

[uncorrected draft – forthcoming in *Midwest Studies in Philosophy: Philosophy & the Empirical*]

## How To Challenge Intuitions Empirically Without Risking Skepticism

Jonathan M. Weinberg

[jmweinbe@indiana.edu](mailto:jmweinbe@indiana.edu)

Abstract: Using empirical evidence to attack intuitions can be epistemically dangerous, because various of the complaints that one might raise against them (e.g., that they are fallible; that we possess no non-circular defense of their reliability) can be raised just as easily against perception itself. But the opponents of intuition wish to challenge intuitions without at the same time challenging the rest of our epistemic apparatus. How might this be done? Let us use the term “hopefulness” to refer to the extent to which we possess a good capacity for the detection and correction of the errors of any fallible source of evidence. I argue that we should not trust putative sources of evidence that are substantially lacking in hopefulness (even if they are basically reliable), and that we are indeed already operating under such a norm in our ordinary and scientific practices. I argue further that the philosophical practice of the appeal to intuitions is, in these terms, badly hopeless. But perception is very hopeful. Thus, recognizing the norm of hopefulness allows the opponents to challenge intuitions without thereby risking skepticism.

## How To Challenge Intuitions Without Risking Skepticism

“The whole strength and value, then, of human judgment, depending on the one property, that it can be set right when it is wrong, reliance can be placed on it only when the means of setting it right are kept constantly at hand.” Mill, *On Liberty*

Intuitions are odd critters: intellectual happenings in which it seems to us that something is the case, without arising from our inferring it from any reasons that it is so, or our sensorily perceiving that it is so, or our having a sense of remembering that it is so. When they occur, they frequently stand out with great psychological salience, but they are not forthcoming about their own origins – envoys to our conscious deliberations from some unnamed nation of our unconscious cognition. But intuitions are also among the chief tools in the analytic philosopher’s argumentative repertoire, in particular intuitions that a particular hypothetical case does or does not fall under some target concept. It can seem that analytic philosophy without intuitions just wouldn’t be *analytic* philosophy. So there is a gulf between our understanding of intuitions and their importance to us, and as a result it is perhaps unsurprising that intuitions have become not just one of philosophy’s tools but part of its subject matter as well. Some philosophers, including Stephen Stich (1990) and Jaakko Hintikka (1999), have argued that intuitions can have no normative epistemic force, are ungrounded in any theory of their correct use, are unreliable, and generally speaking ought to be abandoned with the likes of palmistry and entrail reading. Their arguments have focused on how error-prone such intuitions can be, with a lousy track-record both historically (e.g., Kant on the necessity of the Euclidean geometry of space; the comprehension axiom in naive set theory) and scientifically (e.g., Kahneman, Slovic,

& Tversky (1982); Evans & Over (1996)). Intuitions have also been shown to vary from group to group in a manner inconsistent with philosophers' reliance on them (e.g., Weinberg, Nichols, & Stich (2001); Machery et al. (2004)), and perhaps to be overly sensitive to what other cases have been recently considered (Swain, Alexander, & Weinberg (forthcoming)). In addition to intuition's demonstrable fallibility, some opponents of intuition (e.g., Cummins (1998)) have cited problems with the psychological sources of intuition: our best possible accounts of where intuitions can come from do not square well with our hopes that they are any sort of reliable guide to a truth beyond themselves.

Some defenders of intuition have responded by stressing that intuition, while fallible, is nonetheless not obviously in worse epistemic shape than, say, sense-perception, and like sense-perception is worthy of our Reidian trust (e.g., Bealer (1992); Foley (1998); Sosa (1998, 2005); Tidman (1996); Yablo (1993)). To give up on intuition altogether, others claim, would be to capitulate to a radical skepticism (BonJour (1998); Williamson (2004)). Others (e.g., Jackman (2005); Jackson (1998); Weatherson (2003)) have defended the idea that our intuitions cannot be *totally* mistaken, since they are at least in part constitutive of the concepts which feature in them. Note that these responses are aimed very broadly – their contention is that intuition must be considered at least *basically* reliable and to be *generally* a good source of evidence. The defense is not so much of philosophers' intuitions *per se* but of intuitions at large.

The two sides of this dialectic are therefore talking somewhat past each other. Although the opponents of philosophers' reliance on intuition (henceforth simply 'opponents') have raised specific charges against a *particular* flavor of philosophical intuition, the defenders of intuition-based methods (henceforth simply 'defenders') have offered a defense of intuitions *in general*. My aim in this paper will be to clarify the nature of the opponents' case so that they can take

account of those general defenses, and thereby present an argument whose force must be more directly reckoned with. In order to set up that argument, I will first have to set forth a framework for arguing about the trustworthiness of basic sources of evidence, and I hope that this framework will prove of epistemological value even to those who would oppose my intended result about the untrustworthiness of intuitions.

### I. Can We Challenge Intuitions Empirically Without Risking Skepticism?

Before we can begin to refine the opponents' arguments, we must take a moment to clarify their targets. Most discussions up to this point have focused on intuitions themselves, and in a fairly undifferentiated way.<sup>1</sup> A gloss of "intuition" that comports at all with both specialist and folk usage will take them to be a sort of intellectual seeming, phenomenologically distinct from perception (including proprioception and the like), explicit inference, and apparent memory traces. But this construal includes a rather large and motley class of cognitions. And the opponent would be unwise to keep the conversation focused on so broad a class, since it will include a great deal of cognition that the opponent presumably does *not* want to reject, such as the ordinary application of concepts to particulars (Bealer), or the claim that no object can be red all over and green all over (BonJour), or elementary mathematics (Sosa). The defenders can thus get away with – indeed, can benefit from – a vagueness in the target, as that vagueness lumps together the intuitions the opponents really wants to attack with many others that they really don't, like criminals trying to hide themselves in a crowd of innocent bystanders.

---

<sup>1</sup> See, e.g., Goldman and Pust (1998); Pust (2000); Levin (2004); Goldman (forthcoming); Kornblith (forthcoming); Lynch (forthcoming); Sosa (forthcoming).

One approach would be to define specifically *philosophical* intuitions, and focus the argument on them. But this is unlikely to succeed, as it seems improbable that ‘the philosophical’ forms anything like a natural kind of our cognition. It cannot be in terms of having normative content, or modal content, as such content is still to be found vast amounts of our ordinary cognition (which, again, the opponent is trying to spare). And defining philosophical intuitions simply as “intuitions had by philosophers” seems badly *ad hoc*, and moreover the opponents’ arguments have not generally aimed at any special intuiting incompetence on the part of philosophers. And philosophers have all sorts of ordinary, unobjectionable intuitions all the time, too, which should hardly be guilty by association with whatever intuitional troublemakers the philosophers may deploy in their professional lives.

But the idea that what the opponents really object to is something peculiar to philosophers is an attractive one. Instead of thinking in terms of a problem with something philosophers *have* – in 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. In the extant practice of appeal to intuitions as philosophical evidence, one cites one’s application or withholding of a concept from a given case, usually a hypothetical one, in defense of (or in order to attack) a particular philosophical claim. Such citations thus are meant to carry argumentative, evidential weight, but one is not usually required to offer any further argumentation for the intuition itself. In particular no empirical evidence is required, because one is presumed to have stipulated all the contingencies in the construction of the hypothetical, and one is thus applying only one’s mastery of the concepts involved and not any empirical knowledge. Such citations are thus, in one sense, foundational: although they are used to provide

evidence, one does not, and need not, provide further evidence for them.<sup>2</sup> However, they are not generally taken to be incorrigible or indubitable (BonJour (1998); Bealer (1996, 2000); Weatherson (2003)). We may choose not to endorse an intuition, if the balance of evidence speaks against it; or if one comes to think that the intuition was not formed in a sufficiently truth-conducive way, say, because one did not previously attend to a subtle relation between aspects of the case. However, even if the appeal to an intuition is never taken to be the end of the argument on the whole, it is usually meant to be the terminus of a particular chain of reasoning. This is typical for seemings – an appeal to a percept can similarly terminate a chain of reasoning.

So far, there is little for the opponent to object to; chains of reasoning, like explanations, must come to an end somewhere, and seemings are a standard place for them to do so. Their 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 any given case may be. Everyone is familiar with the likes of Davidson's Swampman and Searle's Chinese room, but one can look at the very recent literature and find the

---

<sup>2</sup> Please note that this very thin sense of “foundational” does not preclude coherentist, infinitist, externalist, and/or localist epistemologies from making sense of them. Any epistemology that attends to our justificatory practices has to say *something* about where we stop offering more reasons and are right to do so, and it may be that one need not read the ultimate structure of justification off of the shape of those practices.

likes of double-lesioned testifiers<sup>3</sup>, new evil demons<sup>4</sup>, and fissioning/fusioning/teleporting pairs (or are they?) of persons.<sup>5</sup> So this anything-goes aspect of the practice is what makes it particularly ripe for the opponents' challenge. It does not of itself comprise the challenge, of course – that's for the rest of this paper to articulate. But 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.<sup>6</sup>

Now that we have fixed the target of the opponents' challenge, we can begin to clarify the structure of their argument, and see just what sort of dialectical resources they do or don't need. The opponents' main sort of premise is a demonstration that intuitions of the sort standardly deployed by philosophers have some sort of less-than-desirable characteristic, such as being error-prone or varying across populations. And their ultimate intended conclusion is some pessimistic evaluation of the trustworthiness of philosophers' appeals to intuition. But as it stands, the inference is an enthymeme. We look to infer from a characteristic of a putative source of evidence to its untrustworthiness, and surely not every epistemically less-than-optimal characteristic warrants such a harsh evaluation. What sort of epistemological principle can the opponents appeal to, in order to derive such a conclusion from such a premise?

---

<sup>3</sup> Lackey (forthcoming).

<sup>4</sup> Lehrer and Cohen (1983).

<sup>5</sup> Most famously Parfit (1984). See also Gendler (2002) for a critique of some intuitions in the personal identity literature.

<sup>6</sup> Note that, given the very broad construal of "intuition" offered earlier, it follows that this practice does not exhaust all the ways in which philosophers may deploy intuitions. They may use them in various ways in their ordinary lives (perhaps when they make quotidian moral judgments about possible actions they are considering undertaking), or also in their professional lives but outside of the PAI (perhaps when they engage in certain forms of basic logical reasoning). Moreover, PAI reflects only a particular subset of our modal judgments, and to impeach PAI need not involve any broader attack on our general capacity for such judgments. The arguments here are thus immune to the complaints of Williamson (2004).

In filling in any enthymeme, one must always navigate between being too weak to get the job done and too strong to be plausible. But the ‘too strong’ side of the dilemma is especially tricky in this case. An overly-strong added premise here risks being not merely implausible, but indeed epistemically disastrous. For the stronger the premise, the greater the proportion of our cognition that will be ruled as untrustworthy. Consider a version of the premise that asserted that any fallible putative source of evidence should not be trusted. Since pretty much all of our sources of evidence, like perception and testimony, are fallible, this version of the argument would force us to give up nigh-well all of our justification about the world. So, even though the opponents’ argument is meant to be skeptical of only a much more limited target, proposing too ambitious a candidate premise would result in a painfully *unlimited* version of skepticism. So, since pretty much all of our sources of evidence about the world are sometimes wrong, the antecedent of the missing premise had better appeal to something more than mere fallibility itself. Similarly, the opponents had better not be demanding that deployers of intuition be able to offer a defense of intuition without any invocation of intuition itself, since it seems that no other source of evidence that we have is similarly capable of such a defense (on an analogy with Alston (1993); see Pust (2000) for a vigorous defense of the claim that intuition and perception are on a par with regard to the availability of noncircular defenses of their trustworthiness). Here is an example of this sort of worry about the opponents’ arguments in action, from Ernest Sosa. He is responding in particular to the results of Swain, Alexander, and Weinberg (forthcoming), which contend that some intuitions about some cases may be inappropriately influenced by what other cases had recently been considered:

But surely the effects of priming, framing, and other such contextual factors will affect the

epistemic status of intuition in general, only in the sort of way that they affect the epistemic status of perceptual observation in general. One would think that the ways of preserving the epistemic importance of perception in the face of such effects on perceptual judgments would be analogously available for the preservation of the epistemic importance of intuition in the face of such effects on intuitive judgments. The upshot is that we have to be *careful* in how we use intuition, not that intuition is useless. It is of course helpful to be shown how intuition can go astray in unfavorable conditions, just as perception can go similarly astray. But the important question is untouched: Can intuition enjoy relative to philosophy an evidential status analogous to that enjoyed by perception relative to empirical science? (Sosa (forthcoming); emphasis original)

If the kind of fallibility identified by Swain, Alexander, and Weinberg is of a sort also found in perception, then perception in science and intuition in philosophy seem to stand and fall together. And one can leave implicit the premise that we – and especially not the empirically-minded opponents! – will not allow science to fall.

Even if the opponent avoids global skepticism, her auxiliary premise may be too strong in another way: it may disable any possible justification for the premise itself. Some (e.g., Bealer (1998)) have argued that intuition will play a role in the defense of any epistemic principle of the sort the opponent requires. If that claim is right, then the argument must not ultimately rule out any appeal to whatever intuitions may be necessary for that defense. Even if that claim is not correct, it is still the case that *something* must be a source of justification for that premise, and that something had better not be impeached by the principle itself. (As the above discussion

about PAI might suggest, though, Bealer's claim 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.)

So we can impose an adequacy condition on the opponents' argument. They owe us a very particular sort of premise:

(i) The premise must be of the form: "Any putative source of evidence with property X ought not be trusted". I will call that property the *epistemically deleterious characteristic*.

Furthermore, of course

(ii) X must apply to philosophers' appeal to intuitions;

or else it will hardly be an appropriate sort of premise for an argument looking to impeach that practice. The next two conditions are required to avoid skepticism on the one hand, and to avoid self-undermining on the other:

(iii) X must not apply to too large and important a set of our evidential sources;

(iv) X must not apply to the source(s) of evidence for the premise itself.

One thing we can quickly conclude from this discussion that the epistemically deleterious characteristic cannot be mere fallibility, or the lack of a non-circular defense, as that would clearly violate (iii) and likely (iv) as well. But what might a good candidate for the epistemically deleterious characteristic look like, then?

## II. Fallibility & Hope

Although simple fallibility is a lousy candidate for the epistemically deleterious characteristic, there's more than one way to be fallible. My contention here will be that what

many uses of philosophical intuition are guilty of, but which our other standard sources of evidence are not, is *unmitigated* fallibility – a fallibility uncompensated by a decent capacity for detecting and correcting the errors that it entails. As we shall see, mitigation is what can make the difference between a successful and unsuccessful epistemic practice.

Given that we don't (except in our most pathologically philosophical moods) let mere fallibility keep us from trusting sense perception, we might still ask: *why* it is that we do not let the fallibility of sense perception worry us overmuch? It is surely not because we do not care about getting the world right. Rather, we have a justified confidence that our practices involving perception will by and large keep us from getting into too much epistemic trouble. (As noted earlier, the opponent had better not think it is a problem that part of that justification comes from perception itself.) We know that perception is fallible, but we also know a great deal about the circumstances and ways in which it is fallible, and about what to do when we find ourselves in such circumstances. As a result we are not held helpless in the face of its fallibility: our practices that rely on perception can take its fallibility into account, and those very practices include measures meant to shield us from its risks.

We can start with the uncontestable fact that we have excellent track-record evidence of perception's successes. Although we are aware of where we have been led astray, such as optical illusions or misheard speech through the noise of a crowd, nonetheless our sense is that these cases are fairly rare. The huge set of our most unnoticeably boring, quotidian cognitions – think of the set of perceptions you use to steer you from your office to the water fountain and back again – contains a vast majority of cases in which we get the world correct.

But is mere on-average reliability, even to a fairly high degree, enough to make a device trustworthy? Let me raise, only to put aside, overwhelmingly reliable faculties or devices: ones

which, though conceivably capable of committing errors, will do so at a frequency that very closely approximates zero, and only in conditions far removed from those in which it will almost always be used. I will stipulate that a source of evidence that is, we might say, *practically infallible* can be automatically trustworthy, even in the absence of evidence that it is so.<sup>7</sup> I will not worry about the practically infallible for the simple reason that so far as I can tell, very few human faculties or human-made devices have this property. Perhaps such capacities as judging of ourselves that we are in severe pain, or that we exist, are overwhelmingly reliable. But certainly neither sense perception nor our intuitive capacities are such.<sup>8</sup>

Sources of evidence whose reliability is of a merely human-scale sort, however, may need something more in order to merit our trust. I am not looking to call the internalists and externalists out to battle here; let me stipulate that we are talking about devices that are generally reliable *and* which we also have good reason to think reliable. Now, it might seem unfair to ask any more from a device than that. After all, in the sciences the norm for reportable results is (roughly speaking) that a scientific study is generally acceptable if its statistics show that there is at worst a 5% chance that what looks like an effect is really just random noise.<sup>9</sup> If it's fine for scientists that 1 out of every 20 reportable results could really be artifacts or just bad luck, why should anyone be held to a higher standard than that? But as they say at Fedex, if you're handling over 6 million packages every business day, a 'mere' 1% loss rate would still be 60,000 packages a day, or over 15 million lost packages a year – not an acceptable figure if you want to

---

<sup>7</sup> Though perhaps not in the presence of (misleading) evidence that it is not. But nothing will turn here on such cases.

<sup>8</sup> Such talk of *faculties* and *devices* is meant to indicate that this condition of practical infallibility is meant only to apply to robustly existing processes – so we can set aside worries about the infallibility of the-token-process-that-delivered-my-belief-at-7:05pm-on-May-21-that-the-sky-is-blue. If you are the sort of philosopher who thinks that the Generality Problem is just wildly unsolvable, then for (iv) X must apply to the intuitions philosophers appeal to. For my purposes you can just consider this practical infallibility condition as satisfied by no real-world process.

<sup>9</sup> Of course, many results have a p value less than 0.05, so this point is *somewhat* overstated. But this should not affect the basic point I am trying to make here. (Thanks to Kelsey Rinella for pointing this out to me.)

use the phrase “absolutely, positively” in your advertising. If we are building a theory using some putative source of evidence, and we expect to need a great many data points, then we will not be reassured simply to be told that our data is 95% accurate. We would be guaranteeing that our theory will be based on what are still a rather large number of errors.

A deeper problem for a ‘mere reliability’ approach to trustworthiness is that it ignores how impotent such a claim of reliability is in the face of *specific* challenges. If you say that you saw the department chair walking across the quad today, but I offer evidence that this could not be so – e.g., he told me he would be vacationing in Italy again this week – then it would be odd for you to respond with, “Well, he may have told you that, but my vision is overall highly reliable”. If reliability were the be-all-end-all of trustworthiness, there would be both nothing more for you to say, and nothing more that we could reasonably ask for you to say. Yet most likely there *is* more for you to say – e.g., “But I saw him clearly, and the light was so good, and he seemed to recognize me and skulk away hurriedly”. Or, if there isn’t anything more that you can say, then your assertion in this instance has not met the reasonable demands I am making upon it. For obvious reasons, I don’t want my discussion here to hang too much on judgments of the reasonableness of such demands, but the key element is that focusing too much on questions of reliability can obscure what other resources we very typically do have available in such discussions.

I would suggest that the reason we tolerate such a degree of fallibility is *not* that our chief norm is a somewhat permissive standard of reliability, but rather that epistemic factors other than reliability are in play here. The line I wish to push is that it is our capacity to detect and correct for errors that makes the difference between the trustworthy and untrustworthy source.

Consider perception again. It is important to note that we have much more to go on in our perceptual practices than reassurance of the on-average high accuracy of our eyes and ears. We also have a well-developed capacity for telling when we are in one of the thankfully rare cases in which our senses deceive us. Each sense modality within itself demonstrates significant intersubjective and intrasubjective agreement: most people see most things in mostly the same way, and their later perceptions are similarly in alignment with their earlier ones. Thus we can use multiple looks in and across perceivers to check each other. If we really want to, we can rely on various sorts of instruments, such as a measuring tape, to double-check what our eyes tell us; though again we would not do so except against a background of very general agreement between our senses and the deliverances of other sources of information. That old saw “measure twice, cut once” recommends to us to exploit this internal and external congruence, and that piece of folk wisdom demonstrates the epistemic relevance not only of vision’s reliability but also of the error-detecting and error-correcting procedures of our broader vision-involving practices.

In addition to vision’s basic reliability and checkability, it has the very useful property of having outputs that carry their own information as to whether the conditions for its use are sub-optimal. We have a very serviceable sense of the range of conditions and applications in which vision is, and is not, reliable. We know to put less confidence in perceptual judgments requiring a particularly exacting degree of precision, or conducted in poor illumination. Our ability to adjust our confidence is partly the product of cultural inheritance and of our lives’ trial and error, but importantly the character of visual experience itself is sensitive to these issues. When vision would provide us with less reliable information, it frequently provides us with less information, period. Dim lights yield dim percepts, and indeed under those circumstances we see the world

(accurately) *as* ill-lit, not (inaccurately) *as* washed-out and gray. So here we have another important mitigating factor for vision's fallibility: our practices with vision incorporate a capacity to detect under which circumstances it will be especially fallible. The fallibility of vision is much less threatening to the extent that one can anticipate where it will not merely be fallible but will, in fact, be prone to failure. When we find ourselves in such circumstances, we can take steps to keep ourselves from being suckered. We can deploy checking procedures like those just discussed or, often more easily, change the circumstances by changing the bulb.

Yet another type of mitigating factor here is our understanding of how vision works -- it is "our" understanding in the sense of being possessed by us collectively, in the distributed and divided way in which we possess the intellectual common property of the sciences. For the science of perception provides a growing ratifying scientific account of the how vision works. We understand better how the visual system is able to use the information it innately contains to extract a robust picture of the world around us (e.g., Marr (1982); Palmer (1999)) This gives us both a deeper understanding of vision's reliability – we can understand not just that it is reliable, but *why* it is – as well as a more subtle and systematic account of ways in which can go awry. We now know, for example, to put less stock in our ability to detect complex changes in our visual environment, as people can often suffer from 'change blindness' and/or 'inattentional blindness'; see, e.g., Simons & Levin (1997). (See also Wells *et al.* (2000) for an overview of how some of our legal testimonial norms have adjusted to our increased scientific understanding of vision and visual memory.)

So, we can trust vision despite its fallibility not just because we are confident that vision tends on balance to get things right, but moreover because we are able to tell when it fails to do

so, and we are able to find out the truth by other means on the (thankfully rare) occasions when vision lets us down. The mitigating factors enable us to detect illusions, recognize perceptual artifacts as such, and in general avoid making any unrecoverable errors. But devices or faculties that are unmitigatedly fallible carry no such assurances. Because “mitigatedly fallible” is so ungainly, let me stipulate the term *hopeful* for such sources of evidence: a source of evidence that is not practically infallible is hopeful to the extent that we have the capacity to detect and correct for its errors. Parallelism (as well as malice aforethought) demands that we 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.

I will therefore propose the following candidate for the opponents' missing premise, with hopelessness serving as the epistemically deleterious characteristic.

(H): *Any putative source of evidence that is hopeless ought not be trusted.*

Given the earlier discussion of perception's high degree of helpfulness, one advantage of using hopelessness as the epistemically deleterious characteristic becomes clear. Namely, it does not obviously have the kind of skeptical consequences that fallibility or no-bootstrapping defendability would have. It thus seems likely to satisfy condition (iii). The principle as formulated also transparently satisfies (i). However, it will take some time to reveal whether it can satisfy conditions (ii) and (iv). In order to do show how (iv) – the non-self-undermining condition – is satisfied, I need to make a case for (H) that does not rely too centrally on anything that

(H) itself would rule out. And in order to show how (ii) obtains – the impeaching-intuition condition – I need to make a case that intuitions of the sort philosophers like to appeal to do not presently possess much hope. In section III, I will build on my discussion thus far to make a case for (H), and if I am successful, we can then see to what extent I have relied upon anything hopeless in doing so. Moreover, in the process of arguing for (H) I will reveal some aspects of hope that will be critically useful in section IV, where I will contend that intuition – or at least our current philosophical practice involving it – is hopeless.

### III. The Case for Hope

#### A. Pragmatic Considerations

Why ought we endorse (H)? That is, why think that hopelessness is particularly worthy of our epistemic scorn in a way that fallible-but-hopeful is not? Let me start to answer that question with the pragmatic point that relying on sources of evidence that are hopeful will make us less likely to make unrecoverable errors in our inquiries. Without good methods of determining when our sources are mistaken, we cannot be as secure in the progress of our theories as we otherwise would be. An early and uncorrectable error can set the entire discipline on the wrong path, and lead to a tremendous waste of efforts or, worse, render barren the entire field of inquiry; such is one of the many epistemological morals we can draw from the sad tale of Lysenkoism. Moreover, when we consider that our sources of evidence will be used by different members of our investigative community, who may be supporting distinct and incompatible theories, we can see a further risk in granting any status to the hopeless: disagreements between rival researchers with hopeless sources of evidence may be impossible to adjudicate. If my lab gets results favorable to my theories, and yours gets different results favorable to yours, then

what we should want is some way to have a principled dialogue about how one of us might have gone wrong, and be able to commit further resources to discovering which of us is right. What we should not want is that our disagreement devolve into fruitless table-pounding and foot-stomping. Even if neither of us will ever change his or her own mind, we at least want the possibility of third parties deciding for themselves which lab has achieved more trustworthy results; and we want to be able to be such third parties ourselves, with regard to others' disputes. If, and only if, our methods possess a significant degree of hope, might that possibility be secured. Otherwise at best we can know that *someone* is wrong, without being able to uncover who is wrong and why.

If we rely on more hopeful sources of evidence, then when we make mistakes we will be able to catch those mistakes more quickly and easily, and when we are arguing about who is correct about a particular result, we will have substantive ways of proceeding in that argument. One could say, following on the model of the Sosa line quoted above, that one ought to use a fallible device only if one can be appropriately *careful* in how one is using it – that way, when worries are raised about particular results, one can go back and check them with greater care and attention. But when a source is hopeless, it follows that *we wouldn't even know what it would be to use it carefully*. The norm of (H) enables us to take risks only where we know how to manage those risks. One might think of it as the epistemic equivalent of not driving without insurance.

If I am right in this practical case for (H), then we would hope that we could, in principle, allow it to form some part of our norms governing investigation. Yet I will claim further that we can also see (H) *already and in practice* fruitfully at work in science. I invoked one relevant type of case in the previous section: the scientific norms that allow for what could seem a disturbingly large number of type I errors. We can see now why the  $p < 0.05$  norm is not

necessarily too permissive. That norm is embedded in a larger set of scientific practices that involve significant components of error-detection and error-correction. Indeed, it is science's capacity to catch and moreover learn from its mistakes that some have taken to determine the dividing line between science and pseudoscience. Such norms as the reproducibility of results, the importance of novel predictions, and the value of free competition of rival research programs, all illustrate our commitment to looking for the errors in our scientific theories, and allowing our own theories to be appropriately scrutinized. (See Mayo (1996) for an illuminating discussion of the value of novel predictions in such terms.)

It's also worth noting a further value of having (H) as a norm, in that it allows us to lessen the damage of type I errors without having to resort to a greater stringency that would automatically raise the likelihood of type II errors. Since we want our norms of inquiry to be not just error-avoiding but also truth-discovering, this is an important result – to slightly mangle a metaphor, it lets those of us fishing from Neurath's boat to cast a wide net without having to keep large clumps of unwanted flotsam cluttering up the deck. We can allow a wider range of putative evidence into the system, because we can have confidence that any false pieces that find their way in will also before too long find their way back out, and hopefulness is the key to such epistemic catch-and-release.

## B. Electron Microscopy

To demonstrate the role of hope in scientific practice, I will briefly discuss two particular instances in the history of science of the adoption or rejection of a putative source of evidence, in which hopefulness or hopelessness were at issue. William Bechtel has recently (2000) illuminated the circumstances under which scientists, particularly biologists, would or would not

be willing to accept the deliverances of a new instrument, and I will borrow here his discussion of the development of electron microscopy. When scientists first look to utilize a new technique, they start with good scientific reasons to at least suspect that the technique may be a trustworthy source of new information. But suspecting that the device may work is a long way from trusting that it does so. Biologists in the middle of the 20<sup>th</sup> century were legitimately concerned that the extreme measures required to fix a specimen for imaging with an electron microscope – either cutting fantastically thin slices of cells, or culturing cells between two barely separated plates of glass – would induce artifacts. The biologists' determination of whether or not they could actually rely on the microscope, claims Bechtel, “are based on whether the techniques produce determinate results ([such as] sharply defined structures in the cell...), results consilient with those generated by other techniques, and results that fit into emerging theories” (Bechtel 152).

Early work with the electron microscope did indeed bring about such results, and thereby led to the general acceptance of the device. Bechtel notes that structures indicated in early studies corresponded well to entities posited in the developing theory of the day, such as the cytoskeleton and the internal membranes of mitochondria. Importantly, there existed evidence for the existence of such structures other than the results from the electron microscope. For example, biologists working with the technique of cell fractionation discovered that respiration took place in mitochondria, and biochemists had established that such respiration would require membranes (Bechtel 154). Moreover, the two very different methods of specimen fixation concurred with each other in these indications, providing a source of consilience (Bechtel 154 – 6). This fact was particularly valuable, since it was unlikely that these two radically different ways of abusing a cell would produce the same artifacts. An error produced by the slicing method might well be caught by an image using the culturing method, and vice versa. Finally,

the observation of already well-established entities (such as the mitochondrion itself, whose existence was already observed in light microscopy) further increased the biologists' growing confidence in their new instrument.

### C. Four Sources of Hope

Before continuing on to a case in which hope failed, let me take a minute to draw a lesson from the discussion to this point, and articulate four basic sources of hope – four ways in which our practices with some would-be source of evidence can allow for the kind of error-detection and error-correction that (H) requires, in a way that we can see in both our scientific and more ordinary epistemic practices (such as those involving perception).

The first of these sources of hope is *external corroboration*. Each sense can at least somewhat be corroborated by the others, and all are corroborated by our more theory-mediated predictions about the world. Similarly, with electron microscopy we saw the consilience of its deliverances with both established biological fact and growing biological theory. Furthermore, we can use such corroboration to adapt our practices to what we know about when and how these devices might go awry, so we can expect a device that is hopeful in this way to become even more so over time.

Second, we have *internal coherence*, in terms of agreement both within and across subjects, or, in the case of electron microscopy, within and across laboratories and methods of specimen fixation. This allows us generally to check our results against further results, in the expectation that the odd-man-out deliverances are mistaken.

Third, both perception and electron microscopy display significant *detectability of margins*: the practices are sensitive to the conditions in which the device is less likely to give

good results.<sup>10</sup> The practice may sometimes detect a device's margins using something external to the device itself; e.g., if you know you have been dosed with a hallucinogen, you will know not to trust your senses. But, more often and more interestingly, this factor is often facilitated by the device itself, when it has a very rich output such that we can learn when certain sorts of outputs are characteristic of suboptimal circumstances. As noted earlier, visual perception in bad conditions very often yields percepts which 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. (See, e.g., Bechtel 154.)

Finally, with electron microscopy, and perhaps to a lesser extent with perception, we possess *theoretical illumination* as to how they work when they do – and, more importantly, why they fail to work when they don't. Bechtel discounts the importance of this factor, in that he claims it is a mistake to require the possession of well-confirmed theory that would exactly predict the device's accuracy. Nonetheless, I believe we can accommodate his concerns: his point is primarily that such a theory ought not be *required*, and we can, and I think had better, agree with that point, since our trust of visual perception significantly antedates (and still exceeds) any even halfway decent scientific theory of it. But such a theory, when and to the extent that we *do* have one, can still serve as a mitigating factor. And typically we will not begin to use a novel device without *something* like the beginnings of such a theory – how else did the scientists come to suspect that their fixation procedures were even remotely likely to result in usable images?

---

<sup>10</sup> Although the notion of margin here has some similarities to that used by Williamson in his (2000), it should be noted that his concept concerns the safety of particular *believings*, whereas mine concerns zones of greater or less reliability of particular *sources of evidence*.

One important further observation to make is that hope does not always come from intrinsic aspects of the source of evidence itself so much as from the particular *practices* of using it. Not that the device's characteristics are irrelevant – wildly and randomly inaccurate devices will be very hard to embed in a hopeful practice, and the richness of a device's signal is of course a property of the device itself – but the ways we use and understand the device do a lot of the heavy lifting here, too. Vision by itself is pretty wonderful, but vision once incorporated into our rich set of checking and correcting practices is far better still. A rich signal cannot help us if we do not avail ourselves of its richness, or do not know how to do so. Similarly, even a device with a thin and only slightly reliable output can be put to good epistemic use, when incorporated into a set of practices that can compensate for its shortcomings (Mayo (1996)).

These four sources of hope seem to exhaust the possibilities: any such possibility for checkability would have to come from within the device's signal (detectable margins); from comparing different deliverances of the device (internal coherence); from comparing deliverances of the device to results from other methods (external corroboration); or from our understanding of the device itself (theoretical illumination). There is no necessary combination of these that any given putative source of evidence must have, and surely different sources have been shown to be hopeful using different combinations in different degrees. But certainly any practice<sup>11</sup> for appealing to a source of evidence that lacked all four would be nigh-well hopeless. And, indeed, I will now turn to another case from the history of science concerning a failed practice, which I will contend failed for precisely the reason of lacking anything like these sources of hope.

#### D. The Hopelessness of Introspectionism

---

<sup>11</sup> Or, at least, any such practice concerning a source of evidence that is not practically infallible; see above.

So – to take an example that perhaps closely parallels philosophers’ intuitive practices – consider the troubled history of introspectionism, in the early days of scientific psychology. Its earliest serious scientific practitioner, Wilhelm Wundt, was clearly concerned with such matters as external corroboration of some sort, and warned that “internal perception could yield acceptable data for science only insofar as experimental conditions permitted a replication of inner experience at will” (Danziger 35). So Wundt only looked to apply introspection to cases like sensation, where the mental entity could be correlated with some repeatable external stimulus. Such followers of his as Oswald Kuelpe and Edward Titchener, however, tried to extend the method into such phenomena as the processes of thought, and quickly ended in a methodological disaster that provided much fodder for their behaviorist critics. They bogged down intractably on such fundamental questions as whether all conscious thought was accompanied by some sort of sensory experience: Titchener’s introspectionists at Cornell took themselves to have demonstrated that such non-sensory conscious thought was possible, whereas the Wuerzburg school took themselves to have shown the opposite (Lyons 21). Such disagreements about what should have been the very core results of the method proved to be intractable. As Edwin Boring recollected later in his career, “There is always to be remembered that famous session of the Society of Experimentalist Psychologists in which Titchener, after hot debate with Holt, exclaimed: ‘You can see that green is neither yellowish nor bluish!’ And Holt replied: ‘On the contrary, it is obvious that a green is that yellow-blue which is just exactly as blue as it is yellow.’ The impasse was an ominous portent of the fate of introspectionism.” (“Mind and mechanism”, *American Journal of Psychology* 59 (1946): 176; I have stolen this wonderful example from Lyons, p.159) Ominous indeed, as the practice became more or less

completely defunct by the 1920s, and turned into something of a bogeyman which later generations of psychologists could use to scare their graduate students.

Why then did introspectionist methodology turn out to be hopeless? Not merely because of the existence of disagreement, for that only indicates fallibility, which we have agreed is not enough to impeach a putative source of evidence. But rather, introspectionism failed because its practitioners could find no way to manage and resolve their disagreements. Contrasting the bad cases like introspection with the good cases like perception and electron microscopy, we see the four basic possible sources of hope in action. For starters, one key possible source of hope was lost once Titchener and Kuelpe abandoned Wundt's norm that would have restricted their practice to cases with some degree of external experimental checkability, and the science of their day had not provided any other theoretical means for an external check. Moreover, the persistence of such disagreements nicely illustrates the lack of internal coherence of the practice. Anecdotes like Boring's also demonstrate that richness of signal, though a prerequisite for one source of hope, cannot by itself constitute such a source -- the variabilities within the signal must be brought under some sort of control with appropriate practices of a sort the introspectionists failed to devise. Finally, introspection lacks the sort of hope that comes from having even a rudimentary folk theory of the sort we all have for perception. We don't know, and more to the point the introspectionists certainly didn't know, under what general circumstances the organs of introspection could or could not be expected to function well.

#### E. Hope in Our Nonscientific Special Epistemic Practices

In addition to a pragmatic argument for (H), and evidence of both scientific and everyday epistemic commitment to it, let me conclude this section by briefly speculating about the role it

plays in our nonscientific epistemic practices as well. We might look to such areas as jurisprudence and journalism, fields with both centrally epistemic interests and sets of well-developed norms. And indeed I suspect that we will find error-detection and error-correction to significantly inform the systems of both professions. In the law, for example, we have the importance of discovery and indeed of the entire appeals process, whose purpose is catch errors made in the lower courts. We see the role of hopefulness in journalistic norms, when reporters are supposed to have corroborating information for most of their claims. Also, under most circumstances reporters can accept the most important claims only when the informant is speaking ‘on the record’ – partly this is to promote reliability, since people will presumably be less likely to lie when their name is to be attached to their statement, but it also serves to increase checkability, and thereby hopefulness, since with a statement on the record one can go to the source and ask them whether they actually said it, or perhaps allow them to clarify their meaning.

I will close this section with a topical example. There was much concern surrounding the 2004 elections about the electronic voting machines manufactured by the Diebold corporation, and that concern continues today, for these machines are now widely used. There had been concerns about the fallibility of the machines, such as when one county in Florida delivered an overwhelming primary victory to Richard Gephardt – well after he had already withdrawn from the race! But the arguments against the machines have not attempted to show that the machines are not *generally* reliable. Indeed, there’s no reason to think that they don’t correctly register the citizen’s vote well over 95% of the time. The central worries, however, are that most of these machines are not currently set up to leave any paper trail of their votes, and that the programs of these machines are kept secret. So the worry isn’t about on-average reliability, but about whether the machines and their results can be checked for errors, and have those errors corrected

once found. Even if they are generally pretty reliable, we nonetheless at this point lack any means to more carefully inspect their results in any particular contested election, or any good way of settling disputes that may arise from any fishy outcomes. In my lingo, these machines are hopeless, and it is their hopelessness that has driven much of the popular concern, and the primary solution called for by those concerned was to render the machines hopeful, primarily by enabling an auditable paper trail of votes cast. So even outside of the norms of the epistemic professions, we can observe a commitment to (H).

I will thus take (H) as established, and allow it to frame the debate over whether we should trust intuitions in epistemology (and elsewhere). I am about to try to make the case that the bulk of philosophical intuitions are indeed pretty much hopeless. Before I do so, let me pause to draw some more general philosophical conclusions from the discussion thus far. First, even if my further arguments do not succeed, I hope that I have at least succeeded up to this point in putting forward a principle that will help frame any future debate on this issue of intuition's epistemic status. Both opponents and defenders of intuition can agree that they are arguing about intuition's hopefulness, and focus their arguments accordingly, attempting to show that our intuition-deploying practices do or do not possess the requisite sources of hope to anything like the degree that scientific instruments or our other basic evidential sources (like sense perception, testimony, or memory) do.

The question of the hopefulness of other basic evidential sources is itself one that of course calls for further investigation. It may explain, for example, how we are justified in trusting so much testimony even when we have obviously no means whatsoever of checking the track-record of our informants. Second, the concept of hopefulness provides a new dimension of epistemic appraisal that has not yet been explored by epistemologists. Most such discussions

have focused on reliability (and its cousins like safety), or the possibility of non-circular demonstrations of reliability. But this particular factor has not, I think, previously garnered much attention. For example, in his recent *Beyond “Justification”: Dimensions of Epistemic Evaluation*, William Alston lists what he clearly aims to be a fairly comprehensive list of the “candidates for the status of epistemic desiderata” (Alston 40). These include such old favorites as the belief’s being based on adequate evidence, or that the subject can defend her claim to the probability of the belief’s truth, and also such relative newcomers to epistemology as the belief’s being the product of a properly-functioning cognitive faculty (Alston 43). But nothing like hopefulness appears on his list of 15 putative desiderata. It is an epistemic concept whose worth is well worth exploring, no matter whether one is a friend or foe to intuitions.

#### IV. How Hopeless is Philosophical Intuition?

If my discussion in the previous sections is correct, we are generally – and properly – committed to a principle like (H). It remains, however, to argue that the practice of philosophical appeal to intuition is not merely fallible but hopelessly so. Mere fallibility is easy to demonstrate – just find a case, any case, where intuition has led us astray. Catherine Elgin does so nicely in two sentences: “Intuition tells me that simultaneity is absolute. Intuition is wrong” (Elgin 54). But to demonstrate hopelessness, we need to attend to the particular nature of the intuitions typically deployed by analytic philosophers, and to the philosophers’ practices with such deployments. Can the opponent of intuition make the case that PAI substantively lacks hope? I will try to do so here, by contending (A) that we lack a robust sense of the margins

of intuition; (B) that our best theories of how intuitions work does not help with the task of incorporating error-detection and error-correction procedures into our philosophical practice; and (C) that we have little information about the overall degree of intra- and inter-subjective agreement and agreement with sources of evidence outside of philosophical intuition, and what information we have does not bode well for the defender of intuition. I will then take up the issue of external corroboration in the final section of the paper.

But before arguing that PAI is currently hopeless, it is worth considering the question of whether my restriction to that target is legitimate – or if I run the risk of running afoul of condition (iii) above, i.e., 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 grandmasters 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.

#### A. Intuition’s Dark Margins

When we examine intuition’s output into our consciousness, we see that it profoundly lacks the richness that typifies sense perception. It’s basically a 1-bit signal: is p possible, yes or

no? Or: does the hypothetical situation fall under the concept, or not? E.g., does the agent know that Nogot has a car, or does he only believe it? So it can be hard for us to detect the margins of intuition, to tell when we are more or less likely to get an accurate intuition. Now, if I may slip into a bit of autophenomenology for a moment, I will grant that it does seem that our intuitions have at least a *little* bit of a gradation in their accompanying feelings of subjective certainty, with some cases striking us forcefully and others more tentatively. But this gradation is largely unexplored – and unexploited – by current philosophical practice. We currently possess no standard reporting procedures for registering any degrees of tentativeness or certainty with intuitions, and even if we did do a better job of reporting whatever limited degree of variation in certainty we may feel, there is still no method in our practice for taking such information into account. Moreover, to whatever extent there is such variation in certainty, it is completely unknown to what extent we have any inter- or even intrasubjective agreement about it. We can find head-scratching intuitions promulgated recently by prominent neorationalists, such as Laurence Bonjour's intuition that a standard inductive premise (i.e., that  $m/n$  of observed A's are B's) is highly likely to have some sort of non-random explanation (Bonjour (1998)); or George Bealer's meta-intuitions that intuition is epistemically basic, and is itself a good source of evidence (Bealer (1996)). I, for one, oscillate between complete agnosticism about these claims, to something like an intuition that they are at least plausible. My hypothesis here is that our senses of certainty are very prone to 'theory contamination': propositions that have any intuitive attraction to us at all, and which we know would support our preferred theory, are more likely to strike us forcefully instead of tentatively. (Note that, for my point here, it's sufficient if you think that *I* am the one who is contaminated, in these cases!) Such well-documented cognitive biases as the confirmation bias<sup>12</sup> and the tendency towards overconfidence, especially in

---

<sup>12</sup> Snyder and Cantor (1979) is a classic study in this area. Also, Jonas *et al.* (2001) is particularly relevant, in that

estimating others' agreement (the so-called 'false consensus effect')<sup>13</sup> further suggest that we will tend drastically to over-estimate the extent to which our intuitions about funny sorts of cases will be shared.

Even if the signal of intuition is too impoverished to make it easy for us to discern its margins, we could at least strive to build into our practices some ways external to the intuitions themselves of locating their margins. 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. Were we to do so, we might at least thereby restrict ourselves to intuitions that receive some of the same sources of hope as everyday intuitions.

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 & Hilbert's poor schnook in their (1997), one of whose eyes has our spectrum, but the other 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. Fans of intuition simply have not yet been interested in exploring what sorts of circumstances will lead to suspect intuitions. So it's not just that intuitions themselves offer us little guide as to their margins; in addition, our practices have not developed a way of ferreting them out, either. Since it's hard to explore the conditions under which intuitions would or would not work well, when we simply don't much know how they work at all, this deficiency is thus

---

the authors found increased confirmation bias when confronting pieces of evidence sequentially – which is surely how philosophers tend to encounter thought-experiments.

<sup>13</sup> E.g., Ross *et al.* (1977); Yates (1990).

surely due at least in part to the generally impoverished state of our understanding of how intuitions work. Speaking of which...

## B. The Generally Impoverished State of Our Understanding of How Intuitions Work

The title of this section says it all: we just do not seem to know much about the underlying psychology of the propositional seemings that we term “intuition”. Part of what makes intuitions *intuitions*, after all, is that they have introspectively opaque origins: it seems to us that p, but we have no sense of from where that seeming arises. This is part of what distinguishes intuitions from, say, explicit inferences (Lynch (2005)). So introspection cannot tell us whether all intuitions come from basically the same source or from a handful of different sources or even from a large hodgepodge of different sources.

And science has not yet done much to fill in what introspection has failed to reveal. There are plenty of potentially relevant scientific results out there, from work on implicit learning (e.g., Reber (1993)) to domain-specific cognition (e.g., Hirschfield & Gelman (1994)) to heuristics and biases in reasoning (e.g., Kahneman, Slovic, and Tversky (1982)) to concepts and categorization (e.g., Barsalou *et al.* (2003); Nichols (2004)). But no psychologist has yet taken an interest in figuring out how to apply such science to the intuitions cited so frequently by philosophers. And in a few instances, psychologists and so-called ‘experimental philosophers’ have taken a direct interest in the sorts of intuitions philosophers have enjoyed mooted (e.g., Greene & Haidt (2002); Nichols (2004); Nahmias *et al.* (2005); Nichols (2006)). This literature shows every sign of continuing to grow, but is still very much in its infancy.

## C. What We Don’t Know About Where We Don’t Agree

At present we have little reason to expect to mitigate the fallibility of intuition by appeal to intersubjective agreement. Recent results (Weinberg, Nichols, and Stich (2001); Machery et al. (2004)) have indicated that ethnic or socioeconomic differences between subjects can predict differences in their intuitions about various cases. For example, those papers reveal that broad disagreements can arise in even such key discipline-defining intuitions as Gettier cases in the theory of knowledge or Kripkean anti-descriptivist cases in the theory of reference – Asian subjects were significantly more likely to attribute knowledge in the Gettier case than Western European subjects, and also more likely to have descriptivist intuitions. Empirically-minded philosophers have really only just started exploring the differences in the intuitions of different demographic groups, but these preliminary results make it clear that we by and large just don't know where it is that different people will or won't agree in their intuitions. Even restricting the subject pool to analytic philosophers, I suspect there is overall less agreement than standard philosophical practice presupposes, because having the 'right' intuitions is the entry ticket to various subareas of philosophy. Since I've started presenting on this sort of material, a number of philosophers (whose names I will omit here, for their professional safety) have confessed to me of various paradigm intuitions that they simply have not shared; one of the projects that my research group hopes to pursue is a systematic survey of the intuitions of professional philosophers. (My current unscientific estimation here is that twin-earth reference cases and Kripkean cases of necessary constitution are particularly suspect.)

The current practices that constitute PAI thus prevent PAI from becoming hopeful: the current structure of the discipline practically ensures that those who share the key intuition may participate in the literature, and those who don't will simply be left out of it – their 'negative results' just won't be published, as it were.

Philosophers may overestimate the amount of intra-subjective agreement that one can expect as well. Several leading theories of the psychology of categorization suggest that we form our categories according to the demands of the moment, task-specifically and on-the-fly (Barsalou (1987); Prinz (2002); Barsalou *et al.* (2003)). Merely performing an arbitrary task involving the concept can change the psychological structure involved in making judgments about membership in the category (Goldstone (2003)). At least in some cases philosophers have reported a certain amount of manipulability or instability in their own intuitions about cases; Block (1971) offers some discussion of this with regard to intuitions about homunculi-headed robots, as does Williams (1970) with thought-experiments about personal identity.

But I suspect that philosophers will generally experience a great deal of intrasubjective diachronic agreement in their intuitions, in particular for cases that are theoretically important for their own projects. But suppose that nonphilosophers, and better yet, philosophers considering cases of less direct theoretical interest to them, were to demonstrate significant variability across time. And, as was mentioned above in §I, some recent experimental results by Swain, Alexander, and Weinberg (forthcoming) indicate that ordinary subjects' intuitions in some standard epistemological thought-experiments are inappropriately sensitive to what other thought-experiments had been recently considered. So we may suspect that the uniformity of philosophers' judgments about personally important cases is an artifact of their theoretical commitments, and not the product of their 'pretheoretic' intuitive capacities. This need not be so; *perhaps* thinking hard about a domain puts one in better touch with one's 'real' intuitions. But this is a mere empirical possibility, and there is some evidence that tells against it. Linguists, for example, often talk about how thinking too hard about the grammaticality of a given sentence can 'burn out' their own initial response to it, and they often undertake to check with other, non-

linguist native speakers to confirm (or reject) their own intuitions. Similarly, philosophers are also familiar with the charge of ‘theory contamination’, in which we accuse someone of now finding *p* to seem true only because they have thought for so long and so hard about their preferred theory that requires *p*’s truth. In addition to such aspects of professional linguistic and philosophical practice, there is also psychological evidence that thinking hard about a problem in which the evidence has been clearly laid out into discrete ‘chunks’ – as is very typical for philosophical arguments from intuitions about thought-experiments – leads to significant primacy effects. It seems that especially thoughtful people tend to lock in their initial judgments, and their very thoughtfulness allows them to rationalize *away* those considerations that might have led to future revisions.<sup>14</sup> (Interestingly, in ‘unchunked’ conditions, being thoughtful *does* have the effect of reducing the *recency* effects that tend to be displayed by the less-thoughtful.)

So, when they appeal to the claim that greater reflection will yield greater convergence, the defenders have at best a weak and untested empirical ‘maybe’. But they need better than that – they need to score a real source of hope, and in the absence of positive confirming evidence that reflection will yield it, the appeal to reflection here does not look like hope so much as it resembles wishful thinking.

## V. External Corroboration & The Status of (H)

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 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

---

<sup>14</sup> Petty *et al.* (2001.)

investigating them other than intuition-mongering. In some of the cases where we clearly have gotten alternative access to the subject matter – e.g., physics’ discoveries about space and time – no effort has been made to use this overlap with philosophy to develop any general means of error-detection within philosophical practice. If I may paint with a very broad historical brush, generally once someone has found another way of working on a domain, the practitioners of intuition-based philosophy have retreated from it. Spacetime in philosophy today is not so much a matter for metaphysics per se so much as a matter for the philosopher of physics – the raw materials being not intuition but the results of the science.

Nonetheless, there are a handful of areas in which we might argue that our intuition-deploying practices may have some hope, because of their regular contact with other sources of evidence about their domains. I will briefly discuss three of them here. First, logic and mathematics are excellent examples of domains with hopeful intuitions. We have cross-checking from different sorts of intuitions: we have intuitions concerning the semantic properties of models; intuitions concerning the syntactic properties of proofs, and indeed our straight-on intuitions about mathematics itself. Perhaps more importantly, we have the kind of checking that comes from the successful integration of mathematics and logic into other ongoing scientific concerns; indeed, into almost all ongoing scientific concerns.

Second, the intuitions that partially constitute folk psychology gain a measure of hope from the fact that they make all sorts of predictions about the world, and that we seem at least somewhat capable of learning from the occasional failure of those predictions. To the continuing dismay of numerous Churchlands, scientific psychology has not generally displaced folk psychology so much as built atop it, and thereby provides a means for testing any claims of folk psychology we may wish to put under close scrutiny. The scientific psychology of folk

psychology is itself a growth industry, and surely could provide some help in debugging some of the trickier philosophy of mind intuitions currently on monger (such as Searle's Chinese Room).

Third, and of clearest application to the arguments here, we have intuitions about epistemic norms. The epistemologists in the audience may scratch their heads at this suggestion – what, other than intuition, could speak to the norms that are to guide our believings, reasonings, deliberations, and so on? My answer will perhaps not be surprising, given my discussion in section III above: we have had a long and fairly well-documented history of trying out different norms to guide our inquiries, and we can learn from our historians which norms have been active when and what results they seem to have yielded. And we can use our best information about the structure of investigative communities both past and present, as well as what we know about the human agents who operate within them, to speculate counterfactually about what results various sorts of norms might or might not generate for us today. I am willing to place some confidence in our intuitions about the norms that govern justification, for example, because I expect that where we need to we can appeal to something outside of those intuitions themselves.

So I hope it is clear, then, how our principle (H) is neither hyperbolically skeptical nor self-refuting. It is not skeptical, because so many of the practices in which our methods, devices, and faculties are embedded, are indeed hopeful. Thus eliminating only hopeless practices, as (H) advises us, would not abandon us to a despairingly impoverished set of epistemic resources. And the opponents' appeal to (H) is not self-refuting, because even if the argument for (H) were to rely somewhat on intuitions, we need not expect that (H) has automatically ruled such intuitions out as untrustworthy. I take it that my reliance here on something like intuition here has indeed been fairly minimal, really only in two places: that we want our epistemic norms to do a good job of producing true beliefs, and that if a norm is well-established in both our folk

and scientific practices, then the dialectical burden is on those who would reject that norm. Neither of these are applications of PAI. Moreover, they have the signs of a hopeful use of intuition, in that disagreement about them can easily result in further conversation. For example, if you fail to share my intuition about the value of truth, then I would suggest for a start that we can look to the history and sociology of science, and see whether truth has played an important role in guiding norm selection (though of course I had better not be committed to the claim that truth has been the *only* guide to such selections).

So, ought we trust intuitions in philosophy? The first part of my answer is: no, when the intuitions are participating in practices that are hopeless, lacking any substantive means of error-detection and error-correction; and yes, when the intuition is embedded in practices that are hopeful. The second part of my answer is to suggest that PAI falls into the first of those categories, and thus ought be considered untrustworthy. But some uses of intuition, including those about logic and math, and about epistemic principles whose merits can be partially tested in the laboratory of the history of science, can reasonably be placed in the second category, and we can trust them for establishing premises to use in our arguments – including (I hope!) my arguments here. In general, though, we can now see a way for the opponent to answer the question from the Sosa quote from §I: “Can intuition enjoy relative to philosophy an evidential status analogous to that enjoyed by perception relative to empirical science?” The opponent may now reply, “No, for intuition, as philosophers tend to appeal to it, lacks the hopefulness that perception has in science (and, indeed, in our ordinary lives). Once we learn *how* to be careful with our philosophical intuitions – that is, when our practices have been rendered hopeful – then we will have a successful analogy between PAI and scientist’s appeals to perception.”

In conclusion, let me stress that much of the hopelessness of philosophical intuitions represents not an intrinsic epistemic shortcoming of our intuitive cognition, but rather flaws in our practices involving those intuitions. We simply do not know enough about how intuitions work, and where we may expect to find agreement. Moreover, our institutional norms on the reporting and weighing of intuitions have neither attempted to create a system which facilitates the detecting and correcting of erroneous intuitions, nor investigated the margins of intuition, to help guide our decisions about which intuitions to trust and which we'd be better off ignoring. But these flaws are all in principle correctable: we can learn more about the psychology of intuitions, and we can undertake to revise our practices to be more in line with (H). So even if my main conclusion here is that philosophical practice of appeal to intuitions is hopeless, let me add as a coda that we need not take them to be *hopelessly* hopeless.

### Bibliography

- Alston, William. 1993. *The Reliability of Sense Perception*. Ithaca: Cornell University Press.
- , 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.
- Barsalou, Laurence. 1987. "The instability of graded structure: implications for the nature of concepts." In U. Neisser, ed., *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge University Press.

- Bealer, George. 1992. "The Incoherence of Empiricism". *Proceedings of the Aristotelian Society Supplemental*. 66: 99-138.
- 1996. "A priori knowledge and the scope of philosophy." *Philosophical Studies*, 81, 121 – 142.
- 1998. "Intuition and the Autonomy of Philosophy". In DePaul & Ramsey
- Bechtel, William. 2000. "From Imaging to Believing: Epistemic Issues in Generating Biological Data". In Creath & Maienschein, eds., *Biology and Epistemology*. Cambridge: Cambridge University Press: 138 – 163.
- Block, Ned. 1981. "Psychologism and Behaviorism". *Philosophical Review*, 90(1): 5-43.
- BonJour, Laurence. 1985. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press.
- 1998. *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Boring, Edwin. 1946. "Mind and mechanism". *American Journal of Psychology*, 59.
- Byrne, Alex & Hilbert, David. 1997. "Color and Reflectances". In *Readings on Color, Vol. 1*. MIT Press.
- Carruthers, Peter. 2003. "Moderately Massive Modularity" in *Minds and Persons: Royal Institute of Philosophy Supplement*.: O'Hear, Anthony (ed). 53: 67-89.
- Chalmers, David. 1996. *The Conscious Mind*. Oxford: Oxford University Press.
- Cummins, Robert. 1998 "Reflections on Reflective Equilibrium" In DePaul & Ramsey, eds. *Rethinking Intuition*.
- Danziger, Kurt. *Constructing the Subject*.
- DePaul, M. & Ramsey, W. 1998. *Rethinking Intuition*. Rowman & Littlefield
- Elgin, Catherine. 1996. *Considered Judgment*. Princeton University Press.

- Evans, Jonathan & Over, David. 1996. *Rationality & Reasoning*. Hove, England: Erlbaum.
- Foley, Richard. 1998. "Rationality and Intellectual Self-Trust". In DePaul & Ramsey.
- Goldman, Alvin. (forthcoming)
- and Pust, J. 1998. "Philosophical Theory and Intuitional Evidence. In Michael R. DePaul and William Ramsey (eds.) *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*. Rowman and Littlefield: MD: pp. 179-200.
- Goldstone, Robert. 2003. "Learning to perceive while perceiving to learn." in R. Kimchi, M. Behrmann, and C. Olson (Eds.) *Perceptual Organization in Vision: Behavioral and Neural Perspectives*. Mahwah, New Jersey. Lawrence Erlbaum Associates. 233-78.
- Greene, Joshua and Haidt, Jonathan. 2002. "How (and where) does moral judgment work?" *Trends in Cognitive Sciences*, 6(12), 517-523.
- Hintikka, Jaakko. 1999. "The Emperor's New Intuitions" *Journal of Philosophy* 96(3):127 - 147
- Hirschfield, Lawrence & Gelman, Susan, eds. 1994. *Mapping the Mind: Domain Specificity in Cognition and Culture*. Cambridge: Cambridge University Press.
- Jackman, Henry. 2005. "Intuitions and Semantic Theory". *Metaphilosophy*. 36(3): 363 – 380.
- Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*.
- Kahneman, D., Slovic, P. & Tversky, A. 1982. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- Klein, Peter. 1999. "Human Knowledge and the Infinite Regress of Reasons". *Philosophical Perspectives*, 13, J. Tomberlin (ed.), 297-325.
- Kosslyn, S. & Osherson, D. 1995. *Visual Cognition*. Volume 2 of *An Invitation to Cognitive Science*. Cambridge: MIT Press

- Lynch, Michael. 2005. "Trusting Intuition". In *Truth and Realism: Current Debates*, ed. P. Greenough and M. Lynch. Oxford: Oxford University Press.
- Lyons, William. 1986. *The Disappearance of Introspection*. Cambridge: MIT Press.
- Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). "Semantics, Cross-Cultural Style". *Cognition*, 92, B1-B12.
- Marr, David. 1982. *A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Henry Holt & Co.
- Mayo, Deborah. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- Palmer, Stephen. 1999. *Vision Science*. Cambridge: MIT Press.
- Petty, Richard; Tormala, Zakary; Hawkins, Chris; and Wegener, Duane. 2001. "Motivation to Think and Order Effects in Persuasion: The Moderating Role of Chunking." *Personality and Social Psychology Bulletin*, 27(3), 332-44.
- Pinker, Steven. 1994. *How the Mind Works*. New York: W. W. Norton.
- Prinz, Jesse. 2002. *Furnishing the Mind*. Cambridge: MIT Press.
- Pust, Joel. 2000. *Intuitions as Evidence*. New York: Garland.
- Reber, Arthur. (1993). *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: OUP.
- Rensink, Ronald. 2000. "When Good Observers Go Bad: Change Blindness, Inattentional Blindness, and Visual Experience." *Psyche* 6(9).
- Samuels, Richard. 1998. "Evolutionary Psychology and The Massive Modularity Hypothesis". *British Journal for the Philosophy of Science*. 49: 575-602.
- Simons, D. J. & Levin, D. T. 1997. "Change blindness." *Trends in Cognitive Science*. 1:241-82.

- Sosa, Ernest. 1998. "Minimal Intuition". In DePaul & Ramsey.
- , 2005. "A Defense of the Use of Intuitions in Philosophy". Forthcoming in Bishop & Murphy, eds., *Stich & His Critics*. Blackwell.
- , Forthcoming in *Philosophical Studies*. "Experimental Philosophy and Philosophical Intuition."
- Stich, Stephen. 1990. *The Fragmentation of Reason*. Cambridge: MIT Press.
- Swain, Stacey; Alexander, Joshua; Weinberg, Jonathan. Forthcoming in *Philosophy and Phenomenological Research*. "Running Hot and Cold on Truetemp: The Instability of Philosophical Intuitions".
- Tidman, Paul. 1996. "The Justification of A Priori Intuitions". *Philosophy and Phenomenological Research* 56(1): 161 – 171.
- Weatherson, Brian. 2003. "What Good Are Counterexamples?" *Philosophical Studies* 115: 1 – 31.
- Weinberg, Jonathan; Nichols, Shaun; and Stich, Stephen. 2001. "Normativity and Epistemic Intuitions" *Philosophical Topics*.
- Wells, Gary; Malpass, Roy; Lindsay, R. C.; Fisher, Ronald; Turtle, John; Fulero, Solomon. 2000. "From the Lab to the Police Station: A Successful Application of Eyewitness Research". *American Psychologist*. 55(6): 581 – 598.
- Williams, Bernard. 1970. "The Self and The Future". *Philosophical Review*, 79(2), 161-80.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press.
- , 2004. "Philosophical 'Intuitions' and Skepticism about Judgement". *Dialectica*, 58(1): 109-153.

Yablo, Stephen. 1993. "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research* 53:1-42.