Practices make perfect: On minding methodology when mooting metaphilosophy[1]

To appear in *Oxford Studies in Experimental Philosophy*

Joshua Alexander
Siena College

Jonathan M. Weinberg
University of Arizona

Recent debates about armchair philosophical methodology, and especially the *method of cases*, have mostly focused on the epistemological status of *philosophical intuitions*; that is, about whether some particular kind of mental state or process has the right kind of characteristics to produce some kind of classic epistemic good such as evidence, justification, or knowledge. This is unsurprising since questions about what to say about *these* kinds of mental states or processes in *these* kinds of terms is perhaps the most natural meeting place for the host of psychological and philosophical issues that are central to these metaphilosophical debates.

This way of thinking about armchair philosophy and the method of cases came to dominate the first decade of metaphilosophical work in experimental philosophy, where the *experimental challenge* was often framed in terms of worries about the *trustworthiness of philosophical intuitions* (see, for history and discussion, Alexander 2012). Joachim Horvath (2010, 448) offers the following general schema for this way of thinking about the experimental challenge:

> (1) Intuitions about hypothetical cases vary with irrelevant factors.
>
> (2) If intuitions about hypothetical cases vary with irrelevant factors, then they are not epistemically trustworthy.
>
> (C) Intuitions about hypothetical cases are not epistemically trustworthy.

In recent years, however, we have seen much needed changes to this way of thinking about the experimental challenge. Jennifer Nado (2016a) has argued, for example, that debates about philosophical methodology should not be framed in classic epistemological terms. In a nutshell, concepts appropriate for the evaluation of ordinary cognitive activity and achievement will often prove inadequate for the distinct, and often more demanding, purposes of philosophical inquiry. And, David Colaço and Edouard Machery (2017) have argued that it has been a mistake to focus so much attention

[1] Our thanks to Shaun Nichols and to two anonymous referees for their generous comments.

on philosophical *intuitions*, per se. What matters, they say, is the assessment of philosophical thought experiments no matter what elements of our psychology are involved in that assessment.

We very much concur with Nado, Colaço, and Machery.[2] In fact, we think that philosophers who want to think carefully about the method of cases should push even a bit further away from the traditional way of thinking about armchair philosophical methodology. When it comes to debates about the method of cases, we think that what really matters has less to do with the psychology involved in reaching verdicts about philosophical cases than it does with the attendant *philosophical practices* that surround and ground how we use those verdicts in philosophical inquiry. We won't argue for that comparative claim here. Instead, what we want to do in this paper is to engage with two recent attempts to respond to the experimental challenge from the perspective of these philosophical practices, and to show that they fail because they are not sufficiently attentive to the particulars of those practices.

The first response comes from Max Deutsch (2010, 2015) and Herman Cappelen (2012), who have recently argued that the experimental challenge involves a fundamental misunderstanding about the nature of philosophical practice. Here's how Deutsch (2015, xv) puts the worry:

> Analytic philosophy is chock-full of hypothetical examples and thought experiments, of course, but analytic philosophers *argue* for their claims about what is or is not true in these cases and thought experiments. It is these arguments, *not* intuitions, that are, and should be, treated as evidence for the claims.

The basic idea is this. Experimental philosophers focus *too much attention* on something that turns out not to matter all that much to philosophical practice, at least from an evidential point of view, namely, philosophical intuitions. What's worse, the way that experimental philosophers talk about philosophy focuses *too little attention* on what really does matter to philosophical practice, namely, philosophical arguments. If this is right, then philosophers can sidestep the experimental challenge to the method of cases altogether. They can grant the conclusion of Horvath's schema, while denying that it is a claim that should disturb practitioners of the method of cases. Nothing that we can learn from studying philosophical cognition is relevant to philosophical methodology since our philosophical methods involve arguments, not intuitions.

The second response comes from Joshua Knobe (2019, 2021), who has recently argued that the experimental challenge fundamentally misunderstands what's been happening in experimental philosophy, in particular recent experimental work that seems to suggest that a surprising number of philosophical intuitions do not vary with irrelevant features. This way of responding to the experimental challenge does not attempt to evade the experimental challenge by cutting it off at the

[2] Having said that, since this is the way that things are framed in the particular debate that we want to talk about in what follows, we will continue to talk about the experimental challenge in these terms, that is, in terms of philosophical intuitions and epistemic goods. We will treat this all in a black-box manner, however, and will take intuition-talk to stand for whatever psychological processes are involved when philosophers use verdicts about thought experiments as evidence on any of the standard conceptions of the method of cases.

pass; instead, it attempts to respond directly by cutting it off at the knees, by rejecting the first premises of Horvath's schema. As Knobe (2021, 69-70) writes,

> If people's intuitions are surprisingly stable, then the whole debate about the implications of instability is moot. As we noted at the outset, there has been an enormous amount of research focused on the question: "If we learn that people's intuitions are unstable, what should we conclude about the use of intuitions in philosophy?" Such research has shown impressive levels of sophistication and ingenuity, but if people's intuitions are not in fact unstable, this is simply not the question we face.

We think that both of these attempts to respond to the experimental challenge fail. That's perhaps unsurprising given the skin we have in this game. What we think is surprising, and our reason for writing this paper, is *why* they fail and what this tells us about the way that philosophers should approach questions about philosophical methodology. In a nutshell, they fail because they take their eyes off the most important part of the game: the methodological practices themselves.

1. Arguments all the way down?

Let's start with the suggestion that case verdicts do not actually figure in our philosophical methods. Are Deutsch and Cappelen right that all, or even most, of the evidential work being done in philosophy is being done by philosophical arguments rather than by what philosophers think about philosophical cases? Deutsch and Cappelen advance this view by proposing reinterpretations of many famous "case-based" philosophical books and papers according to which the case verdicts that seem to play a central role in these papers depend on philosophical arguments. One problem with this strategy is that their "no-intuition" interpretations of these cases have struck many readers, including us, as controversial at best.[3] Here we will focus on two of their most prominent case studies, Gettier (1963) and Lehrer (1990), and argue that their method of, in essence, squinting very hard at the target texts has serious shortcomings and leaves the debate potentially stalemated. In short, because the relevant textual passages are fairly brief, evidence for any interpretation of what's going on in those passages runs out quickly, and there's something inherently unresolvable in debates about long-past mental states of philosophers. After raising worries about the way that they try to establish their interpretation of these cases, we will show what we can learn about these cases by paying closer attention to the broader contours of ordinary philosophical practice. We think that this practice-oriented methodological approach can bring substantially more evidence to bear, and so we hope to be able to achieve the first-order result of pushing back hard against the no-intuition interpretation and the second-order result of underscoring the importance of attending to practices.

Let's start with Gettier, since Deutsch offers such a sustained and thought-through attempt to interpret this classic text in terms of arguments, and not intuitions about the cases. Most people who

[3] David Colaço and Edouard Machery (2017) make a similar observation in their review of Deutsch's book; Avner Baz also does so quite forcefully in his (2017). Cappelen's argument has other problems, most glaringly that he operationalizes "intuition" in an implausibly strong way that makes it almost trivial that philosophers do not use intuitions *sensu* Cappelen. For discussion, see Weinberg (2014).

have written about Gettier's thought experiments, and there have been lots and lots of them, have agreed both that it sure seems an awful lot like Smith lacks knowledge and that it sure seems an awful lot like Gettier thinks that the fact that it seems an awful lot like Smith lacks knowledge counts as evidence that sometimes our justified true beliefs do not count as knowledge. Deutsch disagrees, arguing that although Gettier's case against the view that knowledge is justified true belief centers on the claim that Smith's justified true beliefs do not count as knowledge, Gettier does not rely on this claim because it is intuitive, but instead because he has a good argument that Smith's justified true beliefs do not count as knowledge. Where is the argument? Deutsch points our attention to the very end of the first of Gettier's two cases:

> But it is equally clear that Smith does not *know* that (e) is true; for (e) is true in virtue of the number of coins in Smith's pocket, while Smith does not know how many coins are in Smith's pocket, and bases his belief in (e) on a count of the coins in Jones's pocket, whom he falsely believes to be the man who will get the job.

According to Deutsch, all of the argumentative work is being done by this sentence. On Deutsch's reading of this sentence, Gettier's case that knowledge is not simply justified true belief depends on how we interpret Gettier's use of the word 'for'. Deutsch thinks that Gettier is using the word 'for' as a premise indicator, and this makes it seem rather clear to him that Gettier intends what follows to count as evidence that Smith does not know that (e).

We have substantial worries about Deutsch's interpretation of the passage. But before we get into those, let's suppose for the sake of argument that he is right and Gettier is indeed arguing to, rather than from, the case verdict in question. It is important to see that this concession is not enough to really put all that much pressure on the idea that experimental philosophy is relevant to how we should think about philosophical methodology. Here's why. If Getter is making an argument here, then he is *not making very much of one* (Mallon 2017 presses this point). Deutsch (2017) seems willing to acknowledge this point, and suggests that much of the argumentative work being done when philosophers use thought experiments might be happening off-stage, so to speak, in the heads of the philosophers who are constructing the cases and the readers who are consuming them. This includes what evidence philosophers, both *producers* and *consumers* of thought experiments, think they have for the premises being appealed in those arguments, explicitly or otherwise. But the social and cognitive sciences give us lots of reasons to worry that our evaluations of what evidence we have for our beliefs depends a lot on those beliefs, and so what we happen to believe and what kinds of things influence what we happen to believe is always going to be relevant, although perhaps in a less direct way than is sometimes suggested. This is more than enough to get experimental philosophy, and the kinds of worries that experimental philosophers like us have wanted to raise and address about how philosophers use the method of cases, up and running. Experimental philosophy is relevant to understanding philosophical methodology because, *whatever it turns out to be*, on Deustch's account, that philosophers are tacitly invoking in their arguments, we know that the tacitness of those invocations suggests that they may well be susceptible to unwanted biases and influences, which are just the sort of thing that experimental tools and methods are needed in order to root out. That philosophers would, of course, want to avoid any such errors caused by our unconscious cognition provides plenty of reason to recognize the relevance of experimental philosophy.

Putting this aside, Deutsch's interpretation of what is going on in the passage quoted earlier is far from mandatory. There are other ways to read how Gettier is using the word 'for' in that passage, readings that are equally plausible and that challenge the idea that Gettier intends what follows to count as evidence that Smith does not know that (e). According to one alternative reading, Gettier is using the word 'for' to signal the start of an *explanation* for why Smith does not know that (e).[4] The difference between these two readings is subtle, to be sure, but it is also important. According to Deutsch's reading, Gettier starts with something like the idea that knowledge is incompatible with certain kinds of epistemic luck and, apparently expecting such a generalization to be acceptable as a shared premise, argues from there that Smith does not know that (e). According to the more common reading, however, Gettier starts with the intuitive claim that Smith does not know that (e) and explains why Smith does not know that (e) by highlighting the role that a specific kind of epistemic luck plays in his cases. William Ramsey (2019), offers another alternative reading, according to which thought experiments like Gettier's are intended to provoke an intuition in readers, and the supporting text is offered in order to help *cue* the intuition that is intended. As he explains it,

> thought experiments provide a prompt for an intended psychological reaction. But such a prompt works only if the audience grasps the relevant details of the scenario. To ensure this, philosophers often reiterate salient elements so that the intended target of persuasion realizes the key specifics. This is what Gettier is doing in this sentence [i.e., the one quoted above]. His example is complicated, so he simply reemphasizes the most pertinent aspects for anyone who didn't follow along, and thereby failed to have the intended reaction. In this way, Gettier is acting like someone trying to prod the recall of a crucial memory by relaying noteworthy aspects of whatever occurrence the memory is about. (92-3)

Similarly, consider trying to get students to see an alternate percept in a bistable illusion case, for example, they are stuck seeing the duck-rabbit only as a rabbit. You might say to them things like, "you can see it as a duck, for the rabbit's ears could be like the two halves of a duck's bill, and the duck is facing left instead of right", and in doing so, you would in no way thereby be *arguing* for the anatine visual interpretation over the leporine. You are aiding the student in seeing it for themselves, and coming to their own unargued-for perceptual realization. Which of these readings of Gettier is right? If we just stare hard at this short stretch of text, this question seems, at best, highly debatable. The high plausibility of the two[5] alternative readings that we have sketched here precludes any easy inference to the claim that, while Gettier treats the claim that Smith does not know that (e) as evidence,

[4] Colaço & Machery (2017) also press this point. Cappelen and Deutsch are aware of this distinction, and Deutsch attempts to address the alternative reading in his (2016). We don't find the reasons that he gives in favor of the argumentative reading of the relevant text over explanatory reading compelling for reasons that we get into below.

[5] And indeed, likely more than two -- Bengson (2014) catalogues many further supporting-but-not-supplanting uses of argumentative language in the context of the method of cases.

he only does so because he has an argument for it. This sort of dispute is going to be hopelessly mootable so long as philosophers just engage in hermeneutic squinting, but as we will argue shortly, once we consider the much broader set of evidence about philosophical practices, the question becomes much less moot.

Looking beyond this brief passage itself, what really should matter for these debates is not so much what Gettier took himself to be doing, but what everyone else has taken Gettier to be doing. Not that Gettier's own self-understanding is irrelevant, by any means. But if most epistemologists did not read him as offering an argument, but rather an intuition, and thereupon engaged in a practice that trades in intuitions, and if that practice is the one that continues today, then that clearly will be more salient to how we understand the method of cases, and philosophical methodology more generally, than Gettier's three pages from half a century ago. Deutsch does make an attempt at some of this in his (2016), but we fear his examples do not ultimately work in his favor. At best they suggest one possible way to read some philosophers who, to our eyes, are theorizing about an anti-luck condition, a condition that seems to us best understood as independently motivated by case verdicts, rather than arguing from an anti-luck condition to establish those case verdicts. What's more, it doesn't seem to us that Deutsch's readings of these philosophers and their work are well-supported by the total evidence available. So, for example, Deutsch talks about Alvin Goldman's discussion of Gettier cases in his famous (1967) article "A causal theory of knowing." But in the first passage that Desutsch quotes, where Goldman is describing his strategy for the article, Goldman writes "Michael Clark, for example, points to the fact that *q* is false and suggests that as the reason why Smith cannot be said to know *p*… I shall make another hypothesis *to account for the fact* that Smith cannot be said to know *p*, and I shall generalize this into a new analysis of "*S* knows *p*" (Goldman 1967, 358; emphasis added). It seems clear to us that Goldman's plan is to start from the position that Smith does not know that *p*, explain why Smith's justified true belief that *p* doesn't count as knowledge, and then use this explanation to build a new theory of knowledge. It does not, in other words, seem to us that Goldman is planning to use his theory of knowledge to argue that Smith doesn't know that *p*, which is what Deutsch needs Goldman to be planning to do in order for the strategy that he is taking in his (2016) paper to work.[6]

While Deutsch's strategy is his (2016) paper doesn't seem to us to be persuasive, let's grant it to him as a defensive move, that is, that some epistemologists may have read Gettier the way that he suggests. The problem is that, even when we grant this to Deutsch, it is just not hard to find other, very highly-regarded epistemologists who patently cannot be assimilated to this argument interpretation. Here we will just focus on two. First, Robert Shope (2002, 30-31), a highly engaged participant in the 'S knows

[6] It is actually unsurprising that Goldman seems to start with the intuition that Smith doesn't know that *p* given all of the rather important and influential work (for example, Goldman & Pust 1998, Goldman 2007) he has done on the role that intuitions play in philosophical methodology, and so it is somewhat odd that Deutsch uses him as a prime example here. In fact, the first sentence of Goldman (2017), which in fairness appeared after Deutsch's article, makes it clear precisely what Goldman's take is on this debate: "Reactions to Gettier's (1963) paper demonstrated the powerful role of intuitions in philosophical methodology."

that P' debates and widely considered *the* expert on Gettierology, writes in a handbook piece on the Gettier problem that

> Gettier offered no diagnosis of these examples and no formula for constructing further examples that he was prepared to regard as of the same type. But as other philosophers proceeded to offer additional examples that they regarded as importantly similar to one or another's of Gettier's, the technical label, 'Gettier-type example', sprang into use.

This does not sound like someone who took Gettier, or the crowd of respondents to Gettier, to have agreed on an anti-luck premise that they were just further investigating! It sounds, instead, very much like someone who found himself, and took his colleagues to similarly find themselves, with an unargued-for evidential access to the verdicts about Gettier's and a great many and sundry similar cases.

Second, in the later editions (for example, (1989)) of Roderick Chisholm's *Theory of Knowledge*, he chooses to present only the second of Gettier's cases without the disputably argumentative material at the end of the first case. He describes Gettier as having "noted" - not a verb of argumentation or inference - that the situation in the case "is counter to the traditional theory of knowledge" (91), and in his own presentation of various cases, Chisholm's typical style is to present the case sharply and efficiently, and issue a verdict, with nothing at all offered in support even of the minimal sort that Deutsch proposes to find in Gettier's own paper. Since Chisholm was arguably the most prominent and influential analytic epistemologist of the day, it seems unlikely that his way of approaching Gettier was a highly aberrant one-off. For what it's worth, what would be very helpful at this point would be for some historians of analytic philosophy to offer a scholarly take on just how mid-century epistemologists understood their methods; we suspect that some combination of ordinary language philosophy and Chisholm's own "particularism" was highly conducive to what would now look like an intuition-based methodology, but that is basically just conjecture on our part.[7]

So much for Deutsch's interpretation of Gettier's famous thought experiments, let's turn our attention now to Herman Cappelen's analysis of Lehrer's famous *Truetemp* case. The case involves a person who is a perfectly reliable estimator of local temperatures, but who is unaware of this reliable capacity and has no reason whatsoever to think he has such a capacity. Can this person be said properly to *know* the temperature, when he estimates it? Lehrer, on almost everyone's reading, invites readers to share the intuition that none of Truetemp's true beliefs count as knowledge. And this intuition is used to make a lot of trouble for *externalist* theories of knowledge, since at least some of those theories, especially *reliabilist* ones, would seem to predict that Truetemp does know. On this common reading, Lehrer is appealing to intuitions about the case and arguing from them towards a rejection of externalism. Cappelen construes matters the other way around, using a bit of Lehrer's text:

> The primary argument [that Truetemp doesn't know] goes something like this: "More than the possession of correct information is required for knowledge. One must have some way of

[7] Keep in mind that Gettier was himself a student of Norman Malcolm.

> knowing the information is correct." ([Lehrer] p. 188) Since Mr. Truetemp has no way of knowing that the information is correct, he does not know. (168)

As we saw with Deutsch's interpretation of Gettier, just attending closely to a few sentences of text can make it hard to determine precisely what is the right way to read those sentences. And Lehrer's own words do invite some confusion. For example, after describing the case, Lehrer almost immediately goes on to say: "the sort of causal, nomological, statistical, or counterfactual relationships required by externalism may all be present. *Does he know that the temperature is 104 degrees when the thought occurs to him while strolling in Pima Canyon? He has no idea why the thought occurred to him or that such thoughts are almost always correct. He doesn't, consequently, know that the temperature is 104 degrees when that thought occurs to him*" (187; emphasis added by Cappelen). Now, Cappelen looks at this and declares, "I don't know how to read this other than as Lehrer giving an argument in favor of a certain answer to the question." (170)[8] *Pace* Cappelen, it seems to us entirely clear to us how to read aspects of this passage in terms of the explanatory interpretation that we rehearsed above. That 'consequently' could be inferential, as Cappelen suggests. But it could also easily be explanatory, if we take Lehrer to be laying out how his internalist hypothesis correctly predicts the intuited verdict. (It is admittedly harder to see how to apply the cueing interpretation to the specific term here.) Since Cappelen says nothing more about his interpretive blockage, let's just leave the point here: the method of peering deep into the soul of a couple of quoted passages is not up to resolving such interpretive disputes.

But, once again, when we step back from the vignettes themselves and their immediate textual surroundings, there is plenty more evidence to be found. First, there are *other* parts of Lehrer's text that strongly indicate that he cannot be arguing in the principle-to-case direction, but the other way around. Lehrer is very explicitly embedding his discussion in a dialectic with externalists, and he means to be giving reasons that those theorists could themselves recognize and take to be problematic for their externalist views. It directly follows that we just can't take that "He has no idea why the thought occurred to him or that such thoughts are almost always correct" as any sort of shared premise between Lehrer and the externalist. Indeed, at the start of the relevant chapter, Lehrer says this plainly: "The central tenet of externalism is that some relationship to the external world accounting for the truth of our belief suffices to convert true belief to knowledge *without our having any idea of that relationship*" (177; emphasis added). What Cappelen takes to be a premise is what Lehrer clearly states is the very issue at stake with his opponents. *Contra* Cappelen, once we bring into view the broader argumentative context in which the case is embedded, we don't see how you can read this as Lehrer giving an argument in favor of a certain answer to the question. As Landes (2020, 19) observes, in an excellent elaboration of this argument:

> Lehrer takes himself to have access to what knowledge is independently of the externalist theories of knowledge he is discussing. For this access to justify his rejection of externalism, it cannot be derived from an internalist theory of knowledge. Otherwise, because this passage is

[8] Deutsch concurs: "It is very difficult to see how anyone could read this passage as anything other than an attempt to provide reasons for the judgment" that Truetemp doesn't know (112).

> used to argue against the externalist views of Armstrong, Nozick, Goldman, Dretske, and others, Lehrer would be flagrantly begging the question against his opponents."[9]

Moreover, it's hard for the no-intuitions, arguing-from-principles-to-cases approach to make sense of how Lehrer introduces this whole line of argument (186):

> The general opacity problem with externalism can be seen most graphically by considering the analogy proposed by Armstrong. He suggested that the right model of knowledge is a thermometer. The relationship between the reading on a thermometer and the temperature of the object illustrates the theories mentioned above. Suppose that the thermometer is an accurate one and that it records a temperature of 104 degrees for some oil it is used to measure. … The problem with the analogy is that the thermometer is obviously ignorant of the temperature it records. The question is - why?

There's plainly no argument to the verdict there; just the flat, if utterly plausible, declaration of its obvious truth. And the question then sure seems to be an explanatory one: we have an intuition, and we can now try to make some philosophical headway by explaining it. We take this to be some further evidence of Lehrer's approach to his Truetemp case as well, since Lehrer presents the Truetemp case as in essence an elaboration of the literal thermometer case.

As we also saw with Deutsch's interpretation of Gettier, since we are investigating a widely-deployed philosophical methodology, what the original author of a case may have thought about it is much less important than how the profession picked up and ran with it from there.[10] And here's a useful observation about the analytic epistemology literature since the 1980s-90s: Lehrer's Truetemp case often shows up in discussions juxtaposed with highly similar set of cases due to Laurence BonJour (1985), about a set of clairvoyants who nonetheless lacks any clue that they are, in fact, clairvoyant, and whose clairvoyant beliefs are thus in much the same epistemic boat as Truetemp's meteorological ones. Indeed, the differences between the cases are considered cosmetic enough that epistemologists on the whole have taken Lehrer's Truetemp and BonJour's clairvoyants to be *more or less indistinguishable*. For just one recent example, here's Mylan Engel (2022, 41):

> Internal reasons reliabilism also yields the intuitively correct verdict on both BonJour's Norman [one of the clairvoyants] case and Lehrer's TrueTemp case. Why? Because both Norman and TrueTemp lack internal reasons for their respective beliefs.

Some quick poking about on Google Scholar will demonstrate how widespread is this general assimilation of the two cases. We thus propose that we have good *prima facie* reason to take it that the

[9] Although we don't want to pursue this line of argument here, Landes makes an excellent case that similar concerns can be raised for any attempt to construct a non-question-begging version of the alleged Gettier argument in the Smith case.

[10] Cappelen claims in an aside, "It's important to focus on the original text, not the argument as it is idealized in the later literature." (169) For a scholar of the original author, surely that is so. For methodologists and metaphilosophers, however, Cappelen's advice is perfectly wrong. See also Colaco and Machery (2017, 410-1) on the importance of considering the full historical context of arguments and methods in use.

epistemological community on the whole thinks that these cases work in fundamentally the same way. Most importantly here: if either is a principle-to-case argument, then the other ought be interpreted as such as well, and vice versa.

But, unlike the potential interpretive difficulties induced during a hyper close inspection of Gettier's use of the word 'for', or Lehrer's 'consequently', BonJour is exceedingly clear on how the order of argument is meant to go:

> But it seems intuitively clear nevertheless that this is not a case of … knowledge: Samantha [another clairvoyant] is being thoroughly irrational and irresponsible in disregarding cogent evidence that the President is not in New York City on the basis of a clairvoyant power which she has no reason at all to think that she possesses; and this irrationality is not somehow canceled by the fact that she happens to be right. Thus, I submit, Samantha's irrationality and irresponsibility prevent her belief from being epistemically justified.

Now, the second half of that quoted paragraph, especially that 'thus', certainly can make it look like BonJour is arguing for the conclusion that the clairvoyant doesn't know on the basis of premises about irrationality of a certain sort precluding knowledge. It can seem we're in another argument versus explanation bind. But then see where BonJour goes to immediately from there:

> The case and others like it [i.e., all the clairvoyant cases] suggest the need for a further condition to supplement Armstrong's original one: not only must it be true that there is a law-like connection between a person's belief and the state of affairs that makes it true, such that given the belief, the state of affairs cannot fail to obtain, but it must also be true that the person in question does not possess cogent reasons for thinking that the belief in question is false. For, as this case seems to show, the possession of such reasons renders the acceptance of the belief irrational in a way that cannot be overridden by a purely externalist justification. (60)

The second quoted paragraph reveals why the "arguing for the verdict" interpretation could not actually be a good reading of the first paragraph: the case verdict itself is what is doing the argumentative work, as a premise, in revealing to us the relevant claim about irrationality and knowledge, which is itself the conclusion in turn.

That the cases are leading the way is further substantiated by BonJour's own account of his methodology, when he writes earlier in that paper that the radical nature of externalism makes it hard to find much common ground at the level of general principles about knowledge or justification:

> The problem, however, is that this very radicalism has the effect of insulating the externalist from any very direct refutation: any attempt at such a refutation is almost certain to appeal to premises that a thoroughgoing externalist would not accept.
>
> My solution to this threatened impasse will be to proceed on an intuitive level as far as possible. By considering a series of examples, I shall attempt to exhibit as clearly as possible the fundamental intuition about epistemic rationality that externalism seems to violate. Although this intuition may not constitute a conclusive objection to the view, it is enough, I believe, to shift the burden of proof decisively to the externalist. (56)

This description of his own methodological approach should substantially undermine any attempt to read BonJour, or Lehrer with him, as offering arguments from the epistemic generalities to the cases. He needs the case verdicts themselves to be doing the argumentative work.

And BonJour is not being idiosyncratic here. Our reading of the literature is that it is very common, when philosophers deploy arguments with case verdicts in them, that these case verdicts are methodologically basic in the manner that we have suggested here. There are many other examples where philosophers seem to say that their capacity to offer any further defense of a key premise has run aground. So, for example, Jerry Fodor (1997, 154), writes

> As with most of the metaphysical claims one comes across these days, the one that I just made relies for its warrant on a blatant appeal to modal intuitions. But I think the modal intuitions that I'm mongering are pretty clearly the right ones to have. If you don't share mine, perhaps you need to have yours looked at.

In similar spirit, David Lewis (1996, 561) writes

> I started with a puzzle: how can it be, when his conclusion is so silly, that the skeptic's argument is so irresistible? My Rule of Attention, and the version of the proviso that made that Rule trivial, were built to explain how the skeptic manages to sway us – why his argument seems irresistible, however temporarily. If you continue to find it eminently resistible in all contexts, you have no need of any such explanation. We just disagree about the explanandum phenomenon.[11]

Over and over again we find case verdicts being used in the way that we have described them being used. They are very frequently not, in fact, argued for from independently convincing and available premises, at least not in a way that does not build on their methodological basicness in the first place. Nor is there any ready-to-hand generally available evidence to offer on their behalf, which can be presupposed to be antecedently shared, of a sort that we can find intelligible in terms of observation of testimony from experts, or for that matter, from things like the deliverance of well-calibrated instruments, the products of scientific consensus, and so on. Philosophers seem on the whole fairly happy to use them as evidence, even in the absence of any further evidence on their behalf. The profession seems to be in a state of substantial consensus, then, that we have some sort of cognitive grasp or other on these claims. And this consensus also extends pretty well to the basic contours of this capacity, for example, that it includes, but is not at all limited to, a pretty open-ended range of hypothetical cases with stipulated facts, including perhaps nomologically impossible ones. What the profession seems also to be in a state of substantial dissensus about, however, is just what this grasp ultimately amounts to, as illustrated by the many-hued panoply of extant accounts of case verdicts that

[11] Although Lewis talks about arguments in this passage, it is clear in the article that he doesn't have in mind anything like what Deutsch has in mind when he claims that the method of cases involves arguments. Here's how Lewis describes what he has in mind: "The sceptical argument is nothing new or fancy. It is just this: it seems as if knowledge must be by definition infallible." (1996, 549) That sure looks a lot like an intuition and not an argument.

we rehearsed earlier in the paper. But none of this changes what we have said about the role that the method of cases is meant to play in philosophical practice nor the evidential weight given to what philosophers think about those cases.

We have been suggesting so far that the claims of "intuition deniers" (a term coined by Nado 2016b) come to grief when they confront broader readings of the texts in which thought experiments appear. A further problem arises when we consider, not just our practices regarding thought experiments, themselves, but also our more general practices involving how we are taught both to construct arguments and to critically engage with them on the whole. Construing many of these thought experiments as arguments in the way the intuition deniers suggest flies in the face of our methodological norms of argumentation more generally. So, for example, norms of argumentation in analytic philosophy tend towards the *hyperarticulation of premises, presuppositions, and inferential structure*. And there are good reasons for these kinds of explicitness-encouraging norms. Well-articulated arguments are, well, articulated, and thus allow us to focus attention on each of their attendant parts. We can evaluate each premise, and the ones that seem at all dubious, can be scrutinized even more closely. And, when the premises are presented in such a way that the reasoning becomes as close to transparently valid (or at least cogent) as possible, then we can carefully evaluate the evidential relationship between premises and conclusion. (Notice, by way of underscoring all of this, that it is often considered a very substantial demerit when someone's argument turns out to be enthymemic.) The kinds of arguments that Deutsch and Cappelen have in mind, and which they think can be located in famous verdict-based philosophical arguments, fail to provide any of the methodological benefits that proper and explicit arguments do, and so are not worthy of any particular respect or deference. If philosophers are making arguments in their heads, but *don't actually give us their arguments*, then really all they're doing is asking us to take their word on it. That's strange enough. What's even stranger is that they'd be asking their philosophical opponents to just take their word on it, as well, which is the oddest kind of dialectical trick. Suppose instead that philosophers are not arguing, but are instead issuing case verdicts in accord with the consensus professional norms of the armchair method of cases, which include the presupposition that one's own intuitions will most likely be shared by one's interlocutors. In this case, the overall shape of their argumentative behavior makes much better sense. Of course, it is possible that when philosophers are making arguments in their heads, they have some reasonable expectation that their interlocutors will reconstruct those arguments in *their heads*, and so aren't asking anyone to take their word for it. The problem is that there are limits to any charitable reconstruction of someone else's argument, and so limits to how charitable we can reasonably expect other people to be when they engage with our work. And so a dangerous game is afoot when philosophers leave this much to their readers as our previous discussion about the different ways that philosophers have interpreted what's going on in Gettier's paper. There is a reason that we just don't do this. Anyone with any experience with referees #2 will know how unlikely it is that even their explicitly stated arguments will be uniformly construed charitably, let alone their unstated ones.

The way that Deutsch and Cappelen want us to think about philosophical thought experiments also fails to make sense of how we are properly taught to engage with one another's actual arguments. So, for example, the act of *explaining away* what we think about philosophical cases would constitute a dramatic violation of the principle of charity were we to think that this involves explaining away someone's argument. To see this, consider how John Hawthorne (2005) and Timothy Williamson (2005) attempt to explain away what we think about epistemic cases that involve the possibility that the protagonist is making some kind of mistake or another in terms of the influence that the *availability heuristic* has on how we think about philosophical cases (for critical discussion, see Nagel 2010). For right now, it is important to contrast how this kind of argumentative move seems perfectly good when it takes aim at philosophers' verdicts about philosophical *cases*, with how poorly received would be any version of this move that targets philosophical *arguments*, themselves.[12] When attributing a heuristic to someone's subpersonal cognition, the main constraint in practice is that it be at least somewhat independently motivated, though in practice this need not involve an appeal to the scientific literature. But when proposing a reconstruction of someone's argument, there are significant constraints of charity. "You have asserted that *p*, but I wonder if your argument for *p* is really this patently fallacious one that you would never reflectively endorse?" is *not* a move in good order in our argumentative practices.

We don't take ourselves to have responded here to the whole range of approaches taken by the intuition deniers (for review and discussion of these approaches, see Weinberg 2016, Nado 2016b, and Horvath 2022), or to have come close to a conclusive refutation of the "arguments-not-intuitions" approach. In fact, an important lesson that should be learned from thinking carefully about the ways that Deutsch and Cappelen defend the method of cases is that philosophers need to be considerably more straightforward than they have been about just what they are doing when they use thought experiments. But we do take ourselves to have shown that the "arguments-not-intuitions" approach has so far not attended adequately to key particulars of philosophical practice, not just in the way that verdicts about thought experiments get deployed, but also in how they get argued *about*, and more generally, with the norms for philosophical practice for argumentation itself.

2. Much to *whose* surprise?

As we noted above, the first attempt to respond to the experimental challenge means to do so by evading the challenge altogether. If intuitions do not matter much to analytic philosophers, then any concerns that experimental philosophers might raise about them don't speak to how philosophers use the method of cases. Joshua Knobe, in a set of recent and forthcoming papers, raises a different kind of worry, arguing that the experimental challenge rests on a different kind of mistake. According to him, the experimental challenge doesn't get philosophical practice wrong, it gets the empirical results wrong, or at least is not consistent with the growing body of empirical work that has emerged over the past decade or so, including work on what has, or has not, replicated. It turns out, according to Knobe, that philosophical case verdicts are "surprisingly stable" - that is, less heterogeneous than we

[12] We should distinguish our argument here from that of Climenhaga (2018). Climenhaga argues that the practice of requiring theorists to explain away problematic intuitions is *in and of itself* excellent evidence that intuitions are treated as evidence; and we applaud this argument. Our argument here, however, is about how the Deutsch/Cappelen approach cannot make sense of *the particular sorts of explanations* that typically get offered in this practice. Our thanks to an anonymous referee on this point.

might have first thought and less sensitive to irrelevant factors, as well. And so, the experimental challenge turns out to be no challenge at all.

While Knobe is attempting to cast doubt on the key empirical premise of the experimental challenge, it is important to be clear that he is not trying to endorse armchair philosophical methods, which since Knobe is a noted champion of experimental philosophy, would hardly be expected. He has his own metaphilosophical fish to fry, turning on an interesting distinction between *instability* and *tension* in our philosophical intuitions. Knobe thinks that there's very little of the former sort of thing, but lots of the latter, and he advances a specific research program to investigate these kinds of tensions. Our interest here, however, is just with his assessment of the kinds of stability or instability that are relevant to the experimental challenge. As we shall see, though, his characterization of how much stability there is or is not depends on what perspective one takes on the question.

We should start with a surprising feature of Knobe's papers, which is that their conclusions are framed explicitly in terms of whether the degree of stability found in people's philosophical intuitions is itself *surprising*. The adverb "surprisingly" is right there in the title of both papers, and he's very clear about it in his text, for example, when he writes (2021, 1),

> The evidence now suggests that philosophical intuitions are surprisingly stable. Indeed, the available evidence suggests that philosophical intuitions are surprisingly stable across both demographic groups and situations.

Or, when he writes (2019, 33),

> I have been suggesting that one surprising finding coming out of the experimental philosophy literature is the shocking degree to which demographic factors do not impact people's philosophical intuitions.

Surprising *and* shocking!

In one sense, just about the least surprising thing that we can see in philosophy is the word "surprising" in a paper about experimental philosophy.[13] After all, experimental philosophers *love* to characterize their results in such terms, and for the very good reason that their results often are indeed not ones that we would have antecedently predicted. And after all, much of the point of experimental philosophy is that it can reveal facts about philosophical cognition that are not available from the armchair.[14] But standardly in such papers, we can parcel out the experimental results, themselves, from the further gloss or spin concerning the surprisingness of those results. The study as conducted, the data that is gathered, the statistical inferences from that data - when it goes well, all of that can maybe tell you something about what you were studying, but none of it can tell you whether it adds up to anything surprising. And by and large, in our experience, referees don't seem to worry too much about the surprisingness bit, since in these papers it's mostly just rhetoric.

[13] Just by way of illustration, Google Scholar shows 1540 hits for "experimental philosophy" in articles that were published in 2021, and just over half (777 of them) of them include "surprising".

[14] This idea received some pushback in Dunaway *et al.* (2013), but see Liao's (2016) critical response.

In Knobe's recent arguments, however, it can't be just rhetoric. There is literally no claim to be defended here without *something* setting a threshold for the observed degree of stability being claimed. There's no question, after all, that there's some amount of demographic variation greater than zero and less than absolute. For that matter, we expect everyone in this debate would agree that there's substantially more non-variation than variation, that is, that *most* philosophical intuitions will be invariant to *most* dimensions of possible demographic variation. And so, Knobe's "surprising" modifier is a crucial part of creating a thesis that would be, even in principle, empirically evaluable. Or consider this claim in his (2019): "At an empirical level, the key question is how to explain the surprising robustness of philosophical intuitions. One possible answer would be that the capacities underlying people's philosophical intuitions have an innate basis." (33) Knobe's claim is only intelligible with some sort of characterization of *just how much* robustness there is, a function played here by that "surprising". That is, without the "surprising" or some other threshold-setting locution to characterize just how much robustness has been observed, there just wouldn't be an *explanandum* there for an innateness hypothesis to serve as a potential *explanans*.

But to define a thesis in terms of *p*'s being "surprising" presupposes, at a minimum, some sort of priors regarding *p*. The thesis would then be that the empirical evidence mandates some heavy-duty updating from those priors. For Knobe's purposes, where might these priors come from? Once we recognize that there are several distinct but equally legitimate ways of filling that in, it becomes clear that there are a number of claims that Knobe could be putting into play here - claims that might not all agree in truth-value. Our main contention here will be that, while his claims are at the very least defensible and debatable for some candidate priors, for the source of priors that would be most relevant to debates about methodology, Knobe is badly off-target regarding that state of play in the empirical literature.

The first, and most literal, place to look for priors, in order to determine whether or not philosophical case verdicts are surprisingly stable, would be at what priors experimental philosophers give to the stability of philosophical case verdicts. Now, we'll just speak for ourselves because it's hard to speak for everyone. And, indeed, the fact that it obviously makes no sense to speak for everyone here should raise some preliminary worries that this may not be a great way to proceed. (Maybe we need to do some X-phi on the X-phi experts?) Anyhow, just reporting our own priors, while some of the specific experimental results that have been published in recent years have been surprising, we would say that overall the *trend* is not one that we think falls at all outside our expected range, for two main reasons. First, our view before learning about these experimental results was that philosophers, including experimental ones, didn't know hardly anything at all about just *where* and *how much* of *what sorts* of variations might be found. So the range of plausible distributions of variation was antecedently very open. We take it that what the early experimental philosophy studies showed was not, and was never meant to be, "*look, we've shown that there's rampant variation all over the place, afflicting everyone everywhere all at once*" but something much more like, "*look, we've shown that no one really has much of a clue yet just how much variation there is or isn't out there.*" The early studies revealed our fairly drastic ignorance about intuitional variation by revealing several previously-uninvestigated hypotheses to be live empirical possibilities.

Knobe seems to be coming from a different starting point here, for example, when he writes that because experimental philosophers were studying "intuitions about seemingly abstruse issues, such as the nature of the true self or whether the universe is governed by deterministic laws. There was every reason to expect that such intuitions would differ radically between demographic groups." (2019, 31) We're not sure that there's any useful way to argue about conflicting priors, and so we'll just underscore how far apart our priors are here. It seems to us that there was every reason to think that philosophical intuitions about cases involving these philosophical issues might *differ occasionally, perhaps frequently and perhaps systematically, and in unexpected ways* between demographic groups. And, we take this to be very much weaker than what Knobe reports as his own previous view.

Our second reason to take ourselves to be unsurprised here is that we just don't think that the experimental studies that Knobe musters in his two papers go very far towards removing that ignorance. Here's how Knobe characterizes his argument in the 2019 paper:

> I have been suggesting that one surprising finding coming out of the experimental philosophy literature is the shocking degree to which demographic factors do not impact people's philosophical intuitions. In support of this claim, I have cited 30 studies, by 91 different researchers, comprising a total sample size of 12,696 participants. Many of these results would be highly surprising even in isolation. Taken together, they are downright shocking.

With all due respect to Knobe, we invite the reader to pause for a moment to consider just how huge the Cartesian product is when looking at the space of possible philosophical cases and the different ways in which philosophical case verdicts about these cases might vary. We think it will then seem fairly obvious that 30 studies is just not very many at all, especially when it's not a random sample, and he himself acknowledges that there are other studies that do find demographic variation.[15] People could only be shocked if they expected gobsmackingly wanton demographic variation every which way but loose, but we must confess that that strikes us as an implausible take on the initial experimental philosophy results, even from the very heady early days where effects seemed fairly easy to find.

In short: we already didn't have any very specific expectations, and we don't take ourselves to have received much evidence at all to disconfirm what expectations we did have, in their louche vagueness. It seems to us that the burgeoning body of work has trimmed off the more extreme ends of the distribution of possibilities here. On the one hand, it does not seem likely at this time that, say, East Asian and Western communities have differences in their knowledge attribution so stark as to motivate positing a substantial form of epistemic relativism. That seemed a real possibility in the wake of the original Weinberg *et al.* (2001) study; it no longer seems especially plausible now, especially given the repeated failed attempts at replication (e.g., Kim & Yuan 2015, Seyedsayamdost 2015, Machery *et al.* 2017). But on the other hand, when there were just a handful of results in, it was just as possible that *none* of them could have held up under further scrutiny *and* that no further interesting variation results would surface. Despite these notable replication failures of some of those early findings, so many new

[15] Stich & Machery (2022) argue that Knobe also *misreads* the evidence. Knobe (forthcoming) offers a reply. We will not pursue this line of debate here ourselves, since our arguments are intended to hold even if Stich and Machery are incorrect in this matter. Of course, if they are in fact right, then so much the worse for Knobe's case against the empirical premise.

results have emerged and continue to emerge that it is no longer a live possibility that there are just no meaningful variations here to be concerned with. (See, for discussion, Machery 2017.)

Another place to try to source the priors for a surprisingness claim is from *the state of the scientific literature*. This seems a better claim for Knobe, in terms of getting a claim that may well be true. Perhaps we could construct a history of experimental philosophy and the psychology it tends to attend to, starting in something like the 1990s in big cultural group differences, with work by folks like Richard Nisbett, Kaipeng Peng, or Jonathan Haidt, and then later the research by Joseph Henrich and others on the pervasive differences between Western cognition and that of the rest of the world, and also the explosion of social priming results from John Bargh, Simone Schnall, and others, and the way that the heady first few years of experimental philosophy seemed to be riding that wave - but then observe, quite rightly, that much of the subsequent work on these issues has ranged from the deflationary to the outright debunking of much of the earlier work. It seems to us legitimate to claim that there has been a swing in the big psychological sciences pendulum here, and that the smaller trajectory of experimental philosophy has resonated with it. In that sense, then, the shifting of momentum from reporting all sorts of variation to reporting all sorts of failed replications of variation is certainly noteworthy, and may also legitimately count as "surprising."

Not only is the "direction of the literature" construal of "surprising" a good bet for making Knobe's claim come out true, we would note that questions about the large-scale scientific picture of the nature of the human mind are a perennial interest for Knobe, as evidenced in papers like his 2010 BBS target article. So we think that this way of setting the threshold in Knobe's conclusion is a charitable one, having him making a defensible claim of legitimate relevance to varieties of scientific debates that Knobe has been interested in. We do not take any stance here as to whether that threshold is or isn't met.[16] Our main contention here will be, rather, that questions about the direction of the scientific literature will not serve to set the "surprise" bar *at a spot that will be of much relevance to debates about philosophical methodology.*

To consider how much stability should count as surprising for the purposes of debating philosophical methodology, we should extract the relevant priors from the methodological practices themselves. What sorts of variation, and to what extent, are our methods anticipating? Or rather, to avoid any problematic anthropomorphism, how much variation of what kinds are our practices well-configured to detect, and have been appropriately buttressed to handle? It seems to us the answer is: *approximately none whatsoever!* One quick way to see this is to consider what our philosophical practices would look like if we were at least *trying* to handle demographic and situational variation from the armchair, and to see how there is nothing of the sort in our current practices with the method of cases.[17] For starters,

[16] Though, we suppose that given our earlier point that we don't take the total state of evidence to add up to so very much, probably we would opt for "isn't", if we had to bet. We'd probably opt more for "suggestive" than "surprising". But that plays no role in our arguments here.

[17] To be absolutely clear, we don't mean to be saying that philosophers cannot learn anything whatsoever about the method of cases from the armchair. Surely, that would be incorrect. There has been significant discussion about the shape and structure of the method of cases. The problem is that the philosophical folk wisdom that

if our philosophical practices were anticipating any nontrivial amount of demographic or situational variation, then we would at a minimum see some attention and discussion of such variation, especially in handbooks and textbooks, with debates in the literature about the nature of these variations and how to handle them.[18] For demographic variation, for example, we might expect to find a norm of disclosure about every author's location among various parameters: ethnicity, gender identity, native language, religious upbringing, and so on. That way, we could increase our chances of spotting such variation where it might potentially arise. And, for our methodological practices to monitor for and, ideally, control for situation effects, we would perhaps expect to see the development of specifications for some sort of canonical conditions for using the method of cases. Manuals for philosophical methodology might specify: sit in a quiet room, clear your mind with at least 3 minutes of meditation, then look at the scenario, as printed with black ink on white paper in a clear, sans-serif font. (Maybe Arial rather than Garamond would become *the* font of philosophical thought experiments?) What you would *not* see in philosophical practice are hypothetical cases just dropped into papers wherever it nicely suits the flow of the argument. We leave it to the reader to speculate on what sorts of rules could be adopted for the consideration of cases on the fly, for example, in a colloquium Q&A, in order to screen off such effects, and to observe that, of course, no such rules are even remotely in force, or even under consideration.

We are being fanciful here, but to a purpose: to highlight just how very little our philosophical practices with the method of cases are prepared for really any instability at all. Contrast this with our extensive resources, and well-worked-out norms for their use, for handling sources of error that we *do* expect. We have our norms of argument articulation, as discussed above, in no small part to help us avoid mistaken impressions of validity. Where needed, we use formal machinery such as parentheses, operators, and different forms of quotation, in order to avoid scope ambiguities, use-mention errors, and the like. When matters get particularly fraught and complicated, we can even translate the whole mess into logic and run derivations or provide models.

Indeed, it's worth noting that analytic philosophical practice is not generally cavalier about errors. We don't tend to operate like big data miners, who know that tons of the specific observations in their data set will be mistaken, perhaps many of them quite substantially so, and who are thus counting on their statistical methods to help them filter the signal from the noise. Very much in contrast, our inferential and dialectical practices tend to hold our theories to a very demanding standard. Remember

emerges from this discussion doesn't prepare us for the many different kinds of errors that cannot be detected using only the resources that are available to philosophers from their armchairs.

[18] An anonymous referee points out to us, as indeed many philosophers have over the years, that analytic philosophers are not unaware of some version of these phenomena, perhaps most famously in Williams (1970), or more recently in Gendler and Hawthorne (2005). Our point is that whatever scraps of awareness philosophers have managed to acquire over the years about possible demographic or framing problems with intuitions, it has unfortunately not yielded really any changes whatsoever at the level of actually practiced methodology of a sort that can help us handle such influences. And as we have been stressing here, the shared disciplinary practices, not the minds of individual philosophers, are the proper target of our methodological critiques.

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):[19]

> 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. (p. 1)

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. But to the extent that 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*. 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 the method of cases. It seems that even just a handful of bad intuitive verdicts could spoil an entire line of inquiry.

There is, thus, a clear sense in which we can say that analytic philosophical practice anticipates certain sorts of threats of error, whereas for other threats, such as the potential threats of demographic variation or instability, it is not expecting any such threats in any meaningful way. This fact about philosophical practice is absolutely essential to how we think that philosophers should think about the recent debate about philosophical methodology that has been prompted by the experimental challenge to the method of cases, and to the way that we think philosophers should think about this debate and the relationship between analytic philosophy and experimental philosophy, more generally. From the point of view of our actual philosophical practices, even a very modest amount of demographic and instability would count as a surprising, or even shocking, level. Most importantly, so far as that first premise in the experimental challenge is concerned, there is evidence of plenty enough variation to catch philosophers napping in their armchairs and to reveal just how little we know at this time about where else further variation might yet be found.

Our diagnosis here is that Knobe tried to take a question that can only be answered by attention to the particulars of philosophical practice, and construe it purely as a question about psychology. For we can observe a crucial shifting on Knobe's part in the vocabulary of these discussions, from an ineliminably philosophical one to a straightforwardly psychological one. Those who have pursued the experimental challenge have tended to define *instability* in terms like those that we see in Joachim Horvath's (2010) characterization of that challenge, what he calls its "master argument," namely, variation with "irrelevant factors." This exact phrasing can be found, for example, in the Swain *et al.* (2008) paper with "instability" in its title, where the authors characterize instability in precisely those

[19] See also Nado (2015) on the *epistemic demandingness* of analytic philosophy.

kinds of terms. And, Nado (2015) leads off her critical overview of the experimental challenge with a characterization in highly similar terms:

> Premise: work in experimental philosophy indicates that intuitions vary as a function of such philosophically irrelevant features as order of presentation and cultural background. Conclusion: intuitions are unsuited for their current evidential role in philosophical argumentation. We might call this the 'variation argument' against intuition. (204)

It would be easy to multiply examples. One key feature of this standard construal of instability or variation is that it is *ineliminably metaphilosophical.* You cannot specify the relevant sort of variation without some idea of what sorts of factors or features are relevant to what sorts of philosophical propositions.

Yet, for Knobe, what it means for some philosophical case verdict to demonstrate instability is, simply, for it to be manipulable in ways that do not involve changing the substantive content of the scenarios. Instability is what you get when you only vary either features of the participants themselves (as with demographic variation) or in "studies on the influence of external situational factors. In such studies, researchers give all participants exactly the same case and exactly the same question, but they manipulate some factor in the external situation" (Knobe 2021, 48). Nothing about philosophical truth appears in that construal. It is entirely about what sort of experimental technique, aimed at investigating what sort of variable, is deployed in a given experimental study. To be clear, this is an obviously legitimate notion of instability, and may be just what is needed for investigating certain sorts of large scale questions that are simultaneously scientific and philosophical about the nature of human cognition. It's just not the notion of instability in play in the arguments that Knobe has tried to declare "moot."

And once we take on the perspective of instability *a la* the experimental challenge, how various empirical findings get scored can change rather substantially. For Knobe's version of instability, content effects that are found to be cross-culturally robust get scored as instances of stability. But for the kind of instability that is central to the experimental challenge, any content effects *which arguably are philosophically irrelevant* will get scored as instances of *in*stability, and so much the worse if they are found across the board. There are several candidate effects of this sort listed in Knobe's paper, such as the influence of moral content on Gettier case knowledge attributions and the impact of physical contact in scenarios about moral dilemmas. We would expect *framing effects,* in general, to be examples of this phenomenon, as they will depend on some differences in the language or other framing elements in the contrasted scenarios, yet most typically those differences would not be ones that would be considered philosophically relevant.

Taking the point further, some demographic differences that might not seem very important for Knobe's purposes may be highly charged when we are considering their implications for philosophical practice. To take just one example, the effects of personality type that have been investigated by Adam Feltz and Edward Cokeley (2009, 2019) don't rate a mention in Knobe's 2021 paper, and while he acknowledges them in his (forthcoming), he does so primarily to downplay their significance: "...it is not as though the finding is that there is some pervasive phenomenon whereby all sorts of different philosophical case verdicts are correlated with all sorts of different individual difference measures" (42). Again, it sounds like rather wild and rampant variation is what would be needed to disconfirm intuitional stability *sensu* Knobe. Yet as we noted earlier, we don't think anyone who advanced the experimental challenge ever thought that *that* much variation seemed at all likely. It's enough for the Feltz and Cokely-type effects to be one more spear in the experimental arsenal.

The parameter of demographic variation that produces the starkest mismatch between Knobe and the experimental challengers is that of *philosopher vs. nonphilosopher*. A great many of the findings that are stable across the folk, and scored by Knobe as instances of stability to confirm his thesis, are ones where the folk's consensus verdict diverges from the received verdict in the philosophical literature. For example, Knobe's list of cross-culturally robust findings includes the irrelevance of stakes to knowledge attributions, which, if true, would be a challenging finding for much of analytic epistemology of the last few decades. Philosophers' case verdicts about Gettier cases also seem to diverge in substantial and interesting ways not just from "the folk" but also from our colleagues in basically every other academic discipline (Starmans and Friedman 2020). Furthermore, a significant trend in the X-phi literature that Knobe is commenting on is that philosophers' funny thought experiments *are often kind of meh as instances or non-instances of their target concepts*. Lots of the "stability" cases are like ones where the experimental participants think that, say, Truetemp is a "kinda-sorta but maybe a bit more kinda-sorta-not a case of knowledge" - that is, an *inconclusive* case and thus hardly the kind of result that can be of comfort to armchair practitioners.[20]

From the point of view of the philosophy of human nature, philosophers represent an incredibly small, weird (and highly WEIRD) sub-sub-sample - there'd be no reason to modify anyone's views about cognitive universals based just on what esoteric whackos like us think! But, obviously, from the perspective of philosophical methodology, any "philosopher vs. non-philosopher" results will be highly fraught. We do not think that philosophers should be considered automatically wrong in any such disagreement, and we are sure there will be cases where there is good reason to let our trained philosophical judgment trump the *hoi polloi*. But our point here is just that instances of this particular category of demographic variation are automatically highly salient to the experimental challenge, even if they may be right ignored by someone with Knobe's theoretical interests.

We'll conclude this section with one last extended example. Knobe considers some very well-investigated order effects on trolley cases, particularly regarding the pattern of influences between the

[20] Although, as noted above, we don't intend to engage with Knobe's metaphilosophical project about locating tensions in the stable differences to be found in the experimental philosophy results, we will offer this one observation: we have to be careful in interpreting inconclusive results like a 54%/46% split about P/not-P. It might be, as Knobe takes it, a manifestation of two strong and warring philosophical impulses about P, in tension with each other. But it also might be a manifestation of the folk just not having any strong views about P at all! And these do not exhaust the possibilities; for example, the difference could be the result of some philosophically shallow context-sensitive process. This cannot be read trivially off the distributions or histograms themselves, though we expect that strongly bimodal distributions are more likely to be indicative of tensions, whereas a big shmear centered on the Likert scale midpoint is more likely to mean that the folk just don't much know or care about P.

sidetrack case (aka “switch”; diverting the trolley so it veers onto another track, which will then kill one person) and the footbridge case (aka “push”; shoving the man onto the tracks, killing him but thereby stopping the trolley). In general, it appears that when someone sees the sidetrack case first, they will be much more inclined to find pulling the switch acceptable than they would when they see the scenario after first seeing the footbridge case. Interestingly, there does not seem to be any reverse effect: people just don’t like killing the one to save the five in the footbridge case, regardless of whether they first see the sidetrack case. Wiegmann & Waldmann (2014) systematized and extended that literature, and proposed a model for these order effects, which we are here simplifying greatly: the sidetrack case presents a *cognitive ambiguity* that the footbridge does not. One can view the sidetrack case in a way that really foregrounds the saving of the five, and then parcels out the downstream killing of the one as a separable event, a kind of coda. Or one can view it all as one big event, with the killing included as a constituent element. Depending on which way one parses the events, and in particular whether the killing is construed as part of the action or distinct from it, one may feel differently inclined towards the action. In contrast, the footbridge case doesn’t support such an ambiguity, because no matter how you parse it, you’re going to have to include that killing. Thus, they hypothesize, seeing the footbridge case first leads people to be more likely to parse the sidetrack case in a manner that includes the killing, and thus to judge it as less acceptable. But since there is no corresponding ambiguity in the footbridge case, it won’t experience any such order effect. And indeed, they predict that any other such trolley scenario without such a potential ambiguity of construal will similarly display no order effects, and they confirm that prediction. Knobe sums up this research as follows:

> In the specific case in which participants receive the footbridge dilemma and then receive the sidetrack dilemma, we have strong evidence for an impact of external situational factors: thinking about the first dilemma really does influence judgments about the second. What is this fact teaching us? The obvious first guess would be that it is evidence of a process that leads to some broader form of instability. Perhaps the process affects people's philosophical intuitions more generally, or perhaps just their moral intuitions, or at a very minimum, it surely affects people's intuitions about a wide variety of different trolley problems. The surprising finding coming out of research in this area is that this is not what is happening. The effect observed in this one case seems to be highly circumscribed. Not only does it not emerge in cases that are radically different, it doesn't even emerge in cases that might at first seem extraordinarily similar to the footbridge-sidetrack sequence. (2021, 66)

From the point of view of doing the scientific philosophy of human nature, that may well be right: from that perspective, and for those research interests, it might be appropriate to consider this ambiguity-based mechanism for order effects to be “highly circumscribed,” and of limited further interest. But from the point of view of the armchair method of cases, these effects should be both surprising and *very* disquieting. First, and this really needs to be underscored, even if it only afflicts a small number of cases, these cases are absolutely crucial ones in the normative ethics literature! It would already be problematic if the effects were just found on some of the odder cases, such as the one where the track loops back around on itself (Liao *et al.* 2012). But note that the main effect of interest in Wiegmann & Waldmann (2014) is on the classic sidetrack version of the case, and that case figures prominently in a great many of the arguments in this literature. So much of the trolley-based argumentation relies on making comparisons between the different cases, and it is problematic if we cannot, with the experimental results in, determine what to count as "the" verdict for the sidetrack case. Perhaps from Knobe's particular psychological perspective, oh, it's just this one narrow set of cases, but from the point of view of an ethicist looking to work in this topic, their whole literature will have gone off the rails.

Moreover, Knobe is somewhat underplaying how widely Wiegmann and Waldmann's mechanism for order effects may be expected to generalize outside of a subset of trolley cases. The key idea is that when cases have a particular kind of complex multi-part structure, they may enable different construals with accordingly different emphases. Now, philosophical thought experiments being the wild and chimeric creatures that they are, we ought to expect that other scenarios would manifest this sort of complex, multi-part construction. Indeed, the subclass of Gettier cases that involve switched truthmakers seem like plausible candidates for exploration here, and indeed there is some evidence of order effects on such cases (Machery *et al.* 2018).[21] We wouldn't be (ahem) surprised if other yet-unexamined cases did as well, whenever there's a "one hand giveth and one hand taketh away" sort of structure. Consider Alvin Plantinga's (1993) oft-used case where someone gets a lesion that causes their beliefs to be generally unreliable but *also* produces, using the same mechanism that provides lots of bad beliefs, the belief that they have a lesion. Such cases have a kind of multi-part diverging causal structure that the Wiegmann and Waldmann mechanism for order effects may well be able to get a hold of. (On reflection, it seems to us that a great many tricky reliabilism cases could potentially involve construal issues of exactly the sort highlighted by the generality problem.) Whether or not Wiegmann and Waldmann's mechanism turns out to be "highly constrained" in the specific domain of trolley problems, there is nothing in their specification of the mechanism that would at all indicate it would not be found in many other sorts of cases in various other philosophical domains. The Machery *et al.* (2018) results seem problematic for Knobe's argument either way: either they demonstrate how the Wiegmann and Waldmann mechanism can extend in ways far outside trolley cases, or they demonstrate that there are yet other hitherto-undiscovered mechanisms in our cognition, for other sorts of order effects. This would itself not be surprising, since we already know that there are other kinds of order effects out there, such as recency/primacy effects in causal judgment (Henne *et al.* 2021).[22] And since causal judgments may play a role in many other sorts of philosophical domains

[21] Note that the specific causal framework Wiegmann and Waldmann use for the trolley cases will not carry over to epistemology cases, typically, since ethical dilemmas tend to be about downstream consequences, whereas knowledge attributions tend to be about upstream sources. One way in which the Wiegmann and Waldmann mechanism *might* apply to the switched-truthmaker case in the Machery *et al.* (2018), involves an ambiguity of construal that focuses on the current state of the believer, in which they seem to check all the JTB boxes; and one that reaches further back, and registers the disconnect between the source of the justification, and the distinct object that rendered the belief true.

[22] We should note that this is a different kind of order effect: a content effect of the temporal order of events in the scenario, and not a context effect of the order of presentation. It is nevertheless grist for our mill since the effect is a philosophically irrelevant one, which moreover is susceptible to manipulations that get subjects

(such as attributions of agency), there's every reason to think these could further ramify at least somewhat.

The point of these last empirical speculations is that it's entirely consistent with order effects being basically nugatory from Knobe's perspective of human intuitive cognition on the whole, that they also afflict enough of the cases deployed in the method of cases to raise some very serious methodological headaches for philosophers. The problem cases don't need to comprise anywhere near a vast majority, or even a significant plurality, of the cases that philosophers use, precisely because philosophical practice is so vulnerable to them -- so easily surprised by them, we might say. And the set of such problem cases is much bigger than Knobe estimates, even stipulating that he is right in his overall read of the literature, because the problem cases are delineated in terms of sensitivity to philosophically irrelevant factors, and not just the much narrower (though still decidedly non-null!) set of cases displaying purely situational effects. While we (obviously) think that psychological results are highly important to philosophical methodology, it does not follow that distinctions drawn purely in terms of psychological methodology can do the metaphilosophical work that is needed.

3. Conclusion

We have tried to show that two recent attempts to respond to the experimental challenge to the method of cases from the perspective of these different philosophical practices fail precisely because they are not sufficiently attentive to the particulars of those practices. But the moral of the story is broader than just this. The moral of the story is methodological, or perhaps even *meta-methodological*. It is that philosophers cannot continue to argue about philosophical methodology without really looking at philosophical methodology, as it is in fact practiced. It is not enough to just squint at the putative argument indicators in a handful of texts or defer to distinctions and categories that are well-founded in psychology. Having said this, we want to be careful not to over-represent or over-sell ourselves as observers of philosophical practice. Our claims should be taken as highly empirical ones, for whose defense we have, let's face it, mostly been offering armchair observations. Perhaps we count as "participant observers," but that is hardly an innocent epistemological position. We, thus, conclude by welcoming increased and improved empirical, and indeed experimental, investigation of the particular twists, turns, and contours of actual, in-the-trenches philosophical practices, as an essential component of further methodological debates.

to simulate some rather than other situations in counterfactual reasoning. Moreover the authors urge that their work is relevant to a wide range of philosophical issues, including moral judgment and experimental jurisprudence.

Works Cited

Alexander, J. (2012). *Experimental Philosophy: An Introduction*. Cambridge: Polity.

Baz, A. (2017). *The Crisis of Method in Analytic Philosophy*. Oxford: Oxford University Press.

Bengson, J. (2014). How philosophers use intuition and 'intuition'. *Philosophical Studies*, *171*, 555-76.

BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.

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

Chisholm, R. (1989). *Theory of Knowledge*, 3rd Edition. Englewood Cliffs, NJ: Prentice-Hall.

Colaço, D. and Machery, E. (2017). The intuitive is a red herring. *Inquiry*, *60(4)*, 403-419.

Deutsch, M. (2010). Intuitions, counterexamples, and experimental philosophy. *Review of Philosophy and Psychology*, *1*, 447-260.

Deutsch, M. (2015). *The Myth of the Intuitive: Experimental Philosophy and Philosophical Method*. Cambridge, MA: M.I.T. Press.

Deutsch, M. (2016). Gettier's method. In J. Nado (ed.), *Advances in Experimental Philosophy and Philosophical Methodology* (69-98). New York: Bloomsbury Academic.

Dunaway, B., Edmonds, A., and Manley, D. (2013). The folk probably do think what you think they think. *Australasian Journal of Philosophy*, *91(3)*, 421 - 441.

Engel, M. (2022). Evidence, epistemic luck, reliability, and knowledge. *Acta Analytica*, *37(1)*, 33-56.

Feltz, A. and Cokeley, 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(1)*, 342-350.

Feltz, A. and Cokeley, E., (2019). Extraversion and compatibilist intuitions: a ten-year retrospective and meta-analyses. *Philosophical Psychology*, *32(3)*, 388-403.

Fodor, J. (1997). Special sciences: Still autonomous after all these years. *Philosophical Perspectives*, *11*, 149-163.

Gendler, T. and Hawthorne, J. (2005). The real guide to fake barns: A catalogue of gifts for your epistemic enemies. *Philosophical Studies*, *124(3)*, 331-352.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23(6)*, 121-123.

Goldman, A. (1967). A causal theory of knowing. *Journal of Philosophy*, *64(12)*, 357-372.

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

Goldman, A. (2017). Gettier and the epistemic appraisal of philosophical intuition. In R. Borges, C.de Almedia, and P. Klein, (eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (213-230). Oxford: Oxford University Press.

Goldman, A. and Pust, J. (1998). Philosophical theory and intuitional evidence. In M. DePaul and W. Ramsey, (eds.), *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry* (179-200). Lanham, MA: Rowman and Littlefield.

Hawthorne, J. (2005). *Knowledge and Lotteries*. Oxford: Oxford University Press.

Henne, P. Kulesza, A., Perez, K., and Houcek, A., (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, https://doi.org/10.1016/j.cognition.2021.104708

Horvath, J. (2010). How (not) to react to experimental philosophy. *Philosophical Psychology*, *23(4)*, 447-480.

Horvath, J. (2022). Mischaracterization reconsidered. *Inquiry*, https://doi.org/10.1080/0020174X.2021.2019894

Kim, M., and Yuan, Y. (2015). No cross-cultural differences in the Gettier car case intuition: A replication study of Weinberg *et al.* 2001. *Episteme*, *12(3)*, 355-361,

Knobe, J. (2019). Philosophical intuitions are surprisingly robust across demographic differences. *Epistemology and Philosophy of Science*, *56*, 29-36.

Knobe, J. (2021). Philosophical intuitions are surprisingly stable across both demographic groups and situations. *Filozofia Nauki*, *29(2)*, 11-76.

Knobe, J. (forthcoming). Differences and robustness in the patterns of philosophical intuition across demographic groups. *Review of Philosophy and Psychology*.

Lehrer, K. (1990). *Theory of Knowledge*. Oxford: Routledge.

Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, *74(4)*, 549-567.

Liao, S. (2016). Are philosophers good intuition predictors? *Philosophical Psychology*, *29(7)*, 1004-1014.

Liao, S.M., Weigmann, A., Alexander, J., Vong, G., (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology*, *25(5)*, 661-671.

Machery, E. (2017). *Philosophy Within its Proper Bounds*. Oxford: Oxford University Press.

Machery, E., Stich, S. Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., and Hashimoto, T., (2017). Gettier across cultures. *Noûs*, *51*, 645-664.

Machery, E., Stich, S. Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., Sirker, S., Usui, N., and Hashimoto, T., (2018). Gettier was framed! In M. Mizumoto, S. Stich, and E. McCready, (eds.), *Epistemology for the Rest of the World* (123-148). Oxford: Oxford University Press.

Mallon, R. (2017). Intuitive diversity and disagreement. In J. Nado (ed.), *Advances in Experimental Philosophy and Philosophical Methodology* (99-124). New York: Bloomsbury Academic.

Nado, J. (2015). Intuition, philosophical theorizing, and the threat of skepticism. In E. Fischer and J. Collins, (eds.), *Experimental Philosophy, Rationalism, and Naturalism: Rethinking Philosophical Method* (204-221). New York: Routledge.

Nado, J. (2016a). Experimental philosophy 2.0. *Thought*, *5(3)*, 159-168.

Nado, J. (2016b). The intuition deniers. *Philosophical Studies*, *173(3)*, 781-800.

Nagel, J. (2010). Knowledge ascriptions and the psychological consequences of thinking about error. *The Philosophical Quarterly*, *60 (239)*, 286-306.

Plantinga, A. (1993). *Warrant and Proper Function*. Oxford: Oxford University Press.

Seyedsayamdost, H. (2015). On gender and philosophical intuition: Failure of replication and other negative results. *Philosophical Psychology, 28*(5), 642–673.

Starmans, C. and Friedman, O. (2020). Expert or esoteric? Philosophers attribute knowledge differently than all other academics. *Cognitive Science*, DOI: 10.1111/cogs.12850

Stich, S. and Machery, E. (2022). Demographic differences in philosophical intuition: A reply to Joshua Knobe. *Review of Philosophy and Psychology*

Swain, S., Alexander, J., and Weinberg, J. (2008). The instability of philosophical intuitions: Running hot and cold on Truetemp. *Philosophy and Phenomenological Research*, *76(1)*, 138-155.

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

Weinberg, J. (2014). Cappelen between a rock and a hard place. *Philosophical Studies*, *171(3)*, 545-553.

Weinberg, J. (2016). Intuitions. In H. Cappelen, T. Gendler, and J. Hawthorne (eds.) *The Oxford Handbook of Philosophical Methodology* (287-308). Oxford: Oxford University Press.

Wiegmann, A. and Waldmann, M. (2014) Transfer effects between moral dilemmas: A causal model theory. *Cognition, 131(1)*, 28-43.

Williams, B. (1970). The self and the future. *The Philosophical Review*, *79(2)*, 161-180.

Williamson, T. (2005). Contextualism, subject-sensitive invariantism and knowledge of knowledge. *The Philosophical Quarterly*, *55(219)*, 213-235.