The "Unreliability" of Epistemic Intuitions

Joshua Alexander and Jonathan M. Weinberg[1]

1. Introduction

According to a rather common way of thinking about philosophical methodology, philosophical intuitions play a significant role in contemporary philosophy.[2] On this view, they are an essential part of our "standard justificatory procedure" (Bealer 1998) or the "method of standard philosophical analysis" (Pust 2000), and are part of what makes philosophical methodology unique (Levin 2004, Goldman 2007). We advance philosophical theories on the basis of their ability to explain our philosophical intuitions, and appeal to them as evidence that these theories are true and as reasons for believing as such. Although examples of this way of thinking about philosophical methodology abound, the example most frequently discussed by Kenneth Boyd and Jennifer Nagel (and *passim* in the literature) comes from Gettier (1963), which aims to show that knowledge is not simply justified true belief. Gettier's paper includes two hypothetical cases involving a person who has deduced a true belief on the basis of a justified false belief and, on that basis, formed a justified true belief that doesn't *seem* to count as knowledge. We are supposed to just see this, and this philosophical intuition is in turn supposed to count as sufficient evidence against the claim that a person knows that *p* just in case that person's true belief that *p* is justified.

Whether or not these authors have characterized this practice in exactly the right way, we take there to be a broad consensus that these characterizations are at least in the right neighborhood, and something roughly like this methodological practice is operative with some substantial currency in analytic epistemology today, and for at least the last few decades.[3] Over the past several years, this way of thinking about philosophy has been repeatedly challenged on the basis of both general epistemological reasons (Stich 1988, Cummins 1998, Kornblith 2002) and more specific methodological concerns (Alexander & Weinberg 2007, Weinberg 2007, Alexander 2012). These methodological challenges have focused not on whether philosophical intuitions *could* be a good source of evidence of some sort, but rather whether such intuitions are to be trusted in the ways that they are currently used in actual philosophical practice, given how little we know about them, and given some of the dangerous bits of information that we have about them. While it has become somewhat standard to frame these methodological concerns in terms of the *reliability* of philosophical intuitions, we want to resist this move, arguing instead that the central

---

[1] Authorship is equal.

[2] Here, and in what follows, we remain neutral on the precise psychological nature of philosophical intuitions and whether treating philosophical intuitions as evidence involves treating psychological states (or propositions about psychological states) as evidence or treating the *contents* of those psychological states as evidence.

[3] Williamson (2007) suggests that it is misleading to talk of "intuitions" here, though he basically endorses the methodological picture otherwise. We take our discussion in this paper to apply to philosophical practices as he understands them, needing only suitable terminological tweaks. More radical critics of the descriptive adequacy of this picture include Cappelen (2012) and Deutsch (2011), but we find these philosophers' accounts themselves inadequate to capture and explicate the argumentative strategies in the relevant portions of philosophy, and though this question is worthy of debate, we will not pursue it here. For additional discussion, see Alexander (2010), Alexander (2012), and Weinberg & Alexander (forthcoming).

methodological concern is about intuitional *sensitivity*.

Like Boyd and Nagel, we will here focus on *epistemic* intuitions, and argue that recent concerns about the evidential status of epistemic intuitions can unproblematically acknowledge their overall reliability, at least as 'reliability' is standardly used in epistemology today. If the evidential status of such intuitions in epistemology is going to be defended from this challenge, then it will just not help much to argue that they are on average true more often than false. As we shall argue, any attempts to defend current epistemological practice would require work of a different sort, targeting not baseline accuracy, but rather inappropriate sensitivity.

2. Two Senses of Reliability

Let's distinguish two meanings of "reliability", because we will want to dispute whether epistemic intuitions are currently reliable in one sense, while stipulating to their reliability in another sense. In ordinary usage, "reliable" is basically a synonym either for "trustworthy" or "highly predictable". It applies not only in epistemic contexts, where we talk about things like reliable eyewitnesses, but beyond such contexts as well. We talk, for example, about someone's being a reliable friend or a reliably good guest at a dinner party. Someone who is reliable at *x* is someone can be expected to hit a high *x*-value some suitably large proportion of the time, where what counts as a suitably large proportion will be domain-relative. Reliably safe drivers had better get into a wreck no more than a vanishingly small percentage of the times they get behind the wheel, whereas someone who gets onto base "only" 40% of the time is a fabulously reliable batter. There are perhaps pejorative applications of the ordinary "reliably" as well, for example, someone who is reliably late for dinner is someone who is tardy to the table more often than is ideal. Nonetheless, we will name this the *trustworthiness* sense of "reliability". Compare this with epistemological use of this term, where "reliable" is defined strictly in terms of the propensity of true deliverances of some source of evidence or belief-forming process, and where something's reliability is measured on a somewhat standard scale with a floor somewhere significantly about 50% and a threshold probably not greater than 95% and certainly not as high as 100% – reliability need not require infallibility. We will call this the *baseline accuracy* sense of "reliability", and have in mind something quite like what Goldman (1979) is talking about when he says, "(as a first approximation) *reliability* consists in the tendency of a process to produce beliefs that are true rather than false" (italics original).

The baseline accuracy sense of reliability plays an obvious role in epistemology, where philosophers have used this notion of reliability to build theories of knowledge and justification, but many other debates in philosophy can also be appropriately framed in terms of baseline accuracy. The "rationality wars", for example, may well be fought over whether or not our ordinary inferential capacities are baseline accurate. And even debates about eliminativism about propositional attitudes can perhaps be understood in terms of whether our descriptive and predictive folk-psychological practices are more-or-less accurate or are wholesale mistaken. It is important to note that baseline accuracy is a fairly coarse-grained concept, but these are debates that can operate at a fairly coarse grain. And even champions of propositional attitudes like Fodor would be happy to grant that we can be fairly often wrong in our attributions of beliefs and desires, arguing that these mistakes, as common as they may be, need to be viewed against a background of a pretty high level of success.

Although they do not always explicitly state it in such terms, the nature of their arguments throughout indicates that Boyd and Nagel are clearly interested in a project of defending the overall baseline accuracy of our folk epistemic intuitions, and we have no objections to this kind of project.[4] We think that it is a perfectly respectable project for philosophers to pursue, and for experimental philosophers to pursue in particular, and are broadly inclined to agree with their assessment that our folk epistemic intuitions are generally, and on the whole, reliable. Let's call this the *general reliability thesis* (GRT). It is not unheard of for debates about the evidential status of philosophical intuitions to be framed in terms of GRT. Goldman (2007) defends the use of philosophical intuitions on explicitly reliabilist grounds, as does Bealer (1998) in terms of a modally strong form of reliability, and even some critics have framed their arguments in terms of a deficit of reliability (for example, Machery 2011). Nonetheless, reliability is not the only epistemological machinery that has been used to frame these kinds of debates. Stich (1988), one of the ur-texts in recent debates about philosophical methodology, raises worries about how we'd justify picking between systematically different, but locally coherent sets of philosophical intuitions.[5] Cummins (1998) raises worries about our inability to independently calibrate philosophical intuitions in situations where we are worried that they may have gone awry. And more recently, Weinberg (2007) stipulates to the baseline accuracy of intuitions, but goes on to allege that a paucity of resources for error-detection and error-correction renders much current philosophical practice with intuitions "hopeless".

Those challenges don't require anything like the denial of GRT, and we are inclined to think that any successful attempt to challenge the evidential status of epistemic intuitions on the basis of the kind of work being done in experimental philosophy has to follow suit. Here's why. GRT is a thesis about an overall propensity for correctness – that is, for the proportion of verdicts that are true out of the total set of verdicts. The kind of work being done in experimental philosophy raises worries about our intuitions about specific hypothetical cases, or families of cases. This kind of work does not show anything about what proportion of the total set of relevant cases these specific hypothetical cases comprise, and it is hard to see how it possibly could. Even if there were dozens of such cases – and there are only a handful at this time – that would not go very far at all towards establishing the negation of GRT unless we had good reason to think that the total reference class of cases has a cardinality similarly in the dozens, which does not strike us as antecedently plausible. Moreover, any argument against the GRT needs to make serious estimates of the ratio of bad apples to the total volume of the barrel, and no one has to our knowledge even attempted to make such estimates.[6] We also wholeheartedly agree with those who, in trying to defend epistemic

intuitions, have pointed out that if our intuitive capacities on the whole were unreliable in the baseline accuracy sense, that would be tantamount to an unpalatably strong form of skepticism. That is not a consideration that can preempt any serious empirical attempt to falsify GRT, but so far as we are concerned, it is a reason to want to find a way to raise concerns about the evidential status of epistemic intuitions that is consistent with a broad endorsement of GRT. Put another way, any successful challenge to the evidential status of epistemic intuitions needs to focus on whether they are trustworthy, while respecting the fact that they are reliable in the baseline accuracy sense of reliability at stake in GRT.

3. The Restrictionist Challenge & The Threat of Inappropriate Intuitional Sensitivity

Intuitional diversity has gotten perhaps the most attention, critical and otherwise. And it is easy to see why, since part of what underwrites the practice of pursuing epistemological questions through the lens of our epistemic intuitions is the (often unarticulated) presupposition that these intuitions are more or less universally shared.[7] Moreover, standard machinery from the epistemology of disagreement might make it seem relatively easy to move from intuitional diversity to worries about the reliability of intuitional evidence; after all, evidence of disagreement is *prima facie* reason to worry about evidence.[8] And so we understand why so many authors engaging with this topic, including Boyd and Nagel here, have focused so much of their attention on the diversity results. There is something odd about what they have to say about intuitional diversity, however. They suggest that intuitional diversity is a problem only in contexts where the relevant intuitions are "deeply at odds with one another", going on to argue that there is currently "no robust evidence that the epistemic intuitions of different demographic groups are deeply at odds with each other." But this is a mistake. All that is needed in order for intuitional diversity to be a problem is that the diversity reflects different "epistemic vectors" of the sort that would become encoded in similarly divergent epistemological theories (see Weinberg et al. 2001). If different groups of people have intuitions that are subtly sensitive to different sorts of factors (and if we are going to avoid epistemic relativism) then someone's intuitions have to turn out to be sensitive to the wrong things.[9].

There is also danger here in focusing too tightly on questions of diversity. So far as the larger debate about philosophical methodology is concerned, intuitional diversity is really just a species of a more general category of problematic intuitional *sensitivity*. We want our sources of evidence

---

[7] Some philosophers have argued that intuitional diversity is *not* a problem. Goldman (2007) argues that intuitional diversity reflects conceptual diversity, and that not all forms of conceptual diversity are necessarily problematic; Sosa (2009) suggests that intuitional disagreement might only be superficial; and Zamzow & Nichols (2009) argue that lessons from the history of science, together with work from the social and cognitive sciences, such as Page (2008), suggest that evidential diversity is actually an epistemic good that helps us to overcome several kinds of well-known cognitive biases. For discussion, see Alexander & Weinberg (2007) and Alexander (2012).

[8] Although it is important to say again that, if intuitional diversity is going to threaten GRT, then there needs to be a reference class against which the proportion of disagreement to agreement can be measured, which is something we just don't have.

[9] Boyd and Nagel's misplaced emphasis on "deeply at odds" differences may also explain why they consider only the Gettier case preliminarily reported in Weinberg et al. (2001), and not the several other group differences that are discussed, for example, on a range of varying Truetemp-style cases, which are subtler in effect size but may be particularly salient here as *prima facie* evidence of such differences in epistemic vectors across those groups.

to be sensitive, of course, but we want them to be sensitive to all and only the right kinds of things, that is, whatever is relevant to the truth or falsity of the relevant set of claims. And, it turns out that at least some epistemic intuitions are sensitive to more than just these kinds of things; they are sensitive to aspects of who we are, what we are being asked to do, and how we are being asked to do it. There is a large range of well-motivated and *prima facie* substantiated hypotheses about such sources of noise in various sorts of philosophical intuitions, far more than just ethnicity, gender, and order effects, including such demographic dimensions as personality (Feltz & Cokely 2009), and such seemingly philosophically irrelevant differences as whether people are asked to imagine themselves thinking about the case "in a few days" versus "in a few years" (Weigel 2011), or even what font the case is presented in (Weinberg et al. 2012). Some of these have been observed specifically in knowledge attribution tasks, too, such as an apparent effect of age on fake barn intuitions (Colaco et al forthcoming), and the influence of the moral valence of a proposition on whether or not an agent counts as knowing it (Beebe and Buckwalter 2010) – an effect apparently capable of leading subjects to attribute knowledge in what otherwise should be Gettier cases. These kinds of intuitional sensitivity are both *unwelcome* and *unexpected*, and the very live empirical hypotheses of their existence create a specific kind of methodological challenge to armchair intuitional practices in philosophy, which we have called the *restrictionist challenge* (Alexander & Weinberg 2007).

The restrictionist challenge starts by raising an empirically plausible concern that such sensitivity may very well be found in intuitions of the sorts commonly found in contemporary philosophical practice, and as used in that practice. Given that concern, restrictionists suggest that we must pursue both local methodological restrictions on our uses of intuitional evidence at the practical level, and a global shift in how we think about intuitional evidence and the business of doing philosophy. To understand what this means, think about the frequent comparison between intuitional evidence and perceptual evidence, a comparison endorsed by Boyd and Nagel. Ernest Sosa (2007) argues that, while we know that perceptual evidence displays patterns of unwelcome sensitivity, this simply causes us to be more careful about what perceptual evidence we use and when we use it. But it pays to be careful only when we know what it means to be careful, and here is where the comparison between intuitional evidence and perceptual evidence breaks down. We have a pretty good understanding of when sense perception goes wrong, something that is reflected in our perceptual practices and reinforced by a communal scientific understanding of the mechanisms responsible for our perceptual judgments. This prevents worries about unwelcome perceptual sensitivity from giving rise to global concerns about the epistemic standing of perceptual evidence. The problem is that we aren't in the same position with respect to intuitional evidence. Our comparatively scarce resources for predicting unwelcome intuitional sensitivity puts us in a different epistemic position with respect to intuitional evidence than we are in with respect to perceptual evidence.[10] In a sense, we haven't learned yet what it would mean to be careful. Learning how to be careful will require developing a better understanding of how epistemic intuitions work. If we are going to learn what intuitional evidence can be used and when intuitional evidence can be used, we need to know more about where epistemic intuitions come from, what mechanisms are responsible for producing them, and what factors influence them; and this is going to require looking to the relevant psychology, cognitive science, and empirically-

---

[10] That this important epistemic disanalogy blocks any slide into skepticism is one of the key points of Weinberg (2007).

informed philosophy of mind. It is important to be clear that the restrictionist challenge nowhere involves the argument that philosophical intuitions are evidentially bankrupt, in their entirety and once and for all time.[11] Rather, what is being urged is that the proper evidential role for philosophical intuitions is one that can only be viewed clearly from outside the armchair, and both bounded by and grounded in a scientific understanding of them.

For present purposes, it is particularly important to see that the restrictionist challenge does not take a stand on the overall reliability of epistemic intuitions on the whole; the central methodological problem is that we know that at least some intuitional evidence is problematically sensitive without being able to predict what intuitional evidence is problematically sensitive, and this is a problem regardless of the overall reliability of our epistemic intuitions. Our limited ability to predict when and where a problem will arise can be dangerous, even if we were to have reason to think that the problem will arise only rarely. This is especially true when the stakes are high; for example, in situations where a small number of cases are granted large amounts of power in theory selection.[12] Here, the Gettier cases are again illustrative. Our (allegedly) shared epistemic intuitions about these cases are supposed to count as sufficient evidence that a person's justified true belief need not count as knowledge. This is only one (rather famous) example where our shared epistemic intuitions have been allowed to trump theory, but there are lots of others, and when the stakes are this high, understanding how and when things go wrong becomes more important than knowing that they *usually* go right. This example also provides us with a different way of understanding why GRT doesn't matter all that much in this kind of debate about philosophical methodology. It is a basic constraint in contemporary epistemology that any reasonable theory of knowledge needs to deny that Gettier cases involve knowledge. If epistemologists were wrong about these cases, then pretty much every mainstream theory of knowledge would turn out to be incorrect. Still, Gettier cases comprise a miniscule fraction of potential cases involving knowledge attribution. So even if epistemologists were wrong *only* about Gettier cases, although the overall reliability of our epistemic intuitions would be enormously high, we would still be in a state of complete epistemological disaster.

Setting even these kinds of worries aside, there is another reason to worry that there is an important mismatch between the kind of reliability thesis that Boyd and Nagel defend in their chapter and that the one is actually at stake in recent debates about philosophical methodology. Remember that trustworthiness is relative to domains, and also to purposes. And the restrictionist challenge aims at a practice whose scope and goals may diverge substantially from those of our folk epistemic practices, even despite a fair amount of similarity and overlap between the practices of analytic epistemology and our epistemic folkways. Here, we will focus on three ways kinds of divergence, and point out that each further undercuts the relevance of the baseline accuracy of our folk epistemic attributions to the methodological debate at hand.

First, epistemologists are very often interested in cases that involve extensive details that we suspect are not ecologically valid. Very often, these cases include highly specific information about an agent's mental states and processes, together with some story about how these states and processes connect with other features of the fictional vignette, where this information may not be

---

[11] Restrictionists have been making this point for some time. See, e.g., Swain et al. (2008) p. 153, Alexander & Weinberg (2007), p. 71, and Weinberg (2007) *passim*.

[12] We will suggest below that this is, in fact, a fairly common situation, given current epistemological practice.

available to the agent and is of a sort almost never available in the real world. Gettier cases are again good examples of this, and it is important to keep this feature of these cases in mind regardless of whether non-hypothetical analogs can be constructed for many hypothetical thought-experiments (Williamson 2007).[13] Our folk capacity to think about epistemological issues has likely been shaped to evaluate situations where we typically have rather sparser, noisier access to what might be going on in someone's head. This mismatch between philosophical thought experiments and the proper domain of our folk epistemic capacities will weaken an attempted inference from accuracy on the latter to our trustworthiness in using the former.[14]

Second, epistemologists want to do something with knowledge attributions that the folk are not typically interested in doing, namely, using them to argue for their preferred epistemological theories. And, while Gettier cases are perhaps the most famous paradigm of high-stakes cases in epistemology, they are not the only members of that set. The literature is full of important cases, and our intuitions about such cases are often meant to place substantial constraints on our epistemological theorizing to comply with their verdicts, or face a nontrivial burden of otherwise accommodating them where we wish to dissent from them. Some of these cases are fairly ordinary, like the intuition that one typically does not know in advance that one has lost a lottery simply in virtue of knowing that the odds of one's having won are miniscule. But very often they are more high-flying or esoteric, such as those involving evil demons (old and new), or clairvoyants, or bizarre brain lesions with weird effects not to be found even in the works of Oliver Sacks. There's a good methodological reason that such funky cases are both common and important in analytic epistemology: we are often trying to get some evidential traction in the slippery effort of rationally preferring one very good epistemological theory over rival, very good epistemological theories. This helps show how Boyd and Nagel get matters somewhat the wrong way around in their discussion of ordinary versus subtle cases; it doesn't really matter much whether epistemologists sometimes, or even very often, deploy ordinary cases, so long as they also rely crucially on more esoteric cases. Yet the very factors that make intuitions about unusual cases methodologically crucial also leads us to expect them generally to be more error-prone. They will be more susceptible to subtle effects of context, for example, because they will often involve splitting apart distinct features that commonly go together in knowledge attribution and using them against one another, such as the baseline accuracy of a piece of cognition and the availability of at least some considerations that speak in favor of that accuracy. Different contexts may cue up different weightings of such features in our unconscious categorizing systems, and thus have an increased potential to produce different attributions.

As for group differences in cognition, that stands as a vibrant and hotly-debated topic in psychology these days, and, we think, a rather more open question than Boyd and Nagel take it to be. But we do not need to take sides in that ongoing debate here. For our purposes, it is enough to note that in order to raise worries about the relevant philosophical practices, one need not go as far as Nisbett et al. (2001) or Henrich et al. (2010), and claim strong and pervasive differences between 'WEIRD' Westerners and the rest of the world. For example, although he is a prominent critic of those authors and their claims of deep group differences in cognition, even Mercier (2011) suggests that we should still expect to find fine-grained cultural differences in knowledge

---

[13] Although we don't want to press this concern here, we seriously doubt that ordinary folks perform knowledge attributions about real-world analogs in anything like their full Gettier-ized complexity.

[14] See Machery (2011), pp. 201-205, for a similar line of argument.

attributions, even against a widely-shared background of cognitive convergence. And we conjecture that these more subtle differences in folk epistemologies will be more likely to manifest in the unusual and marginal sorts of cases that are popular with epistemologists, than in ones that are evaluated and discussed in some frequency in civilian life. It is thus consistent with high baseline accuracy for epistemic intuitions in general that the sorts of intuitions that are important in analytic epistemology may suffer more from unwanted and unanticipated sensitivities.[15]

Third, a key feature of analytic epistemology involves the kinds of inferences in which our epistemic intuitions play a role.[16] Boyd and Nagel conclude with the claim that epistemic intuition is "reliable enough". But one question that must be asked here is, reliable enough *for what cognitive purposes*? For most of the purposes involved in everyday cognition, we are inclined to agree with Boyd and Nagel that epistemic intuitions are probably trustworthy enough to guide us in our quotidian transactions with the world. But our philosophical purposes are far more exacting, and one clear way to see this is in the kinds of inferential uses to which epistemic intuitions are put in epistemology, but not in ordinary life. It is not an exaggeration to think that *every* epistemological theory that is based at all on our epistemic intuitions is based on sophisticated inferences driven substantially by those intuitions – no one thinks that they can just read the correct epistemological theory directly off of the cases, non-inferentially. As such, to the extent to which we are interested in questions about baseline accuracy here, it seems that we should not particularly care about the baseline accuracy of the epistemic intuitions themselves, but rather on the inferences that will be based on those intuitions. We don't just want our theories to be based on premises that are mostly true; we want our theories themselves to be mostly true. What is at issue is not the obvious point that inferences should be conditionally reliable, but whether conditional reliability is enough, and there is reason to worry that it is not, and that some additional dimension of epistemic evaluation is needed.

To see why, consider two inferences rules that are equally reliable on all true inputs but whose reliability diverges sharply as soon as the quality of the inputs begins to degrade: one inference rule that takes ten propositions as inputs and outputs their ten-way conjunction, and another inference rule that takes the same ten propositions as inputs and outputs their ten-way disjunction. These two rules are maximally conditionally reliable; when all of the inputs are true, both will produce true outputs. And yet when the quality of the inputs degrades, so does their reliability – but not in the same way, and that is where this other dimension of evaluation can be seen. For the

---

[15] It is important to note that these concerns do not apply back to the more ordinary sorts of cases. This is one reason why an "Austinian" version of restrictionism may be well worth considering, in which the restriction in question is to cases that are well-attested in everyday discourse. However, though we think that such a restriction would be a good option to consider, it is neither necessary nor sufficient as a response to the restrictionist challenge. It is not sufficient because there is still likely *some* effect of context, demographic variation, etc. on at least some ordinary cases. And it is also not necessary, so long as one undertakes the restrictionist challenge's methodological recommendations, and explores more closely just where these effects do – and do *not* – apply. That is, we may learn that certain sorts of weird cases are not susceptible to any of these effects, and could thus be appealed to in greater confidence. The main point of the restrictionist challenge in all of this is simply that, even if that turns out to be true, then it is a fact that needs to be *learned*, and not something we can yet count on.

[16] Even if – as is surely the case – epistemological theorizing encompasses substantial resources beyond such intuitions, that does not blunt the worry here that very often the case verdicts play a real and substantial role as premises in epistemological arguments. It goes no distance to avoid the restrictionist challenge to point out the existence of such resources; rather, one would need to show that such resources render the case verdicts evidentially inert.

ten-way conjunction rule becomes maximally conditionally *un*reliable as soon as any of its inputs deviate from the truth, while the ten-way disjunction remains maximally conditionally reliable so long as at least one of its inputs is true. This dimension of the evaluation of rules of inference closely resembles a related issue in the evaluation of models, where we evaluate them in terms of *robustness*, or how well they withstand alterations to their basic assumptions and parameter-settings. So, let's call this dimension of inference evaluation *error-robustness*. Here we want to suggest that part of determining whether our epistemic intuitions are trustworthy involves determining how error-robust are the epistemological inferential practices that take epistemic intuitions as inputs. The more error-fragile one's inference rules are, the less tolerant of unwanted sensitivities one can afford to be in one's premises.[17]

And there is good reason to worry that the inferential practices of analytic epistemology are highly error-fragile. Remember that we are talking about a practice whose operating norms allow counterexamples to trump theory. Weatherson (2003) provides a nice description of this feature of analytic epistemology (and other areas of analytic philosophy):

> In epistemology, particularly in the theory of knowledge, and in parts of metaphysics, particularly in the theory of causation, it is almost universally assumed that intuition trumps theory. Shope's *The Analysis of Knowledge* contains literally dozens of cases where an interesting account of knowledge was jettisoned because it clashed with intuition about a particular case. In the literature on knowledge and lotteries it is not as widely assumed that intuitions about cases are inevitably correct, but this still seems to be the working hypothesis.

Weatherson immediately goes on to claim that epistemologists (and other philosophers) are wrong to let counterexamples trump theory, and indeed to argue very cleverly for this claim over the course of his paper.[18] But to the extent to which his description of the inferential practices of analytic epistemology is correct, and we think he is largely on target here, this suggests that the inferential practices of analytic epistemology will be highly error-fragile.[19] And this means that very little threat of error is needed in order to generate the kinds of methodological concerns that drive debates about the evidential status of epistemic intuitions.

4. Defending the Evidential Status of Epistemic Intuitions: Is There Life Beyond the Armchair?

Since the restrictionist challenge does not take a stand on the overall reliability of our epistemic intuitions, it makes little difference whether the overall baseline accuracy of our epistemic intuitions turns out to follow from some widely shared, ordinary folk capacity to think about epistemic issues. This suggests that a different kind of defense of the evidential status of epistemic intuitions in philosophical practice is needed, one that goes beyond merely affirming the baseline accuracy of folk epistemic intuitions, and is instead pitched more directly both as the specific sorts

---

[17] For further discussion of error-robustness, particularly in the context of Goldman's epistemology, see Gonnerman and Weinberg (forthcoming).
[18] Weatherson does also claim that, consistent with his arguments there, nonetheless epistemological practice can continue on *largely* as it has, and we think he is perhaps too optimistic about this (see Weinberg & Crowley 2010).
[19] Of course other forms of inference used by epistemologists may be less error-fragile. We are, again, only targeting current sorts of intuition-driven methods in epistemology.

of intuitional sensitivity that experimental philosophers are warning about, and at the specific contours of analytic epistemology. But what could that defense look like? There seem to be at least two options. One option involves taking the second part of the restrictionist challenge seriously by engaging in the kind of careful empirical work needed to understand where epistemic intuitions come from, what mechanisms are responsible for producing them, and what factors influence them. This means *more* empirical work for philosophers, not less. It also means that philosophers need to continue to improve the methods used to study philosophical cognition, combining survey methods with more advanced statistical methods and analyses, and supplementing survey methods with a wider variety of methods from the social and cognitive sciences (Alexander et al. 2010; Alexander 2012; Scholl 2007; Weinberg forthcoming). Perhaps most importantly, it means that philosophers have to resist the temptation to jump too quickly to broad philosophical conclusions on the basis of what individual studies seem to show. Science is slow business, and we need to resist the urge to make it go faster simply because that would better suit our philosophical goals.

Now, Boyd and Nagel are at least somewhat on the same page here, in that they happily acknowledge the importance and relevance of experimental work. And some of Nagel's own scientific work represents a high water mark in exactly the kind of careful investigation of epistemic intuitions that we are advocating here. Where we sharply diverge from their take on the current state of play, however, is on the *necessity* of such work. Because they only consider the question of baseline accuracy of epistemic intuitions, primarily with a focus on the question of "deep" intergroup differences, they are comfortable taking current philosophical practice to be more or less in good order as it stands. Although these practices can be usefully supplemented by experimental work, on their view, such work would basically be methodologically supererogatory. Intuitions are already, they say, "reliable enough". However, once we recognize that baseline accuracy is insufficient for methodological trustworthiness, it should also become clear that current intuitional practice cannot be adequately defended on the terms that Boyd and Nagel offer. There is too much threat of error from too many possible sources of sensitivity, and we're still at such an early stage of these investigations that mostly what we're learning is how much we yet have to learn, with strange new candidates for unwanted effects popping up every year. (For example, Nagel et al. (forthcoming) can be viewed as identifying a surprising new pattern of intuitional sensitivity in terms of empathy.) Given the nature of the restrictionist challenge, then, defending the reliability-as-trustworthiness-for-epistemological-purposes of intuitions will *require* the sorts of work we are advocating, and on a fairly expansive scale.

The understandable desire to avoid needing to shoulder such a burden may explain the popularity of another family of defenses: arguing that experimental philosophers haven't actually been studying the right *kind* of thing. These kinds of defenses look to preempt the restrictionist challenge before it can even get off the ground: if there are no appropriate studies that have yet been performed, then our armchair methods are not (yet) in a state of challenge. As Boyd and Nagel note, one particularly popular version of this kind of response, which we have called the *expertise defense*, involves arguing that philosophers are interested in expert philosophical intuitions rather than folk philosophical intuitions (Hales 2006, Ludwig 2007, Williamson 2007). But who has expertise about what and under what circumstances turns out to be a rather complicated empirical question. It seems that only certain kinds of training help improve task performance and, even then, only for certain kinds of tasks, and there is reason to worry that

philosophical training isn't the right kind of training and that philosophical thought-experimenting isn't the right kind of task (Weinberg et al. 2010, Alexander 2012). What's more, what empirical work has been done on expert philosophical intuitions doesn't look particularly promising for proponents of the expertise defense. These studies seem to suggest that expert intuitions display much of the same kind of intuitional sensitivity that folk intuitions display (Schwitzgebel & Cushman 2011, Schultz et al. 2011, Machery 2012, Knobe & Samuels 2013, Tobia et al. forthcoming). There is plenty of room for further empirical investigation here, and we would expect that philosophical training (or selection) will ward off at least *some* of the unwanted effects that may afflict folk epistemic intuitions. But we can see no intellectually respectable way at this time, given the current state of the literature, to hold to the expertise defense without any such further results in hand.

There is an illuminating comparison here to linguistics, especially if we connect the expertise question up with the issue of error-fragility discussed above. Boyd and Nagel cite some interesting recent work that suggests that syntactic intuitions of professional linguists are *generally* well-shielded from unwanted influence of the theoretical background of the intuiting linguists. But the sorts of inferences that linguists make are so error-fragile, that even the fairly small remaining degree intuitional error on the part of linguists poses a dire methodological threat. For example, Gibson et al. (2013) point out that even the low intuitional error rate of 2%-5%, as estimated by the authors cited by Boyd and Nagel, means that once you have more than about a half dozen to a dozen data points that are crucial to your linguistic inference, the odds of such an inference being correct start to drop rapidly below any minimally acceptable threshold of inferential accuracy. And the problem becomes compounded when, of course, we lack useful ways of determining which few of some large set of intuitions are going awry – in a nice convergence with the arguments of the restrictionists in experimental philosophy, Gibson et al. also emphasize the importance of experimental methods in recognizing and correcting errors in linguists' own intuitions. Expertise can at best only improve our intuitions so far, and the greater the error-fragility of our inferences, the smaller the consolation that can be provided by even a generous stretch of expert improvement.

A different way of trying to reduce the significance of experimental philosophy has been to shift attention from *whose* philosophical intuitions are relevant to *what* philosophical intuitions are relevant, a move that we have called the *thickness defense* because it typically involves adopting a "thick" conception of philosophical intuition, that is, a conception that includes specific semantic, phenomenological, etiological, or methodological conditions on what counts as a genuine philosophical intuition (Weinberg & Alexander forthcoming). The basic idea is both simple and attractive. If only certain kinds of mental states count as genuine epistemic intuitions, and these aren't the kinds of things that have been studied by experimental philosophers, then the kinds of methodological worries that have been raised aren't worries about the actual methods used in philosophical practice (Ludwig 2007, Kauppinen 2007, Cullen 2010, Bengson forthcoming). In our forthcoming paper, we argue that this strategy works only if certain conditions are met, conditions having to do with the propensity of "genuine" philosophical intuitions to avoid error and with our ability to successfully identify in practice which mental states count as genuine philosophical intuitions. If it turned our that philosophical intuitions were no more likely to track philosophical truth than the kinds of mental states studied by experimental philosophers, or that they were prone to other sorts of errors, then there would be little comfort in finding out that experimental philosophers haven't been studying genuine philosophical intuitions all along.

Likewise, there would be little comfort in knowing that experimental philosophers haven't been studying the right kind of mental states unless we already have ready at hand the resources needed to pick out those right kind of mental states. And, we argue that no version of the thickness defense currently on offer comes close to meeting all of these conditions: some versions treat philosophical intuitions in such a way that we either have good reason to worry that they won't track philosophical truth or else lack the means to distinguish them from other kinds of mental states, while other versions simply leave it an empirically open question whether or not the conditions have been met, which again means more experimental work will be needed, not less. As a conjectural roadmap to future research in experimental philosophy, some versions of the thickness defense may be promising, but as a means to avoid the challenge raised by experimental philosophy, they have thus far proved inadequate.

These conditions are almost certainly not exhaustive, but they do provide a place to start. For present purposes, what is important is that these conditions provide a framework for understanding what a successful defense of the evidential status of epistemic intuitions must look like. They also serve to remind us how important it is to understand what is at stake in this debate, and help us move beyond the tendency to measure success in the tedious terms of shifting burdens. When both sides start out with a clear sense of what moves (and countermoves) are available, we hope that the debate can begin to bring more light and less heat. We do think that one result is coming clearly into view at this point, though: there are vanishingly few moves left available to those who wish to defend the armchair while remaining in it. Too many questions pertinent to evaluating the trustworthiness of epistemic intuitions can only be addressed properly with some substantial reliance on scientific methods.

5. Conclusion

In the end we suspect that some ambiguity in the term "reliability" has muddied the debates about how these kinds of methodological debates should be framed, and consequently how we should score specific moves in these debates. Boyd and Nagel frame concerns about the evidential status of epistemic intuitions in terms of the overall baseline accuracy of our epistemic intuitions. Given this way of framing things, it makes perfect sense to argue in defense of GRT, and to do so in a largely defensive manner by pointing out that experimental philosophers haven't done enough to show that GRT is false. We hope to have shown that this is all rather beside the point, and that the restrictionist challenge should not be framed in terms of anything so strong as GRT, and moreover that the relevant experimental philosophy papers should not be read as trying to pursue such a target. Given the purposes and needs of analytic epistemology, all that is needed in order to establish the "unreliability" of epistemic intuitions – not once and for all, but only under our current state of substantial ignorance about them – is something much weaker. All that is needed to raise the restrictionist challenge is the existence of an empirically plausible threat of unwelcome and unexpected intuitional sensitivity. Boyd and Nagel successfully argue that the very high bar of falsifying the GRT would be a perilously difficult leap to attempt. Thankfully, restrictionists can safely operate at altitudes much closer to the ground, and still make plenty of trouble for current philosophical practices with epistemic intuitions.

Bibliography

Alexander, J. 2010: Is experimental philosophy philosophically significant? *Philosophical Psychology* 23: 377-389.

Alexander, J. 2012: *Experimental Philosophy: An Introduction*, Polity Press.

Alexander, J., and Weinberg, J. 2007: Analytic epistemology and experimental philosophy. *Philosophy Compass* 2: 56-80.

Bealer, G. 1998: Intuition and the autonomy of philosophy. In M. DePaul and W. Ramsey (eds), *Rethinking Intuition*, Rowman and Littlefield, pp. 201-240.

Beebe, J., and Buckwalter, W. 2010: The epistemic side-effect effect. *Mind & Language* 25: 474-498.

Bengson, J. Forthcoming: Experimental attacks on intuitions and answers. To appear in *Philosophy and Phenomenological Research*.

Colaco, D., Buckwalter, W., and Stich, S. Forthcoming: Epistemic intuitions in fake-barn thought experiments. To appear in *Episteme*.

Cappelen, H. 2012. *Philosophy Without Intuitions*, Oxford University Press.

Cullen, S. 2010: Survey-driven romanticism. *Review of Philosophy and Psychology* 1: 275-296.

Gettier, E. 1963: Is Justified True Belief Knowledge? *Analysis* 23: 121-123.

Cummins, R. 1998: Reflections on reflective equilibrium. In M. DePaul and W. Ramsey (eds), *Rethinking Intuition*, Rowman and Littlefield, pp. 113-128.

Deutsch, M. 2011: Intuitions, counter-examples, and experimental philosophy. *Review of Philosophy and Psychology* 1: 447-460.

Feltz, A., and Cokely, E. 2009: Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition* 18: 342-350.

Gibson, E., Piantadosi, S., and Fedorenko, E. 2013: Quantitative methods in syntax/semantics research: A response to Sprouse and Almeida. In press at *Language and Cognitive Processes*.

Goldman, A. 1979: What is justified true belief? In G. Pappas (ed) *Justification and Knowledge*, D. Reidel, pp. 1-23.

Goldman, A. 2007: Philosophical intuitions: Their target, their source, and their epistemic status. *Grazer Philosophische Studien* 74: 1-26.

Gonnerman, C., and Weinberg, J. Forthcoming: Why messy minds make for messier armchairs: Goldman and the epistemology of categorizer pluralism. To appear in H. Kornblith and B. McLaughlin (eds) *Goldman & His Critics*. Wiley-Blackwell.

Hales, S. 2006: *Relativism and the Foundations of Philosophy*. MIT Press.

Henrich, J., Heine, S., and Norenzayan, A. 2010: The weirdest people in the world? *Behavioral and Brain Sciences* 33: 1-75.

Kauppinen, A. 2007: The rise and fall of experimental philosophy. *Philosophical Explorations* 10: 95-118.

Knobe, J., and Samuels, R. 2013: Thinking like a scientist: Innateness as a case study. *Cognition* 126: 72-86.

Kornblith, H. 2002: *Knowledge and Its Place in Nature*, Oxford University Press.

Levin, J. 2005: The evidential status of philosophical intuition. *Philosophical Studies* 121: 193-224.

Ludwig, K. 2007: The epistemology of thought experiments: First person versus third person approaches. *Midwest Studies in Philosophy* 31: 128-159.

Machery, E. 2011: Thought Experiments and Philosophical Knowledge. *Metaphilosophy* 42: 191-214.

Machery, E. 2012: Expertise and intuitions about reference. *Theoria* 27: 37-54.

Mercier, H. 2011: On the universality of argumentative reasoning. *Journal of Cognition and Culture* 11: 85-113.

Nagel, J., San Juan, V., and Mar, R. Forthcoming: Lay denial of knowledge for justified true beliefs. To appear in *Cognition*.

Nisbett, R., Peng, K., Choi, I., and Norenzayan, A. 2001: Culture and systems of thought: holistic versus analytic cognition. *Psychological Review* 108: 291-310.

Page, S. 2008: *The Difference*. Princeton University Press.

Pust, J. 2000: *Intuitions as Evidence*. Routledge.

Scholl, B. 2007: Object persistence in philosophy and psychology. *Mind & Language* 22: 563-591.

Schultz, E., Cokely, E., and Feltz, A. 2011: Persistent bias in expert judgments about free will and moral responsibility: A test of the expertise defense. *Consciousness and Cognition* 20: 1722-1731.

Schwitzgebel, E., and Cushman, F. 2012: Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language* 27: 135-153.

Sosa, E. 2007: Experimental philosophy and philosophical intuition. *Philosophical Studies* 132: 99-107.

Sosa, E. 2009: A defense of the use of intuitions in philosophy. In M. Bishop and D. Murphy (eds), *Stich and His Critics*, Wiley-Blackwell, pp. 101-112.

Stich, S. 1988: Reflective equilibrium, analytic epistemology, and the problem of cognitive diversity. *Synthese* 74: 391-413.

Swain, S., Alexander, J., and Weinberg, J. 2008: The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophical and Phenomenological Research* 76: 138-155.

Tobia, K., Buckwalter, W., and Stich, S. Forthcoming: Moral Intuitions: Are Philosophers Experts? To appear in *Philosophical Psychology*

Weatherson, B. 2003: What good are counterexamples? *Philosophical Studies* 115: 1-31.

Weigel, C. 2011: Distance, anger, freedom: An account of the role of abstraction in compatibilist and incompatibilist intuitions. *Philosophical Psychology* 24: 803-823.

Weinberg, J. 2007: How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy* 31: 318-343.

Weinberg, J. Forthcoming: The promise of experimental philosophy & the inference to signal. To appear in J. Beebe (ed) *Advances in Experimental Epistemology*, Continuum Press.

Weinberg, J., and Alexander, J. Forthcoming: The challenge of sticking with intuitions through thick and thin. To appear in A. Booth and D. Rowbottom (eds), *Intuition,* Oxford University Press.

Weinberg, J., and Crowley, S. 2010: Loose constitutivity and armchair philosophy. *Studia Philosophica Estonica* 2: 177-195.

Weinberg, J., Gonnerman, C., Buckner, C., and Alexander, J. 2010: Are philosophers expert intuiters? *Philosophical Psychology* 23: 331-355.

Weinberg, J., Nichols, S., and Stich, S. 2001: Normativity and epistemic intuitions. *Philosophical Topics* 29: 429-460.

Williamson, T. 2007: *The Philosophy of Philosophy*, Blackwell Publishing.

Zamzow, J., and Nichols, S. 2009: Variations in ethical intuitions. *Philosophical Issues* 19, 368-388.