The Calibration Challenge to Philosophical Intuitions

Paul O. Irikefe (UC Irvine)[1]

Abstract

To several critics of the philosophical method of cases—Robert Cummins, Jonathan Weinberg and his colleagues, and Avner Baz—the fact that philosophical intuitions cannot be calibrated means that we cannot rule out the skeptical hypothesis that the outcome of our theorizing based on these intuitions is deeply distorted by our cognitive artifacts. Moreover, they take this hypothesis to license the negative conclusion that we are unable to have much of the armchair knowledge we typically attribute to ourselves when philosophizing based on appeal to philosophical intuitions. This paper addresses this skeptical argument and shows how this negative conclusion can be resisted.

Keywords: Calibration. Method of Cases. Intuition. Thought Experiments. Ordinary Language Philosophy.

1. Introduction

The philosophical method of cases constitutes an important aspect of analytic philosophy. In this practice, in the course of inquiring into some subject matter of philosophical interest, such as, knowledge, belief, free will, promising, justice, reference, obligation, responsibility, and so on, a given philosopher presents a description of an actual or hypothetical scenario. And then the philosopher uses the intuition (or the content of this intuition) that some facts hold or fail to hold in the described

---

[1]**Secondary affiliation:** Research Associate, African Centre for Epistemology and Philosophy of Science, University of Johannesburg, South Africa.

scenario as negative or positive evidence for some particular claims or theory about knowledge, belief, free will, justice, reference, and so on.

Roughly two and half decades ago, Robert Cummins drew attention to a peculiar problem with this practice. Unlike observational sources in science, which supply knowledge or evidence in scientific inquiry, philosophical intuitions lack one epistemically useful feature: calibration. That is, it is impossible to calibrate philosophical intuitions, or so Cummins (1998) argues. This is because all the normal strategies for doing so are unhelpful. And the only option that might work, relying on a well-worked-out theory of the domain of philosophical intuitions would paradoxically make relying on philosophical intuitions useless or pointless since it would enable us to answer all questions about the target domain. Philosophical intuitions thus function like a pseudo-scientific instrument or even worse. Indeed, he thinks that "the most plausible account of the origins of philosophical intuitions is that they derive from tacit theories that are very likely to be inaccurate" (Cummins, 1998, p. 125). And he says that these putative origins explain why philosophical intuitions can always be explained away. One can always assume with great plausibility that an opponent's intuitions are products of her bad implicit theory. And that opponent has no way of ruling out that possibility precisely because there is no way of indicating that philosophical intuitions reliably track what is being investigated as we would if we were relying on a "genuine observational source"(Cummins, 1998, p. 117). Call this the *calibration challenge to philosophical intuitions*.

The calibration challenge has been revived by several theorists in the recent metaphilosophical debate. For example, Baz (2023), a leading theorist of ordinary language philosophy argues that some recent defenders of the practice of appealing to cases in analytic philosophy such as Cappelen (2012) and Deutsch (2015) fail to reckon with the calibration challenge posed by Robert Cummins. On his view, they do this because they hold that philosophical intuitions are useless apart from the support of arguments, which on their own rendering often take the form of the very theory that is supposed

to be under assessment. And by so doing, Baz claims, they have invited upon themselves the worry about theory contamination, which Cummins raised in connection with the calibration challenge. More importantly, Baz (2023) uses the insight of ordinary language philosophy to explain how the answers we give in our normal and everyday employment of philosophically interesting terms are "calibratable" whereas the answers we give when we are asked whether those terms apply to cases are not. Further, he argues that they are not precisely because the philosophical context lacks some vital social constraints present in our ordinary linguistic contexts. As another example of the lack of these social constraints, he references the practice of contemporary English-speaking moral philosophers who apply words such as "ought" or "must" to cases. And this must mean, he says, that "the morality reflected in their work is, in effect, that of the moralizer" (Baz, 2023, p. 83).

Similarly, Weinberg, Crowley, Gonnerman, Vandewalker, and Swain (2012) speak of the dangers of theory-contamination in our reliance on philosophical intuitions, citing the work of Goldman and Pust (1998) as an appropriate acknowledgment by mainstream philosophers of the danger. And they think that this danger reinforces the need for the calibration of philosophical intuitions along the line they proposed, namely, extrapolative calibration. What is interesting here is that extrapolative calibration presents a more feasible picture of the calibration challenge to philosophical intuitions, more feasible in relation to that of Cummins. Hence, they say: "We find Cummins' argument overlooks the potential for extrapolative calibration and underestimates the variety of resources potentially available for calibrating intuitions. So philosophers ought not at this time preemptively foreclose on the possibility of an epistemically fruitful calibration of intuition" (Weinberg et al., 2012, p. 266). Despite this concession, however, they think that although it might be possible to calibrate philosophical intuitions, none of the available strategies deemed acceptable by extrapolative calibration is fruitful as things currently stand in the discipline.

A few general points about the challenge so summarily expressed. The challenge in analytic philosophy is brought into sharp relief by the worry about cognitive artifacts (the preconceptions of philosophers, but also framing effects, order effects, etc.) distorting philosophical theorizing in unforeseen ways. If we take the preconceptions of philosophers as an example, this worry is that these preconceptions might bias philosophers' judgments about cases, the very judgments that are then used to support the theory in contention (see also Goldman & Pust, 1998; Knobe & Nichols, 2007; Turner, Nahmias, Morris, & Nadelhoffer, 2007). Given this, the implicit premise of the argument (i.e., the master argument) crisscrossing the views of the proponents of the calibration challenge is that since philosophical intuitions cannot be calibrated, we are not in a position to rule out the skeptical hypothesis that analytic philosophers are in that bad epistemic situation. In fact, both Cummins and Baz claim that analytic philosophers are actually in that skeptical situation, a situation where the outcome of their theorizing is deeply distorted by cognitive artifacts. With these initial clues, we can present more fully the master argument driving the calibration challenge to philosophical intuitions:

C1. Philosophical intuitions cannot be calibrated, that is, certified as reliably tracking the world or certified as reliably attuning us to our shared world.

C2. If we are not able to calibrate philosophical intuitions, we cannot rule out the skeptical hypothesis that the outcome of our theorizing based on appeal to philosophical intuitions as evidence is deeply distorted by cognitive artifacts.

C3. If we are to have much of the knowledge we typically attribute to ourselves in the context of using the philosophical method of cases we need to be able to rule out this skeptical hypothesis.

CC. Therefore, we are unable to have much of the armchair knowledge we typically attribute to ourselves. (Indeed, what seems to stand in place of those knowledge attributions or knowledge claims we make in the context of the philosophical method of cases may be no more than the mere distorted products of our cognitive artifacts).

Notice that the argument so presented ignores some key differences in the views of the proponents of the calibration challenge. For example, apart from their distinctive notions of calibration, one key difference is the following. Whereas Cummins (1998), Weinberg et al. (2012), and other participants in the debate think that the questions we are prone to ask in the context of the philosophical method of cases (does a particular term apply to the protagonist of the story?) can be answered truthfully or correctly, Baz challenges that assumption. Instead, he speaks merely about the sort of calibration that makes it more likely that what we say is competent, that is, comprehensible, intelligible, pertinent, responsible, helpful, and so on (Baz, 2023). In the light of this divergence in perspective, therefore, one might wonder about the rationale of presenting their views as specific instances of a master argument and as posing a distinctive calibration challenge to philosophical intuitions.

Despite this divergence in perspective, presenting their views this way is justifiable. For one, there are shared concerns, insights, and argumentative moves among the proponents as represented by the master argument. Crucially, they all share a thin conception of the calibration challenge, that is, the view that philosophical intuitions cannot be certified as "measuring" their intended target as observational sources and empirical judgments normally can. Their disagreement, in essence, lies in the thick description of the challenge, namely, the intricate details and contexts through which the challenge is brought into sharp relief and made sense of. For instance, Baz (2023) unpacks the challenge within the broader context of our linguistic practice, where calibration answers to the question of our linguistic competence, and asks whether we are intelligible, comprehensible, and so on. And Cummins (1998) and Weinberg et al. (2012) unpack it within the broader context of scientific inquiry where calibration answers to the question of correct or incorrect, true or false judgments, and so on. It is of course true that Baz (2016, 2017, 2023) has good reasons to think that we should not apply the model of the correct and the incorrect, the true and the false to our linguistic context,

including the context of the philosophical method of cases. But this particular debate bordering on the thick description is something I shall set aside in this paper in order to make progress on the narrow front of our inquiry. Indeed, unless inquiry is restricted in this manner, it is almost impossible to make progress in addressing any particular question. Moreover, my later engagement with the views of the proponents of the calibration challenge would not be done on terms incompatible with their own framework and context.

Now the main contention of this paper. I argue that the first premise of the master argument is false: philosophical intuitions can be calibrated along the methodological line proposed by Weinberg and colleagues, namely, extrapolative calibration. That is, the calibration challenge mounted with that very notion can be met. Admittedly, philosophical intuitions cannot be calibrated along the methodological line proposed by Robert Cummins or Avner Baz. But those notions of calibration are not plausible. This is because the relevant notion of calibration yields the wrong conclusion for a standard putative source of evidence as in the case of Cummins, or there are empirical and theoretical considerations that tell strongly against it, as in the case of Baz. If this is so, we can resist the second premise of the master argument and deny the conclusion as well. More positively, we can affirm that we are able to have some of the armchair knowledge we typically attribute to ourselves when philosophizing with the method of cases.

I proceed as follows. In section 2, I begin with Cummins' calibration challenge to philosophical intuitions and show why the view is not plausible. Then in section 3, I consider Baz's claim that our normal employments of words are calibratable in everyday contexts but not in the contexts of the philosophical method of cases and mount theoretical and empirical arguments against it. Finally, in section 4 I take up the calibration challenge as presented by Weinberg and colleagues and argue that this specific challenge can be met even as things currently stand in analytic philosophy. The paper

ends with some concluding remarks on the import of these discussions for skepticism about armchair or philosophical knowledge as raised in the master argument.

Let me say something here about the scope of this paper. This paper is only about the calibration challenge and not about the restrictionist challenge. The restrictionist challenge is the argument to the effect that there is some empirical evidence[2] showing that philosophical intuitions are unsuitable to play the role they normally play in philosophical inquiry. And further, that as a result of this, philosophers must at the present moment restrict their appeal to philosophical intuitions or suspend the practice altogether (Alexander, 2012; Alexander & Weinberg, 2006; Machery, 2017; Swain, Alexander, & Weinberg, 2008; Weinberg, 2007; Weinberg, Nichols, & Stich, 2001). The restrictionist challenge has received a great deal of attention in the literature.[3] Given this, I shall not discuss it or attempt to address it here. That said, let us begin with Robert Cummins.

2. Robert Cummins and the Calibration Challenge

Cummins was the first to highlight the problem underlying the calibration of philosophical intuitions. He begins his "reflection on reflective equilibrium," the title of his famous paper by noting that reflective equilibrium as a traditional philosophical methodology has the same structure with scientific methodology. Indeed, he says it is the recognition of this similarity of structure, implicitly or explicitly, which provides the authority for reflective equilibrium in philosophy.[4] They are both strategies of adjusting theory in the light of some data and vice versa. But then he notes, this source of authority for reflective equilibrium means that philosophical intuitions are epistemologically useful if they can play the role that observational data play in science (Cummins, 1998, pp. 113-115). Further, he notes

---

[2] See Machery (2017) and Stich and Machery (2023) for an excellent summary of the empirical evidence.

[3] The most prominent responses to the challenge are the meaning divergence response (Goldman, 2010; Lycan, 2006; Sosa, 2010), the expertise response (Grundmann, 2010; Hales, 2006; Horvath, 2010; Irikefe, 2020b; Kauppinen, 2007; Ludwig, 2007; Williamson, 2007, 2011), the mischaracterization response (Cappelen, 2012; Deutsch, 2015; Molyneux, 2014; Williamson, 2007), and the reflective response (Bengson, 2013; Kauppinen, 2007; Ludwig, 2007).

[4] For an alternative approach to the authority of reflective equilibrium, see Irikefe (2020a).

that observational data play a role in science because observational sources that generate the data can be calibrated. For example, it was fitting for the defenders of the theory Galileo was attacking to ask whether the mountains seen through the telescope were mere artifacts of the telescope. And "the proper response was to point the telescope at something of known size, shape, distance, color, and so on to determine what distortions it introduced; to calibrate it, in short" (Cummins, 1998, pp. 116-117).

By contrast, Cummins claims it is not possible to calibrate philosophical intuitions. This is because meaningfully doing so requires a test key that (a) gives access to the domain of philosophical intuitions, (b) does not in turn come from the deliverances of the instrument being calibrated (i.e., philosophical intuitions), (c) does not come from any source that itself stands in need of calibration, for example, the intuitions of the experts in the field[5] except the experts rely on something else entirely, namely, the best theory of the domain, and (d) does not rely on cases everyone already agrees on beforehand (Cummins, 1998, p.117-118).[6]

So, Cummins (1998) thinks that we could calibrate philosophical intuitions "but only on the assumption that there is some non-intuitive access to its targets" (p.117), plausibly in the form of a fully-worked-out theory of the domain of intuitions since such a theory can derive that certification on the strength of its theoretical virtues, such as its elegance, simplicity, fertility, and so on. But here is the further twist. He says that once we have such a theory, which can provide an independent access to the domain of philosophical intuitions, and which would help us to settle all questions about the domain, appealing to philosophical intuitions would become pointless. Concluding he says: "Philosophical intuition, therefore, is epistemologically useless, since it can be calibrated only when it

---

[5] In Cummins' words, "For, of course, they [the intuitions of the experts] have not yet been calibrated" (Cummins, 1998, p. 117).
[6] As Cummins (1998) pointed out about consensus intuitions, "if we know that everyone, including our subject, agrees on the test items, there is no point in administering the test" (p.117).

is not needed . . . But if we are not in a position to do this, philosophy should not have any faith in it" (Cummins, 1998, p. 118).

Cummins' analysis is very interesting. But there are two major worries with his calibration challenge. The first obvious worry is that his suggestion of how philosophical intuitions can be calibrated and the damning conclusion about it he arrives at are at odds with his liberal treatment of the calibration of material instruments like the telescope in the Galileo episode. For example, the known values in the domain of the telescope he appealed to, such as, known size, shape, distance, colors, etc., are known by means of sources that stand in need of calibration in clear violation of condition (c) above. Of course, it is true that Cummins does not presuppose that all calibration requires the kind of conditions he articulates for philosophical intuitions. But that just underscores the point I am making about how well-motivated those conditions are especially given the far-reaching conclusion he draws from them and given that his claim is that "philosophical intuitions can be calibrated only when it is not needed" in the light of the conditions he articulates for doing so (Cummins, 1998, p.118). The "only" here reinforces this worry because it highlights just how Cummins is really thinking about the conditions he uses in evaluating philosophical intuitions. Moreover, as Machery (2017) points out, even in the Galileo episode matters were not that simple:

> Galileo did much engineering work on the telescope, turning it into a scientific
>
> instrument, and the telescope was calibrated against objects on earth (e.g., towers), but
>
> it was obviously not calibrated against features of the moon's landscape (craters and
>
> mountains), and observations of the moon's landscape, of Jupiter's moons, and of the
>
> Pleiades and other constellations, all of which were reported in the *Starry Messenger*
>
> (1610), could have been qualitatively different from those of the earth—a possibility
>
> Galileo's opponents such as Martin Horky exploited to undermine his observations
>
> and conclusions (Machery, 2017, pp. 100-101).

Second, and more importantly, since perception in scientific inquiry is what really stands analogous to philosophical intuitions in philosophical inquiry rather than material instruments like the telescope as Cummins seems to suggest; it is not clear how applying this standard of calibration would vindicate perception, which is clearly a source of knowledge or evidence if anything is. To see this, consider the following very ordinary way people go about calibrating visual perception. One goes to an eye specialist, say an ophthalmologist, to check how good one's vision is and whether or not one needs some corrective measure such as a pair of glasses. The ophthalmologist administers a standard vision test using a Snellen chart that shows objects in various sizes. And based on one's visual performance on those objects the specialist pronounces one's vision as reliable within a normal human range, or else recommends some corrective measures for fixing it, say some medication and a pair of glasses. But this would not constitute calibrating one's vision under Cummins' account for dismissing philosophical intuitions since the certification of the reliability of that vision depends on the deliverances of a visual capacity, which has not been calibrated, namely, that of the ophthalmologist in clear violation of condition (c). Or if the ophthalmologist's vision has been calibrated, we run into the same problem since at some point in the chain of the calibration of visual perception we would have to admit the input of an uncalibrated perceptual capacity.

Perhaps, this gets the demand for calibrating visual perception wrong. As Brian Weatherson remarked in distinguishing intuition from perception: "There is a distinction to be drawn here, since perception divides into natural kinds, visual perception, tactile perception, etc., and we can use each of these to calibrate the others. It is hard to see how intuitions can be so divided in ways that permit us to check some kinds of intuitions against the others" (Weatherson, 2003, p. 4). Notice, however, that to rely on olfactory or on auditory perception, which have not been calibrated to calibrate visual perception, whether in intra-person or in inter-person context is once again relying on sources that stand in need of the very certification they seek to confer on other sources. Under the framework

Cummins uses for diagnosing the problem with philosophical intuitions, this would be like relying on an uncalibrated mercury thermometer to calibrate a platinum thermometer.[7]

The right lesson to draw here, though, is not that perception can never be calibrated. We do so all the time in very ordinary ways. Visual perception cannot be calibrated on the notion of calibration that Cummins suggests for calibrating philosophical intuitions.[8]

Next, let us consider a different formulation of the calibration challenge inspired by ordinary language philosophy.

3. Avner Baz and the Calibration Challenge

In a recent contribution to this debate, Baz (2023) presents what can be reasonably read as a calibration challenge to philosophical intuitions drawing from the resources of ordinary language philosophy, a challenge at the heart of which is a claim of fundamental discontinuity between two contexts, namely, the ordinary/empirical context and the philosophical context of theorizing. This specific challenge

---

[7] See also Sosa (2007) and Overgaard, Gilbert, and Burwood (2011).

[8] One might worry that the claim that perception in scientific inquiry is what really stands analogous to philosophical intuitions and that perception can only be calibrated on the basis of perception is too quick. And this is so because one, it ignores the role of measurement and of measuring the same magnitude by different methods in science. And two, it ignores all of the myriad ways we have of "calibrating" both perception and scientific theories: we interact with the world we perceive, we get into trouble or danger when we misperceive, and we build machines, send rockets to the moon, treat diseases (successfully or unsuccessfully). Given this, one might even ask, what is the equivalent of any of that in the case of philosophical intuitions? I reply as follows. First, the advancement made in metrology and our ability to measure the same magnitude by different methods in science, laudable though it is, does not really constitute a way of calibrating our basic perceptual modalities such as visual perception since claims about the reliability of those measuring instruments fundamentally presuppose their reliability. Second, one might ask, suppose that we are able to build machines with our scientific theories, send rockets to the moon, treat diseases and so on, do these by themselves supply full enough evidence that perception and our scientific theories are well calibrated, and that they are reliably helping us to track the world? There is a form of skepticism Kitcher (1992) calls sociological skepticism that answers in the negative and questions our right to be confident about such claims based on our interaction with the world (pp. 262-263). In citing Kitcher here, I am interested not just in what the sociological skeptic says about perceptual judgment and scientific theories but also the solution he proposes for dealing with the skeptical challenge. For both the problem and the solution illustrate why philosophical theories and intuitions on the one hand, and perceptual judgments and scientific theories on the other hand are in the same predicament. Here is Kitcher's solution after going through several proposals that fail to do justice to the sociological skeptic's argument: "Just as there is no showing that our perceptual experience is reliable without taking for granted a substantial part of the belief system that rests on perceptual experience, so too we cannot find a tradition-free standpoint from which to validate the deference to authorities that permeate our cognitive life. The perceptual skeptic might be right. So, too, might the sociological skeptic. However, once that concession has been made, we can take up an epistemological enterprise that is both important and manageable" (Kitcher, 1992, p. 266).

shall be my present focus, and not the full spectrum of Baz's objections to the philosophical method of cases, which he fully outlines in his previous work on the subject (Baz, 2016, 2017).[9]

Here is Baz:

> Outside philosophy—by which I here mean, outside philosophizing by way of the method of cases—to say of someone that she knows or does not know this or that (for example) is to *situate ourselves* significantly in relation to others and incur commitments and liabilities. If I tell you that so and so knows that such and such, in a context where it is taken for granted that such and such, and what I say is rightly understood to mean that so and so has already learned (realized, figured out, heard...) that such and such, and that you can, or should, rely on so and so, or expect her, to act as if such and such, and you do and she doesn't, then you might, under certain circumstances, reasonably complain against me for telling you that; under different circumstances, my telling you that could not reasonably be faulted, even if it ended up misleading you... And so on and so forth. *In these and similar ways, our ordinary and normal employment of 'know' (for example) is open to the feedback of reality, is subjectable to rational criticism or assessment, and in that sense is calibratable, as the answers we give to the theorist's questions are not* (Baz, 2023, p. 83).

The thought here is that the risks, the liabilities, and the commitments we undertake in giving our answers to some particular questions or in claiming that so and so *knows* (is free, responsible, etc.) act as a kind of guardrail in making it likely that what we say is competent (comprehensible, intelligible, pertinent, responsible, helpful, and so on). Further, this guardrail, it is claimed, is absent in the context of the philosophical method of cases. Here is Baz again: "For when, in the course of everyday life,

---

[9] In previous work, I have provided several responses to these other aspects of Baz's objections to the philosophical method of cases (Irikefe, 2022a, 2022b).

you misperceive, you are liable to bump your head against reality—in any number of ways, and sometimes literally; whereas if, in the theorist's context, you find yourself inclined to say that the protagonist in some Gettier-style story knows the proposition in question (for example), the worst that could befall you is bumping your head against the philosophical consensus" (Baz, 2023, p. 83).

The implication of this is a claim of a "damning" and "fundamental" discontinuity between the context of philosophers' appeal to intuitions and the context of everyday employments of our words. To put this differently, it means that our performances in the world of philosophical intuitions are in general disconnected from what else we may do or say or be expected to do or say in the world beyond philosophical intuitions (see Baz, 2023, p. 10). As Baz puts it in a previous work using Knobe's harm and help case (Knobe, 2003a): "There is, for example, no reason to expect those who say they 'agree' with the sentence 'The chairman of the board harmed [helped] the environment intentionally' to respond systematically differently to the chairman, or to an actual person who has done *just* what she did, from those who say they 'disagree' with the sentence" (fn. 27).

I shall mount two arguments against Baz's challenge. The first argument is less direct. It tries to show that Baz's guardrail cannot plausibly serve the function it is recruited to serve here, namely, pronouncing the context of philosophers' appeal to intuitions defective and the context of the everyday non-defective. And this is because it systematically makes people less competent in what they say in everyday contexts, that is, less comprehensible, intelligible, pertinent, responsible, helpful, and so on. If this is so, we cannot rely on it as plausible explanation of the calibration of our linguistic practice in the intended sense, and for establishing the sort of discontinuity Baz aspires to draw between philosophy and the everyday. The second argument I shall mount is more direct, and it is one I take to be more devastating for Baz's claims. It is this. There is empirical evidence indicating that Baz's claim of fundamental discontinuity between the two contexts is not plausible, and that our

performances in the world of philosophical intuitions can (in principle) be systematically connected to our performances in the world beyond philosophical intuitions, that is, in the world of the everyday.

3.1. Guardrail and Systematic Incompetence in What we Say

Begin with the first argument. Imagine that we have this measuring instrument that we take to be properly functioning, that is, measuring what it is supposed to measure in the laboratory in the very domain it is most suited for. Moreover, it gives us a straightforward way of calibrating other instruments and pronouncing them in the laboratory as being properly functioning or improperly functioning in that particular domain. Heretofore, it has led us to pronounce some instruments as properly functioning and others as not. Now, to our greatest dismay, we found out that this instrument is deeply problematic. More than half of the time, it is improperly functioning unbeknown to us and fails in measuring what it is supposed to measure. In fact, reports are now coming in that indicate that this instrument has led many scientists astray when they relied on it in their own investigations, although some other scientists have been very lucky in this regard.

I take it that among other things we may do, we should stop using this instrument as a means of pronouncing other instruments as properly functioning or improperly function. In fact, all the pronouncements we have made using this instrument can no longer stand. The instrument as we now know it is not just fit for *that* kind of work.

I claim that this applies to Baz's guardrail since the guardrail is in the business of distinguishing the context of philosophers' appeal to intuitions from the context of everyday linguistic practice and showing the former to be defective but the latter to be non-defective.

To see this, consider the following. In a notable survey by Brian Schaffner and Samantha Luks, they asked almost 700 American adults a very simple question: which of two photos showed a larger crowd, one, a clearly half-empty inauguration photo of Donald J. Trump and the other, a crowd-filled inauguration photo of Barack Obama? The result: 15% of voters for Trump said the half-empty

inauguration photo displayed more people, despite perceptual evidence indicating the contrary. However, among non-voters, both for Trump and Obama, or for anyone else, only 3% gave that problematic answer (Schaffner & Luks, 2017).

The question posed to the subjects was empirical, and there is something they can do in this case to give the right answer—look at the photos more closely, perhaps. Also, the question is not some idle question: it is deeply connected to the political commitments, and cares of the participants, indeed to two important political figures that have loomed large in American political space for the past decade or so. Moreover, this is a question that were the subjects to give the wrong answer, they would be bumping against reality, and the empirical evidence before them. All in all, this is a scenario where there is guardrail. However, the guardrail was no guarantee that a significant portion of Trump's voters give the appropriate answer in this case, one that is sensitive to the evidence before them or the way the world actually is. If anything, the guardrail systematically militates against doing so.

Nor is there anything awkward about the case. One can draw similar examples from issues ranging from global warming, vaccination, gun violence, tobacco smoke, and many other issues of contemporary social life. For example, one can imagine a case where subjects are asked about whether or not there is global warming, and where the relevant subjects are climate scientists, politicians, and ordinary citizens who have high stakes, interest, risks, and liability on whether or not the answer to this question lies in the affirmative or in the negative. And who precisely because of the risks, and liabilities that they would incur by answering the question in the affirmative become less competent in what they say, that is, less helpful, less responsible, less intelligible, and so on (see Oreskes & Conway, 2011).

Note that these people, that is, people in various sorts of echo chambers[10] as the participants in Schaffner and Luks' study seemed to be and people who strongly disagree or who are opinionated on issues about global warming, vaccination, gun violence, tobacco smoke, and so on because of some particular interests or risks they have on the subject may be millions of people on the planet. And guardrail systematically makes them less competent in what they say. In other words, in some circumstances, guardrail makes some cognizers unreliable sources of information.

A sympathizer to Baz's view might protest as follows: "I guess the thought is that social context can make people worse at gauging reality as well as better? But the sorts of accountability Baz mentions are in place in these cases too—we (rightly) criticize Trump supporters for their weak grasp on reality. The judgments Trump supporters make are calibratable, they're just refusing to adjust their beliefs in the light of widely available calibration-based evidence that they are wrong."[11]

It is helpful here to make a distinction on a first-order level between a subject who situates herself in a particular speech context in relation to another subject or audience who relies on her saying such and such in an actual speech context, and a second-order level between those subjects in the actual speech contexts and readers or observers who imaginatively project themselves into the actual speech context and seek to evaluate or criticize the subjects at the first-order level. When Baz talks about the sort of calibration that is available to us because of our linguistic guardrail, that calibration is presented as being in play at the first-order level. It is therefore not quite right to say that *our* ability to criticize these subjects at a second-order level is evidence that the guardrail is *effectively* and *rightly* present and doing its work at the first-order level. More importantly, invoking criticism here does not really address the problem. If the *risk* and *liabilities* of downgrading a big champion of one's political

---

[10] Echo chamber is an informational enclave where contrary voices, opinions, and experts have been systematically discredited. See Nguyen (2020) for an excellent analysis of the phenomenon.

[11] A similar but differently worded objection says that it is exactly the feedback of reality *that allows us* (italics, mine) to detect the bias in judgments made by the respondents. What I say here applies to this objection as well.

tribe is so powerful that subjects are willing to doubt their own very eyes and become less competent in what they say, they are also likely to dismiss such criticisms out of hand or better still use their ignoring them to signal their loyalty to their political tribe or echo chamber and become less competent in what they say at a more egregious level.

Here is another objection a sympathizer to Baz's view might make. "The example shows what hardly needs showing—that people are prone to deceive themselves when they have some vested interest in things being otherwise than they actually are. Granted. But how exactly is that supposed to undercut Baz's argument for discontinuity between our everyday employment of words such as 'know' or 'ought' and their 'attempted employment' in the theorist's context? Again, if there is a disagreement about crowd size, we could settle that disagreement by, for example, 'counting.'" The example does undercut Baz's argument since it shows that the guardrail so described by Baz (2023) does not give us a good story of how we are systematically competent in the answers we give to everyday empirical questions but not the answers we give to the sort of questions that arise in the context of philosophers' appeal to intuitions. Indeed, the guardrail Baz's account identifies seems not to be the sort of things we should rely on for that kind of explanation and demarcation.[12]

Can we settle the disagreement by some methods such as counting as suggested in the objection? That is highly unlikely. If the subjects in those studies and many others in various echo chambers are disposed not to believe their eyes and become less helpful, less cooperative, and so on when something is shown to them that conflicts with their *interests* and *commitments,* they would be disposed to manifest the same attitudes when the results of the counting conflict with their *interests*

---

[12] It is surely true that I have taken my example from everyday "playground" and not from our philosophical "playground." But precisely because the account provides a way to gauge not just our being competent in our linguistic performances in our philosophical playground but also in our being competent in our linguistic performances in everyday playground, showing that the account fails to tell a systematic and principled story in the latter case is enough to cause trouble for the account.

and *commitments* too. We may seek to employ some other epistemic sources or call on the experts. But that is unlikely to work either. Like other subjects in various echo chambers, sources or experts that do not align with our subjects' *interests* and *commitments* are likely to be preemptively discredited.

3.2. Against the Claim of Fundamental Discontinuity

Let us turn to the second argument. There is in fact a growing body of evidence showing that our performances in the world of philosophical intuitions can (in principle) be systematically connected to our performances in the world beyond philosophical intuitions, that is, in the world of the everyday. The evidence is from several experimental studies focusing on "utilitarian" judgments in sacrificial moral dilemmas such as the Footbridge case (Thomson, 1985) in which participants have to choose to sacrifice one person to save the lives of a greater number of people. When participants endorse this kind of choice, they are taken to be expressing a utilitarian perspective/judgment and when they refuse to endorse it, they are taken to be expressing a non-utilitarian or a deontological perspective/judgment. Several recent studies indicate that these utilitarian judgments made in the world of philosophical intuitions are correlated with anti-social traits in the world beyond philosophical intuitions, that is, in the world of everyday morality such as the exhibition of psychopathic traits (Bartels & Pizarro, 2011; Glenn, Koleva, Iyer, Graham, & Ditto, 2010; Wiech et al., 2013) and diminished empathetic concern for others (Choe & Min, 2011; Crockett, Clark, Hauser, & Robbins, 2010). I report on a similar study by Kahane, Everett, Earp, Farias, and Savulescu (2015), which built on those results. Beyond that, however, they seek to challenge the assumption that utilitarian judgments express "the kind of general impartial concern for the greater good that is at the heart of utilitarian ethics" (Kahane et al., 2015, p. 194), a view they attribute to Greene (2007) and Singer (2005).

Participants in the study were asked to complete a questionnaire with two parts, the first consisting of *four sacrificial moral dilemmas*, the Footbridge case including, and the second consisting of individual differences measures. On moral dilemmas, participants were presented with sacrificial moral

dilemmas in moral philosophy such as the Footbridge case and asked, "from a moral point of view, should you perform the action, that is, push the stranger in the Footbridge case?" Afterwards, they were asked to rate the wrongness of this action on a scale of 1 to 5. On the measurement of *business ethics*, they use the scale from Cooper and Pullig (2013) and included six items describing ethics violations, such as "an underpaid executive padding his expense account by about $3,000 a year." And for each of the scale items, participants were asked to rate the acceptability of the conduct described, with higher scores indicating more lenient assessment of the conduct. On the measurement of *primary psychopathy*, the chosen scale consisted of sixteen items, including "success is based on survival of the fittest" and "I am not concerned about the losers." And on the measurement of *empathetic concern*, the chosen subscale from Interpersonal Reactivity Index consisted of seven items, including "When I see someone taken advantage of , I feel kind of protective towards them."

Kahane et al. (2015) found that:

a. The endorsement of "utilitarian" solution in personal moral dilemmas such as the Footbridge case was associated with a lower wrongness ratings of the "utilitarian" action (r = .68, p < .001).

b. Lower wrongness ratings of business ethical violations correlated with greater endorsements of "utilitarian" solutions to the sacrificial moral dilemmas (r = .25, p < .001) and lower wrongness ratings of the "utilitarian" action (r = .31, p < .001).

c. Lower wrongness ratings of business ethical violations correlated with increased primary psychopathy (r = .58, p < .001) and reduced empathic concern (r = .29, p < .001) (p.196).

They conclude that "'utilitarian' judgment was associated with more lenient assessment of immoral behavior in the Business Ethics measure. This association is directly between 'utilitarian' judgment and an amoral pattern of judgment rather than, as in prior studies, only between 'utilitarian' judgments and reduced empathic concern or measures of antisocial personality traits" (pp.196-197). I take these

results to undercut the claim that in general there is a *fundamental* discontinuity between our performances in the world of philosophical intuitions and our performances in the world of the everyday. More specifically, utilitarian judgment in the world of philosophical intuitions whose hallmark is moral expediency is systematically correlated with moral expediency in everyday contexts, for example, in the business dealings of subjects in ordinary contexts.[13]

How should we extrapolate this to every domain of intuitions in the absence of similar studies in every domain of intuitions? Here is how. I claim that these results support the following general conclusion. Our (linguistic) performances in the context of philosophers' appeal to intuitions and our performances in the context of the everyday can (in principle) be systematically correlated. Notice that this is a modal claim. It posits only the possibility of this connection in every domain of philosophy, which is enough to fend off the kind of skepticism that Baz's argument leads to. To refute this weak modal claim, one would have to show that *the connection is impossible in every domain*, which is clearly not the case.

Next, I turn to a third version of the calibration challenge.

4. Weinberg and Colleagues and the Calibration Challenge

The final version of the calibration challenge is due to Weinberg and colleagues, and presents a more realistic outlook of the challenge of calibrating philosophical intuitions, more realistic in relation to Robert Cummins. They express the marked difference this way:

---

[13] One might object as follows: Unlike cases of perception, it is unclear what would be the analogue of "bumping one's head against reality" in the moral domain. Response: "Bumping one's head against reality" would be an obstacle to calibrating philosophical intuitions in the moral domain if "bumping one's head against reality" provides a straightforward way of calibrating perception and the equivalent facts or reality are not available in the moral domain. But the first argument against Avner Baz's challenge shows clearly that bumping one's head against reality is not an indubitable standard for calibrating perception because there are cases in perception where cognizers are directly bumping their heads against reality and yet they are incompetent in what they say. Given this, the prospect of bumping one's head against reality provides no indubitable standard to ruling out intuitions, including moral intuitions even if we concede that there are no analogous facts or reality to bump one's head against in the moral domain.

Our concern is that Cummins has overlooked important resources and strategies for calibration in general, and intuitions in particular. For not all calibration requires a certified basis that covers the whole target domain. Sometimes one can calibrate using a combination of information about some proper subset of the target domain on the one hand, and about the nature of the instrument and how it is meant to be able to track that domain on the other hand. We will call the first sort of information *the partial certified basis*, and the other *the theory of the instrument*, and together they license an *extrapolative inference* to the instrument's accuracy over the target domain. We thus call this calibration strategy *extrapolative calibration*, which one might contrast with the *exhaustive calibration* strategy that Cummins considers (Weinberg et al., 2012, p. 264).

Note that the "partialness" in the designated subset of the target domain here is a specification that the relevant basis should not exhaust the entirety of the target domain so that our calibration of an instrument dependent on that basis would not be epistemologically pointless. Rather, it would enable us to learn something new about the domain.

In the case of philosophical intuitions, they think the following are candidates for partial certified basis: ordinary intuitions, consensus intuitions,[14] reflective intuitions, expert intuitions, and clear/forceful intuitions. That said, they argue that although it is possible to calibrate philosophical intuitions, possibility does not entail feasibility. And here is why. Ordinary and easy intuitions lack "a theory of the instrument even a folk one that will ground the extrapolation of the trustworthiness of their basis to a majority of the intuitions at work in philosophical contexts" (Weinberg et al., 2012, pp. 271-272), and clear/forceful, reflective or expert intuitions do not form a proper partial certified bases

---

[14] Following Cummins (1998), Weinberg et.al (2012) define "consensus intuition" as the intuition everyone agrees on. But it is hard to think of any philosophical intuition, or case judgment *everyone* literary agrees on. So along with Weinberg and colleagues, this category can conveniently be set aside.

of the domain "even though their candidate theories of the instrument do generalize to philosophical contexts" (Weinberg et al., 2012, p. 272).

It is noteworthy that this view gives us something much more nuanced and balanced than the view of Cummins, which also relies on drawing close analogy between philosophical practice and scientific methodology. For example, the present view allows, in principle, that intuition can be calibrated using intuition. And if we extend this logic, it would also allow that perception can be calibrated using perception, for example, using a partial certified basis, say that of visual perception and some folk theory about the working of perception that would allow us to extend its trustworthiness or reliability to the domain of visual perception. Thus, the familiar difficulties that plague Cummins' view does not seem to arise here.

Let us grant for the sake of the argument that Weinberg et.al (2012) are right that clear/forceful, reflective, or expert intuitions[15] do not form partial certified bases. Nonetheless, there is still some room for the prospect of calibrating philosophical intuitions using some partial certified basis, contrary to Weinberg and colleagues.

4.1.   The Prospect of Extrapolative Calibration of Philosophical Intuitions

To start with, we can depend on ordinary and easy intuitions, the bulk of intuitions that belong to the world of the everyday as our partial certified bases. These intuitions include the classificational judgments of everyday mental states (mindreading, simulation, or folk psychology) and the classificational judgments of moral conducts and misconducts, such as, that a particular action is kind or cruel, and so on. On the view of Weinberg et.al (2012), however, the only missing piece to using easy and ordinary intuitions as a basis for calibrating philosophical intuitions is that we lack "a theory of the instrument even a folk one that will ground the extrapolation of the trustworthiness of their

---

[15] See Schwitzgebel and Cushman (2012).

basis to a majority of the intuitions at work in philosophical contexts" (Weinberg et al., 2012, pp. 271-272).

But there is actually such a folk theory, I claim. It is called the claim of epistemological continuity, recently defended by a number of theorists sympathetic to armchair philosophy. And it says that the basis by which we classify philosophical cases is the same basis by which we classify everyday mental states, such as states of knowledge and states of ignorance, having nothing distinctively philosophical about it. For example, Williamson argues that the epistemology of philosophical thought experiments is merely the application of our capacity to evaluate ordinary counterfactual conditionals, and this involves the use of imaginative simulation.[16] Jennifer Nagel also defends a version of the claim of epistemological continuity, brilliantly captured in the title of her paper: "Mindreading in Gettier cases and Sceptical Pressure Cases." In particular, she says: "Ascribing knowledge [in Gettier cases] is like recognizing a friend's face, or seeing that she looks sad" (Nagel, 2012b, p. 171). However, unlike Williamson's account where off-line simulation plays the most distinctive role in the relevant epistemology, her account is ambivalent about the right model of our folk psychological capacities at work in the process. Elsewhere too (Author, 2022), I have argued that what underwrites the claim of epistemological continuity is not Williamson's off-line simulation or Nagel's mind-reading capacity but our socially acquired skill or competence to know the requirement of reasons for or against a given judgment (or conduct), for example, in the moral domain to know that cruelty is wrong through our sensitivity to the salient facts that obtain or fail to obtain in the situation. There I presented an account too of how we acquire this competence, the various features

---

[16] Williamson says imaginative simulations are "the most distinctive cognitive features of the process of evaluating them [i.e., such counterfactual conditionals]"(Williamson, 2007, p. 152).

of it that make it a suitable epistemological basis and how the account is not vulnerable to some objections mounted against the claim of epistemological continuity in the recent literature.[17]

Further, it is good to note that the claim has both descriptive and normative dimensions. It is descriptive in the sense that it seeks to capture the underlying competence in our classificational practices in analytic philosophy using a non-tendentious and non-mysterious everyday capacities. The motivation for this as Williamson rightly noted is that since humans did not evolve under the pressure to do philosophy, postulating a unique capacity exclusive to armchair philosophizing makes the whole enterprise looks suspicious, and susceptible to the temptation that it can be obliterated without any collateral damage (Williamson, 2007). The claim is also normative: it gives us a rough diagnostic of the kind of cases we ought to be relying on. We should be relying on cases with "mundane realism" understood here in terms of how easily the event or combination of events described in the case can occur given the way the world is right now. Notice that mundane realism so understood does not track the frequency or infrequency or probability or improbability of an event or combination of events. An event can be mundane in this sense and yet the event does not frequently occur and probabilistically unlikely to occur. A good example is having a winning ticket in a national lottery. If I am a lottery player, the extraordinary event in which I have the winning ticket is something very probabilistically unlikely to occur, indeed, in the order of several million-to-one. And yet, it could

---

[17] Such as that of Baz (2016, 2017) and Nado (2015). Admittedly, Baz (2016, 2017) was the first to recognize the central place of the claim of epistemological continuity in the recent metaphilosophical literature. Baz specifically challenges the linguistic version of this thesis, the claim that what we are invited to do linguistically in the context of philosophical intuitions is not significantly different from "something that we routinely do in the course of everyday experience, and which underlies our ordinary and normal employment of our words"(Baz, 2017, p. 32). In effect, Baz argues that there is a "fundamental" and "damning" discontinuity between the context of philosophical intuitions and the context of everyday employments of human words. In the previous section, I provided an empirically informed argument against the plausibility of the claim. In previous work too (Irikefe, 2022b), I have also shown that my view of the claim of epistemological continuity does not take on board some of the problematic assumptions Baz attributes to some recent defenders of the linguistic version of the claim of epistemological continuity such as Williamson (2007), Cappelen (2012), and Jackson (2011).

easily occur because very little changes need to be made to how the world is right now for my ticket to be the winning ticket, namely, just a few colored balls falling with a little different combinations.

That a case should be mundane in this sense captures a platitude in the recent debate about what a good case should be, namely, that such a case should not be fantastical. The idea comes out nicely in an exchange between Alvin Plantinga and Linda Zagzebski on the brain lesion case. Plantinga had introduced the brain lesion case to show that reliabilism does not give a sufficient condition for knowledge. Plantinga imagines an agent zapped by rays from the outer space that gives the agent a brain lesion, which disrupts the agent's normal cognitive processes but leaves the agent with a reliable process in the formation of true beliefs in a very limited region, for example, the belief that the agent has a brain lesion (Plantinga, 1993). Commenting on this case, however, Zagzebski makes the following illuminating point:

> There are several reasons why we should be dubious of such an example. In the first place, I think that an epistemic theory has the methodological right to take for granted that the theory offers an account of the nature of human knowledge within a normal human environment. Counterexamples based on very abnormal conditions have much weaker force against such a theory than those that arise from normal circumstances, although this is not to say that they have no force at all (Zagzebski, 1996, p. 304).[18]

As an alternative to the brain lesion case, Zagzebski suggests the case of *Oblonsky*, a character described by Tolstoy in *Anna Karenin* who has a reliable but defective belief forming processes that prevent the character from knowing. The point here is that part of the reason this move is pointing us in the right direction (besides the one that Zagzebski gives) is that we want a case to guide us to something *deep* and *interesting* about the world or about our conceptions of the world and not to some *ad hoc* feature/quality we have identified in the world or in our conceptions of it. And the worry is

---

[18] For a similar point in the philosophy of language, see Devitt and Porter (2021).

that by relying on fantastic examples we risk missing out on this important element. And this is so because fantastic examples typically portray *ad hoc* constructions and combination of events, not ones that are likely to occur given the way the world is right now.

The usefulness of such a notion (i.e., mundane realism) is that it allows us to distinguish the brain lesion case (Plantinga, 1993), the Swampman case[19] (Davidson, 1987), and cases of fissioning/fusioning/teleporting pairs (Parfit, 1987) on the one hand, and the Oblonsky case (Zagzebski, 1996), the hospital version of the Gettier case (Machery et al., 2017), the stopped watch version of the Gettier case (Nagel, 2012a) and similar cases in analytic philosophy on the other hand.[20] The world in which the events described in the former cases can occur are very much unlike our world and it would take some alterations of the laws of our biological and physical processes for those events to occur whereas the events described in the latter cases can easily occur. For example, all it takes for the combination of events described in the hospital case to occur is for Paul Jones to have one hell of an evening in which Mary Jones (i.e., Mrs. Jones) has a heart attack on her way back home (perhaps, given her medical history), she is admitted to a hospital and the hospital Mr. Jones calls to inquire about her has a patient bearing the name "Mary Jones." So, despite both classes of cases pulling apart contingently occurring features and how infrequently we encounter or read texts about them (Machery, 2017, pp. 113-120); there are fundamental differences between both classes of cases that make cases of fissioning/fusioning/teleporting pairs and similar ones defective and the hospital case and similar ones good for theorizing. For example, the pulling apart in both classes of cases are not in the same order of magnitude.

---

[19] Bracketing for a moment their dialectical uses (see Machery, 2017).

[20] Both the stopped watch and the hospital versions of the Gettier case would be instances of a class of Gettier case Blouw, Buckwalter, and Turri (2017) call "failure to detect accompany by similar replacements", that is, in these cases "the agent fails to detect the truth, but her judgment is nonetheless made true by a state of affairs similar to what she based her belief on" (p. 247). Further, these cases are unlike fake barn version of Gettier cases that yields a high likelihood of knowledge attributions (Blouw et al., 2017; Colaço, Buckwalter, Stich, & Machery, 2014; Ziólkowski, 2016).

4.2. Moving from Claim about our Reliability in the Ordinary and the Easy to Claim about our Reliability in Philosophical Contexts

Given the foregoing explanation of the folk theory required for extrapolative calibration, we can now explain how we can project from the reliability of intuitions in the ordinary and the easy to reliability of intuitions at work in philosophical contexts. In our philosophical "playground" the environment for the normal operation and reliability of our everyday intuitions are still intact because, for example, in mundane cases, the cases describe events that involve the same biological and physical processes of the universe. To borrow a metaphor here, in mundane cases we are like ordinary chess players (not necessarily chess masters!) who can be counted upon to comprehend new moves and draw interesting connections and associations, even ones we rarely encounter. And we can be counted upon to be reliable in doing so because the fundamentals of our playing environment have not been altered. We are still playing and engaging our opponents using the same chess board positions and pieces, and so on. Human cognitive competences are adaptive in this limited sense, including our competence to recognize what facts obtain or fail to obtain in cases.[21] Indeed, given how pivotal events or situations that can easily occur albeit rarely occur or infrequently occur have been to our evolutionary history, it would be strange if our competences are not suited by adaptative pressures to anticipate, evaluate, and judge reliably situations or events in this general class.[22]

---

[21] Machery argues that because of the difference between unusual situations and everyday situations, the readers of philosophical cases involving unusual situations cannot rely on heuristics that work well for everyday situations. Given what I have said so far, we see what is missing here (Machery, 2017, p. 115). The claim ignores fundamental differences between philosophical cases and the way in which most philosophical cases that are mundane are really "modally close" to the actual world. I am thinking of modality here in terms of similarity ordering of possible worlds as spelt out by Lewis (1973, 1986).

[22] For example, tsunami and events that led to the extinction of the dinosaurs. One might protest that the events or combination of events in philosophical cases are not about survival but really about some idle combination of events constructed by philosophers. But the hospital case, for example, although involves questions about whether someone knows describes combination of events bordering on someone's survival.

But note that this claim of reliability and claim about our projection are not to the entirety of philosophical intuitions, for clearly the experimental data do not support such a generous leap.[23] The claim rather is that we should project to the class of mundane cases, which involve scenarios that are likely to engage reliably our cognitive capacities with the *proviso* that whether a philosophical case does in fact generate reliable judgment is a matter to be finally settled by results from experimental philosophy and psychology. Thus, work in experimental philosophy and psychology provides valuable data for the process of calibrating philosophical intuitions. Notice too that because mundane realism is only a good but not a conclusive indication that a case would generate reliable judgments, we can have a situation where the mundane realism of a case indicates that it would be reliable, but actual experimental work shows that it is not. The reason for this is that being mundane is not the only factor that determines reliability. Reliability also depends on the absence of information that generates framing effects, culturally or demographically sensitive information, etc., in the linguistic composition of a specific case.[24] Moreover, using a rule of thumb or a folk theory such as mundane realism as a rough indication that something has a given property (that is, reliability or trustworthiness) is only valuable in the absence of empirical knowledge. Once we have established empirical knowledge about a case, a rule of thumb about such a case ceases to serve its function. This is true not only about philosophical intuitions and our process of calibrating it but also true of perception and our process of calibrating it too. Once we have well established scientific knowledge about vitiating circumstances where perception yields unreliable data, using rule of thumb or folk theory as a rough indication to know when it is delivering reliable data in that specific circumstance ceases to be meaningful.

If I am right about the foregoing analysis, the disturbing characteristics about philosophical intuitions Machery (2017) identifies such as pulling apart features that contingently occur, infrequency

---

[23] Again, see the excellent summary of the experimental data by Machery (2017), and Stich and Machery (2023).
[24] See example below in the discussion of a recent work on Kripke's Gödel-style case.

in our encountering of such cases, and so on are no barriers to the certification and use of philosophical intuitions. The trick is to recognize that we are only after a subset of philosophical intuitions. And indeed, there are philosophical intuitions that have been reliable in the light of the experimental data while having these disturbing characteristics. Here, I focus on results from experimental studies on Gettier cases, the Gödel and Jonah cases, and the Truetemp case. My focus on the first two is to affirm reliability in the light of claims about demographic effects; and my focus on the last case is to affirm reliability in the light of claims about contextual effects in the recent literature.

In their paper, Weinberg et al. (2001) report results showing cross-cultural variations in intuitions when subjects were presented with the American car case (the only Gettier case used in the study). But subsequent experimental studies failed to replicate this. First, there is the result obtained by Nagel (2012a) who presented subjects drawn from different ethnic backgrounds (White, South Asian, East Asian, Latin American, Black) with eight Gettier-style cases, including the stopped watch case. The results show that roughly two-thirds of participants shared the Gettier intuition, that is, classified the Gettier-style cases as involving the absent of knowledge. And there is no statistically significant correlation between knowledge attribution and ethnicity or gender (Nagel, 2012a; Nagel, San Juan, & Mar, 2013). Similarly, Turri (2013), Kim and Yuan (2015), and Seyedsayamdost (2015a, 2015b) report experimental results that show that the Gettier intuition is uniform across culture, socio-economic group, and gender. More remarkably are the results obtained by Machery, Stich, et al. (2015) and Machery et al. (2017). Machery, Stich, et al. (2015) present subjects drawn from four different cultural groups with distinct languages (Brazil, India, Japan, and the USA) with Gettier-style cases, including the hospital case. They report results indicating that Gettier intuition is uniform across culture. Similarly, Machery et al. (2017) present the hospital case to subjects located in 23 countries and 17 languages. The results show that between roughly 70% and 90% of participants give

answers in line with the Gettier intuition. The results also show that the Gettier intuition does not vary along gender differences. Machery et al. (2017) view these results as providing convergent support for the hypothesis that the Gettier intuition is part of a core folk epistemology.

Next, in one of the landmark experimental philosophy papers, Machery, Mallon, Nichols, and Stich (2004) report results indicating cross-cultural variation in semantic intuitions, when subjects were presented with Kripke's Gödel-style and Jonah-style cases. While 56.5% of Western participants give response in line with Kripkean intuition, that is, in line with Kripke's causal historical view when they evaluated the Gödel-style case, only 31.5% of the Chinese participants do so, that is, a majority of these subjects tended to have descriptivist intuitions. These results were initially replicated and bolstered by the results obtained by Machery et al. (2010), Machery, Sytsma, and Deutsch (2015), Sytsma, Livengood, Sato, and Oguchi (2015), and Beebe and Undercoffer (2015, 2016).

However, in a more recent work, Izumi, Kasaki, Zhou, and Oda (2018) identified an overlooked confound in the experimental studies that have sought to replicate the findings of Machery et al. (2004) using East Asian languages (i.e., Machery et al., 2010; Machery, Sytsma, et al., 2015; Sytsma et al., 2015). These studies, they argue, depend on the use of articleless "bare noun phrases," in Chinese and Japanese, which are known to be multiply ambiguous, and which may have distorted the answers of the Chinese and Japanese participants in those studies who are fluent in Japanese. Izumi et al. (2018) conducted two new experimental studies in Japanese to replicate Machery et al. (2004) taking the observation into account. They report significant differences in the results of probes contingent on the phrasing of the probes. In response to the replication probes with the bare noun phrases, 30.6% (N=111) of the participants answer in line with causal-historical view. And in response to the replication probes without the bare noun phrases, 50% of the participants give answer in line with the causal-historical view (N=110). Thus, results without the bare noun phrases correspond to the results obtained by the original experimental studies of Machery

et al. (2004) in which 56.5% of the Western/American participants selected the causal-historical view. The implication is that the results of Machery et al. (2004) have not been replicated.

I take my last example from experimental studies on contextual effects in philosophical intuitions. Swain et al. (2008) report experimental results showing that judgment on the Truetemp case (Lehrer, 1990) is susceptible to priming effects: laypersons are less likely to attribute knowledge in the Truetemp case when first shown an instance of clear case of knowledge, and more likely to attribute knowledge when first shown a clear case of the absence of knowledge. However, in a recent experimental study to replicate those results involving three experiments, one devised in Polish, and the other two devised in English, Ziółkowski (2021) found no evidence of priming effects for knowledge ratings regarding the Truetemp case. Even more striking are three experiments conducted by Ziółkowski, Wiegmann, Horvath, and Machery (2023) that show that prediction of order effects on the Truetemp case by Swain et al. (2008) could not be consistently and robustly replicated.

In presenting these results, I am not trying to show that all philosophical cases that have been shown to generate unreliable results would follow this pattern. My claim is more modest: it is not completely all gloom and doom despite many of our current cases suffering from worrisome demographic and contextual effects. Moreover, we can and should view some of the empirical data such as those about contrast effects[25] (Gerken & Beebe, 2014; Schaffer & Knobe, 2012) and Knobe effects[26] in moral psychology (Knobe, 2003a, 2003b, 2004) and epistemology (Beebe & Shea, 2013) as supplying us valuable and interesting information about circumstances where a straightforward appeal to philosophical intuitions is not legitimate. Indeed, in the light of so much work that has been done in experimental philosophy and psychology over the last two decades or so, it is fair to say that our understanding of the workings of philosophical intuitions has never been in a better shape.

---

[25] "That folk inclinations to agree with ascription of knowledge to a subject depend not only on the subject's epistemic position but also on what aspects of it are 'in contrast'" (Gerken & Beebe, 2014, p. 1)
[26] That the goodness or badness of side effects influences intuitions about intentional actions.

5. Conclusion

The skeptical conclusion in the master argument we began with depends on the premise that philosophical intuitions cannot be calibrated. But given the prospects presented by extrapolative calibration and the options defended in the previous section of the paper, we can see why that premise is false and why we are able to resist the skeptical conclusion that follows from it. More positively, we can affirm that we are able to have some of the armchair knowledge we typically attribute to ourselves when philosophizing with the method of cases. I should also note that what all these portend for the calibration of philosophical intuitions is that this process is an on-going process rather than a once-off event done by members of the philosophical community. That may mark one way in which the calibration of a basic evidential source differs from the calibration of our ordinary scientific instruments.

Acknowledgements

Declarations

I hereby declare that I have no conflict of interest in connection with this submission.

Reference

Alexander, J. (2012). *Experimental Philosophy: An Introduction*: Polity.
Alexander, J., & Weinberg, J. (2006). Analytic Epistemology and Experimental Philosophy. *Philosophy Compass, 2*, 56-80. doi:10.1111/j.1747-9991.2006.00048.x

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*(1), 154-161. doi:https://doi.org/10.1016/j.cognition.2011.05.010

Baz, A. (2016). Recent Attempts to Defend the Philosophical Method of Cases and the Linguistic (Re)turn. *Philosophy and Phenomenological Research, 92*(1), 105-130.

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

Baz, A. (2023). On philosophical idling: the ordinary language philosophy critique of the philosophical method of cases. *Synthese, 201*(3), 1-20.

Beebe, J., & Shea, J. (2013). Gettierized Knobe Effects. *Episteme, 10*(3), 219-240. doi:10.1017/epi.2013.23

Beebe, J., & Undercoffer, R. (2015). Moral valence and semantic intuitions. *Erkenntnis, 80*, 445-466.

Beebe, J., & Undercoffer, R. (2016). Individual and Cross-Cultural Differences in Semantic Intuitions: New Experimental Findings. *Journal of Cognition and Culture, 16*(3-4), 322-357.

Bengson, J. (2013). Experimental attacks on intuitions and answers. *Philosophy and Phenomenological Research, 86*(3), 495-532.

Blouw, P., Buckwalter, W., & Turri, J. (2017). Gettier Cases: A Taxonomy. In R. Borges, C. de Almeida, & P. Klein (Eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (pp. 242-252): Oxford University Press.

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

Choe, S. Y., & Min, K.-H. (2011). Who makes utilitarian judgments? The influences of emotions on utilitarian judgments. *Judgment and Decision Making, 6*(7), 580-592. doi:10.1017/S193029750000262X

Colaço, D., Buckwalter, W., Stich, S., & Machery, E. (2014). EPISTEMIC INTUITIONS IN FAKE-BARN THOUGHT EXPERIMENTS. *Episteme, 11*(2), 199-212. doi:10.1017/epi.2014.7

Cooper, M. J., & Pullig, C. (2013). I'm Number One! Does Narcissism Impair Ethical Judgment Even for the Highly Religious? *Journal of Business Ethics, 112*(1), 167-176. doi:10.1007/s10551-012-1239-0

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences, 107*(40), 17433-17438.

Cummins, R. C. (1998). Reflection on Reflective Equilibrium. In M. DePaul & W. Ramsey (Eds.), *Rethinking Intuition* (pp. 113-128): Rowman & Littlefield.

Davidson, D. (1987). Knowing One's Own Mind. *Proceedings and Addresses of the American Philosophical Association, 60*(3), 441-458. doi:10.2307/3131782

Deutsch, M. (2015). *The myth of the intuitive: Experimental philosophy and philosophical method*: The MIT Press.

Devitt, M., & Porter, B. C. (2021). Testing the Reference of Biological Kind Terms. *Cognitive Science, 45*(5), e12979. doi:https://doi.org/10.1111/cogs.12979

Gerken, M., & Beebe, J. (2014). Knowledge in and out of Contrast. *Nous, 50*(1), 133-164.

Glenn, A. L., Koleva, S., Iyer, R., Graham, J., & Ditto, P. H. (2010). Moral identity in psychopathy. *Judgment and Decision Making, 5*(7), 497-505. doi:10.1017/S1930297500001662

Goldman, A. (2010). *Philosophical naturalism and intuitional methodology*. Paper presented at the Proceedings and Addresses of the American Philosophical Association.

Goldman, A., & Pust, J. (1998). Philosophical Theory and Intuitional Evidence. In M. Depaul & W. Ramsey (Eds.), *Rethinking Intuition: The Psychology of Intuition and its Role in Philosophical Inquiry*: Rowman & Littlefield.

Greene, J. (2007). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*: MIT Press.

Grundmann, T. (2010). Some hope for intuitions: A reply to Weinberg. *Philosophical Psychology, 23*(4), 481-509.

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

Horvath, J. (2010). How (not) to react to experimental philosophy. *Philosophical Psychology, 23*(4), 447-480. doi:10.1080/09515089.2010.505878

Irikefe, P. (2020a). An Agent-Based Account of the Normativity of Reflective Equilibrium. *Philosophia, 48*(1), 217-225.

Irikefe, P. (2020b). A fresh look at the expertise reply to the variation problem. *Philosophical Psychology, 33*(6), 840-867. doi:10.1080/09515089.2020.1761541

Irikefe, P. (2022a). Avner Baz's Ordinary Language Challenge to the Philosophical Method of Cases. *Dialectica, 999*(1).

Irikefe, P. (2022b). The epistemology of thought experiments without exceptionalist ingredients. *Synthese, 200*(3), 191. doi:10.1007/s11229-022-03690-2

Izumi, Y., Kasaki, M., Zhou, Y., & Oda, S. (2018). Definite descriptions and the alleged east–west variation in judgments about reference. *Philosophical Studies, 175*(5), 1183-1205. doi:10.1007/s11098-017-0902-9

Jackson, F. (2011). On Gettier Holdouts. *Mind and Language, 26*(4), 468-481.

Kahane, G., Everett, J. A. C., Earp, B. D., Farias, M., & Savulescu, J. (2015). 'Utilitarian' judgments in sacrificial moral dilemmas do not reflect impartial concern for the greater good. *Cognition, 134*, 193-209. doi:https://doi.org/10.1016/j.cognition.2014.10.005

Kauppinen, A. (2007). The rise and fall of experimental philosophy. *Philosophical Explorations, 10*(2), 95-118.

Kim, M., & 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. doi:10.1017/epi.2015.17

Kitcher, P. (1992). Authority, deference, and the role of individual reasoning in science. *The social dimensions of science*.

Knobe, J. (2003a). Intentional action and side effects in ordinary language. *Analysis (United Kingdom), 63*(3), 190-194.

Knobe, J. (2003b). Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology, 16*(2), 309-325.

Knobe, J. (2004). Folk Psychology and Folk Morality: Response to Critics.

Knobe, J., & Nichols, S. (2007). An experimental philosophy manifesto. In J. Knobe & S. Nichols (Eds.), *Experimental Philosophy* (pp. 3--14): Oxford University Press.

Lehrer, K. (1990). *Theory of Knowledge* (Vol. 42): Westview Press.

Lewis, D. (1973). Counterfactuals and comparative possibility. In *IFS: Conditionals, Belief, Decision, Chance and Time* (pp. 57-85): Springer.

Lewis, D. (1986). *On the Plurality of Worlds*: Wiley-Blackwell.

Ludwig, K. (2007). The Epistemology of Thought Experiments: First Person versus Third Person Approaches. *Midwest Studies in Philosophy, 31*(1), 128-159.

Lycan, W. G. (2006). On the Gettier problem problem. In S. Hetherington (Ed.), *Epistemology Futures* (pp. 148--168): Oxford University Press.

Machery, E. (2017). *Philosophy within its proper bounds*: Oxford University Press.

Machery, E., Deutsch, M., Mallon, R., Nichols, S., Sytsma, J., & Stich, S. P. (2010). Semantic intuitions: Reply to Lam. *Cognition, 117*(3), 361-366. doi:10.1016/j.cognition.2010.08.016

Machery, E., Mallon, R., Nichols, S., & Stich, S. (2004). Semantics, cross-cultural style. *Cognition, 92*(3), 1-12.

Machery, E., Stich, S., Rose, D., Alai, M., Angelucci, A., Berniūnas, R., . . . Zhu, J. (2017). The Gettier Intuition from South America to Asia. *Journal of Indian Council of Philosophical Research, 34*(3), 517-541. doi:10.1007/s40961-017-0113-y

Machery, E., Stich, S., Rose, D., Chatterjee, A., Karasawa, K., Struchiner, N., . . . Hashimoto, T. (2015). Gettier Across Cultures. In (pp. 645-664): Nous.

Machery, E., Sytsma, J., & Deutsch, M. (2015). Speaker's Reference and Cross-Cultural Semantics. In (pp. 62-76).

Molyneux, B. (2014). New Arguments that Philosophers don't Treat Intuitions as Evidence. *Metaphilosophy, 45*(3), 441-461.

Nado, J. (2015). Intuition, philosophical theorizing, and the threat of skepticism. In *Experimental Philosophy, Rationalism, and Naturalism* (pp. 212-229): Routledge.

Nagel, J. (2012a). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research, 85*(3), 495-527.

Nagel, J. (2012b). Mindreading in Gettier Cases and Skeptical Pressure Cases. In J. Brown & M. Gerken (Eds.), *Knowledge Ascriptions*: Oxford University Press.

Nagel, J., San Juan, V., & Mar, R. A. (2013). Lay Denial of Knowledge for Justified True Beliefs. *Cognition, 129*(3), 652-661.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme, 17*(2), 141-161.

Oreskes, N., & Conway, E. M. (2011). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*: Bloomsbury Publishing USA.

Overgaard, S., Gilbert, P., & Burwood, S. (2011). *An introduction to metaphilosophy*: Cambridge University Press.

Parfit, D. (1987). *Reasons and persons*: Oxford University Press.

Plantinga, A. (1993). *Warrant: The current debate*: Oxford University Press, USA.

Schaffer, J., & Knobe, J. (2012). Contrastive knowledge surveyed. *Nous, 46*(4), 675-708.

Schaffner, B., & Luks, S. (2017). This is what Trump voters said when asked to compare his inauguration crowd with Obama's. *Washington Post, 25*.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & language, 27*(2), 135-153.

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

Seyedsayamdost, H. (2015b). ON NORMATIVITY AND EPISTEMIC INTUITIONS: FAILURE OF REPLICATION. *Episteme, 12*(1), 95-116. doi:10.1017/epi.2014.27

Singer, P. (2005). Ethics and Intuitions. *The Journal of Ethics, 9*(3/4), 331-352. Retrieved from http://www.jstor.org/stable/25115831

Sosa, E. (2007). *A Virtue Epistemology: Volume I: Apt Belief and Reflective Knowledge*: Oxford University Press UK.

Sosa, E. (2010). Intuitions and meaning divergence. *Philosophical Psychology, 23*(4), 419-426. doi:10.1080/09515089.2010.505859

Stich, S., & Machery, E. (2023). Demographic Differences in Philosophical Intuition: a Reply to Joshua Knobe. *Review of Philosophy and Psychology, 14*(2), 401-434.

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

Sytsma, J., Livengood, J., Sato, R., & Oguchi, M. (2015). Reference in the Land of the Rising Sun: A Cross-cultural Study on the Reference of Proper Names. *Review of Philosophy and Psychology, 6*(2), 213-230. doi:10.1007/s13164-014-0206-3

Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal, 94*(6), 1395-1415. doi:10.2307/796133

Turner, J., Nahmias, E., Morris, S., & Nadelhoffer, T. (2007). Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research, 73*(1), 28-53.

Turri, J. (2013). A conspicuous art: putting Gettier to the test. *Turri, J.(2013). A conspicuous art: putting Gettier to the test. Philosophers' Imprint, 13*(10), 1-16.

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

Weinberg, J. M. (2007). How to challenge intuitions empirically without risking skepticism. *Midwest Studies in Philosophy, 31*(1), 318–343.

Weinberg, J. M., Crowley, S., Gonnerman, C., Vandewalker, I., & Swain, S. (2012). Intuition & calibration. *Essays in Philosophy, 13*(1), 15.

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29*(1-2), 429-460.

Wiech, K., Kahane, G., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2013). Cold or calculating? Reduced activity in the subgenual cingulate cortex reflects decreased emotional aversion to harming in counterintuitive utilitarian judgment. *Cognition, 126*(3), 364-372. doi:https://doi.org/10.1016/j.cognition.2012.11.002

Williamson, T. (2007). *The Philosophy of Philosophy*: Wiley-Blackwell.

Williamson, T. (2011). Philosophical expertise and the burden of proof. *Metaphilosophy, 42*(3), 215-229.

Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry Into the Nature of Virtue and the Ethical Foundations of Knowledge*: Cambridge University Press.

Ziółkowski, A. (2016). FOLK INTUITIONS AND THE NO-LUCK-THESIS. *Episteme, 13*(3), 343-358. doi:10.1017/epi.2015.49

Ziółkowski, A. (2021). The Stability of Philosophical Intuitions: Failed Replications of Swain et al. *Episteme, 18*(2), 328-346.

Ziółkowski, A., Wiegmann, A., Horvath, J., & Machery, E. (2023). Truetemp cooled down: The stability of Truetemp intuitions. *Synthese, 201*(3), 1-19.