevant conversational norms are violated. For example, it's not too hard to think about disjunctions that are true, but which are unassertable by the standards Stalnaker gives us. Take a disjunction  $A \vee B$  where *A* is known to be true and *B* is known to be false. It should be easily recognized as true but unassertable. Does it support the or-to-if apparent validity?

$A \vee B$

So, if not-A, B.

When we think, as Stalnaker's diagnosis inspires us to do, of examples of the schema where  $A \vee B$  is true but unassertable in the way indicated, the intuitive validity is greatly diminished; indeed, it's not difficult to find intuitive counterexamples: let *A* be *I have more than \$100*, and *B* be *I am a millionaire*. This is an especially fine feature of Stalnaker's explaining-away: it tells us that we're confusing one property with another, and, once we think to separate the two, it becomes intuitively clear that our intuitions are sensitive to the factor he's pointed us to.

Notice, too, that the key cases are unusual ones: we test our intuitions about a particular sort of disjunction, which our linguistic conventions guarantees to be a deviant one. Sticking to everyday cases will never distinguish entailment from reasonable inference. That we must consider such unusual cases to test the theory is some pressure in favor of classifying Stalnaker's project here as a PAI one; it also underscores the value of PAI.

Stalnaker's project is, I think, a paradigm of first-rate philosophy. If so, it gives an instance of traditional armchair philosophical practice that successfully detected cases of intuitive error,

diagnosed them, and corrected them. In Weinberg's terms, it showed that armchair philosophical appeals to intuition are, at least to some extent, hopeful. In particular, the observed inconsistency between the validity intuition with respect to the or-to-if inference, on the one hand, and the truth-value intuition with respect to the deviant conditionals with false antecedents, on the other, provided reason from internal coherence considerations to look deeper into these intuitions. And theoretical understanding of the genesis of the validity judgment gave an effective explaining-away.

Furthermore, it is perhaps worth mentioning that my high opinion of Stalnaker's project here is not an idiosyncratic one; Stalnaker's argument has, I think, been widely accepted—far more widely accepted than, for instance, Chalmers' arguments against physicalism. If there are still proponents of the truth-functional view of the indicative conditional, they now recognize that the simple argument at the top of this post is not nearly enough to establish it. Insofar as the practitioners of armchair analytic philosophy have, on the whole, accepted Stalnaker's diagnosis, the practice on the whole demonstrates some degree of hopefulness.

## §9. Conclusion

Jonathan Weinberg's (2007) helpfully advances the dialectic of the experimental philosophy critique in a number of ways. For one, it recognizes one of the most important challenges facing attempts to discredit traditional philosophical methods, and lays out the nature and scope of that challenge in an admirable way. For another, it articulates the epistemically interesting category of hopefulness; I agree with Weinberg both that this category is interesting in its own right, and that philosophical appeals to evidence, if hopeless, ought not to be trusted. The point of disagreement, then, between defenders and experimentalist critics of traditional armchair methodology, then, has been given a badly needed focus: is traditional philosophical methodology hopeless?

The answer to this question, I suggest, requires a careful study of actual philosophical practice that is absent from the experimentalist literature. In his new paper, Weinberg does more than he and his colleagues had done before towards characterizing the target of his critique, and connecting experimental results with that target; but I have argued that it is still not nearly enough, either to characterize an appropriate target in PAI, or to argue that it is hopeless. The cursory survey through philosophical methodology I have offered here has, I hope, demonstrated that at least some PAI encodes practices that confer it with hopefulness. And I see no reason to believe that a more thorough investigation into philosophical methods would find the examples I've cited to be importantly atypical.

Weinberg's project represents the clearest attempt to articulate the pressure that experimental philosophy is meant to put on traditional philosophical methodology. I have argued, however, that his case is not convincing. Weinberg thinks our practices are hopeless; that the fallibility of our intuitions dooms us to wallow in unmitigated error. He tells us that we have no means to tell when we've gone wrong; that our methods don't give us resources for internal or external calibration; that we can't, consistent with current practices, differentially respond by giving greater weight to stronger intuitions, grounded in more reliable tendencies, or exploit our psychological understandings of the sources of our intuitions to weed out ones likely to be erroneous. I contend, however, that an examination of philosophy as actually and traditionally practiced delivers a message of hope: yes we can. (And indeed, we do.)

I conclude that experimental philosophy has demonstrated no serious threat to traditional philosophical methodology.<sup>13</sup>

## References

- Bealer, G. (1992) “The Incoherence of Empiricism”, *Proceedings of the Aristotelian Society Supplemental* 66: 99-138.
- Cohen, S. (1984) “Justification and Truth”, *Philosophical Studies* 46: 279-96.
- Cummins, R. (1998) “Reflection on Reflective Equilibrium”, in M. R. DePaul and W. Ramsey (eds.), *Rethinking Intuition*. New York: Rowman & Littlefield Publishers.
- Currie, G. (1998) “Art Works, Ontology of”, E. Craig (ed.), *The Routledge Encyclopedia of Philosophy*. London and New York: Routledge.
- Devitt, M. (Forthcoming). “Experimental Semantics”. *Philosophy and Phenomenological Research*.
- Engel, M. (1992) “Personal and Doxastic Justification”. *Philosophical Studies* 67: 133-51.
- Goldman, Alvin I. (1979). “What is Justified Belief?” In Pappas, G. (ed.), *Justification and Knowledge*, D. Reidel, 1-23. Reprinted in Sosa, E., Kim, J., Fantl, J., and McGrath, M., *Epistemology: An Anthology*, 2<sup>nd</sup> Edition. Malden, MA: Blackwell Publishing.
- Howard-Snyder, F. (2007). “Doing vs. Allowing Harm”, E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2009 edition), URL = <http://plato.stanford.edu/archives/spr2009/entries/doing-allowing/>.
- Ichikawa, J. (forthcoming) “Explaining Away Intuitions”, forthcoming in *Studia Philosophica Estonica*. URL = <http://philosophy.jollyutter.net/papers/eai.pdf>
- Kauppinen, A. (2007). “The Rise and Fall of Experimental Philosophy”, *Philosophical Explorations* 10(2): 95-118.
- Lehrer, K. & Cohen, S. (1983). “Justification, Truth, and Coherence”, *Synthese* 55: 191-207.
- Lewens, T. (2007) ‘Functions’. In Matthen, M. and Stephens, C. (eds.) *Philosophy of Biology*, Elsevier.
- Lewis, D. ‘Causation’. In *Philosophical Papers*, v. II, New York: Oxford University Press.
- Machery, E., Mallon, R., Nichols, S. and Stich, S. “Semantics, Cross-Cultural Style”, *Cognition* 92: B1-B12.
- Maudlin, T. (2007) *The Metaphysics Within Physics*. New York: Oxford University Press.
- Nichols, S., Stich, S., and Weinberg, J. (2003) ‘Metaskepticism: Meditations in Ethno-Epistemology’. In Luper, S. (ed), *The Sceptics*. Ashgate Publishing: 227-47.
- Pust, J. (2000) *Intuitions as Evidence*. New York: Garland.
- Russell, B. (1986) “The Philosophy of Logical Atomism”, in *The Collected Papers of Bertrand Russell*, v. 8: 157-244.
- Schaffer, J. (2000) “Trumpton Preemption”, *The Journal of Philosophy*, 97(4): 165-181.
- Sosa, E. (Forthcoming) “Experimental Philosophy and Philosophical Intuition,” *Philosophical Studies*.
- Stalnaker, R. (1976) “Indicative Conditionals”, reprinted in Harper, W. I., Stalnaker, R.,

---

<sup>13</sup> I am grateful for many discussions on these and related topics with Derek Ball, Jessica Brown, Herman Cappelen, Yuri Cath, Torfinn Huvenes, Dilip Ninan, Ernest Sosa, Brian Weatherston, and Jonathan Weinberg. Many of the ideas in this paper were first developed out of conversations on the Arché Philosophical Methodology Weblog, <http://www.st-andrews.ac.uk/~armeth/>; I am grateful also to the participants in those discussions there, including many of those mentioned above.

and Pearce, G., *Ifs: Conditionals, Belief, Decision, Chance, and Time*. London: D. Reidel Publishing Company: 193-210.

- Stich (1990) *The Fragmentation of Reason*. Cambridge, MA: The MIT Press.
- Swain, S., Alexander, J., and Weinberg, J. (2008) “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp.” *Philosophy and Phenomenological Research* 76(1): 138-55.
- Weinberg, J. (2007) “How to Challenge Intuitions Empirically without Risking Skepticism”, *Midwest Studies in Philosophy*, 31: 318-43.
- Weinberg J., Nichols, S., and Stich, S. (2001) “Normativity and Epistemic Intuitions”, *Philosophical Topics* 29: 429-60.
- Williamson, T. (2007) *The Philosophy of Philosophy*. Malden, MA: Blackwell Publishing.