Why Variation Matters to Philosophy

Edouard Machery

Abstract: Experimental philosophers often seem to ignore or downplay the significance of demographic variation in philosophically relevant judgments. This article confirms this impression, discusses why demographic research is overlooked in experimental philosophy, and argues that variation is philosophically significant.

Following the groundbreaking research by Jonathan Weinberg, Shaun Nichols, and Steve Stich (2001), part of experimental philosophy has endeavored to assess how much philosophical judgments vary across cultures, languages, religions, generations, age groups, genders, or socioeconomic groups (a research tradition that I will call "comparative experimental philosophy").[1] Recent large-scale projects such as the Geography of Philosophy Project have extended comparative experimental philosophy in a more systematic direction (Kiper et al. 2022). This project, generously funded by the John Templeton Foundation, has been examining since 2017 how people understand three concepts of philosophical interest, viz. the concepts of knowledge, understanding, and wisdom, across cultures, religions, and languages using a variety of empirical methods: qualitative interviews, computational text analyses of large linguistic corpora in many languages, experimental studies using verbal and visual stimuli, and experimental paradigms drawn from experimental economics.[2] The results of the Geography of Philosophy Project and, more broadly, of the research tradition started by Weinberg et al. (2001) are naturally relevant for psychology and anthropology,[3] but their philosophical relevance might not seem obvious: why does it matter for philosophy if judgments of philosophical interest such as those about beauty, fairness, and love vary? The goal of this article is to offer a response to this question.

Here is how I will proceed. In Section 1, I examine why the study of variation has been relatively neglected: I examine various plausible causes, and I also assess whether they justify this neglect. In Section 2, I review the

[1] This goal is sometimes put in terms of concepts instead of, or in addition to, judgments, the goal being then to assess how much concepts vary.

[2] More information can be found on the project's webpage: www.geographyofphilosophy.com and on its YouTube's channel: www.youtube.com/@geographyofphilosophyproje9275.

[3] That said, anthropologists might have some concerns with the meaningfulness of asking philosophical questions to lay people across cultures (see, e.g., Clark Barrett in the following video: www.youtube.com/watch?v=xxMFDZ__MXY).

Res Philosophica, *Vol. 100, No. 1, January 2023, pp. 1–22*
https://doi.org/10.11612/resphil.2264

evidence suggesting that judgments of philosophical interest vary across and, equally important, but more neglected, within populations.[4] In Section 3, I answer the question about the philosophical significance of this variation. In Section 4, I examine two responses.

1 Neglect of Cross-Cultural Research in Experimental Philosophy

There is no agreement about how much judgments of philosophical interest vary across populations such as cultures and generations. Unexpectedly, the controversy has taken place within experimental philosophy rather than between experimental philosophers and their critics (Knobe 2019; Stich and Machery Forthcoming; Weinberg and Alexander Forthcoming). In particular, Joshua Knobe has argued that research has found surprisingly little variation (2019, 31):

> Work in experimental philosophy is often concerned with 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. Yet actual research on this topic has yielded a surprising result. Again and again, studies find that effects observed within one demographic group can also be found in a variety of others.

Stich and Machery disagree (Forthcoming):

> We think Knobe's account is seriously mistaken, and that it is based on a radically misleading portrait of recent work in experimental philosophy and cultural psychology. We are concerned that Knobe's inaccurate account of the literature may have a negative impact on the sort of research that is done in experimental philosophy, and that this may impede attempts to address the cultural insularity that characterizes much of recent philosophy in the analytic tradition.

This controversy (which I will call the "invariance controversy") takes place against the backdrop of a seriously incomplete empirical record. Before examining how the empirical record bears on the invariance controversy despite its limitations, I consider the reasons why research on variation in experimental philosophy is so limited. Experimental philosophers rarely sample from different populations, and typically extrapolate on the basis of convenience samples from online pools of participants examined in a

[4] I use "population" in its statistical sense (the idealized group of human beings from which a given sample is drawn from, however this group is individuated) instead of in its commonsensical sense (roughly, groups of human beings living in different countries and having different cultures). In the statistical sense, men and women can be two distinct populations.

single language (Machery et al. 2021). To provide evidence for this point, I randomly selected 10% of the articles (36 articles including 88 studies) on the list of experimental-philosophy studies developed by Stuart et al. (2019), which includes all the original experimental philosophy studies we could locate until 2017 (available at https://osf.io/2z87f/).[5] I coded whether the samples in each of the 88 studies were intentionally drawn from different populations (e.g., different countries, different religious groups, different genders, different age groups, etc.; Y/N/unspecified), and for what purpose (i.e., which variation was hypothesized), whether they were made of students (Y/N/unspecified), whether they were drawn from online pools of participants (Y/N/unspecified), and whether they were a community sample (Y/N/unspecified). Figure 1 reports the distribution of the 36 articles per year.

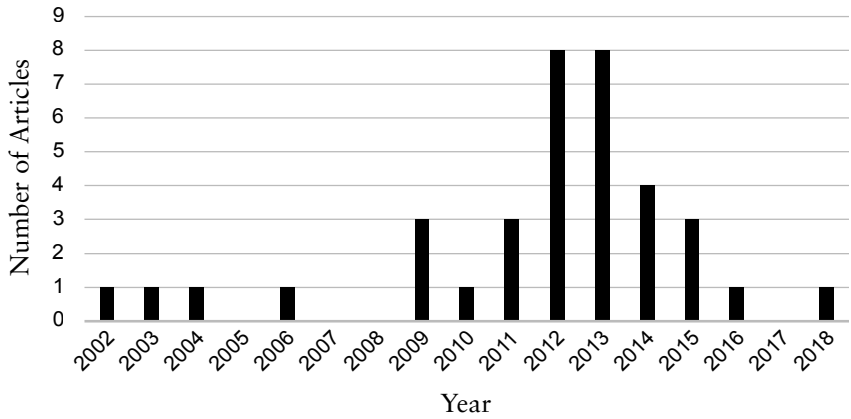

FIGURE 1. Distribution Per Year of the 36 Articles Sampled

As can be seen in Table 1 (first column), few studies aim at examining whether the results reported vary across some populations.

	Intentional Study of Variation	Student Sample	Online Participants	Community Samples
Percentage of Yes	10.2%	30.7%	43.2%	13.6%

TABLE 1. Goal of the 88 Studies and Characteristics of their Samples

Most studies also rely either on samples of students or of online participants (Table 1, columns 2–4 and Figure 2).

[5] The spreadsheet file for the analysis reported in this article can be found here: https://osf.io/r6dk5/.

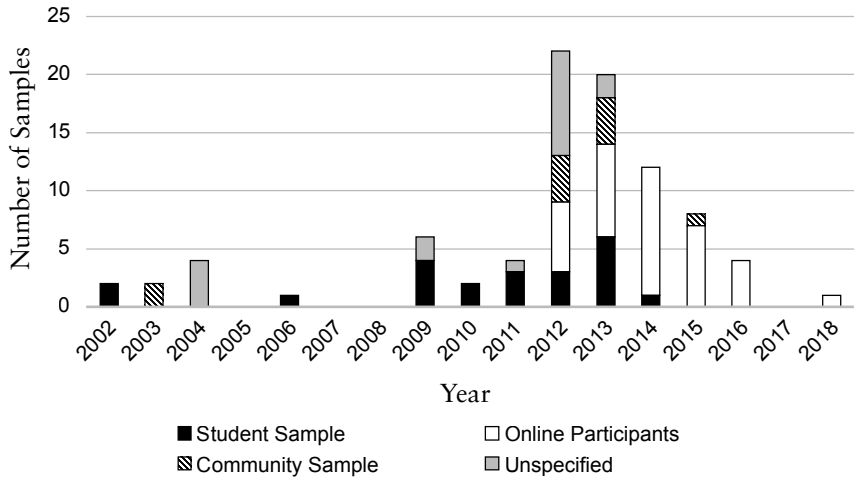

FIGURE 2. Distribution of Sample Types Per Year for the 88 Studies in the Sample

Among the 88 studies that were examined, all but 7 (92.0%) appear[6] to have examined English speakers (although some of them might not have been native speakers[7]), all but 10 (88.6%) participants located in the USA (although not all of them were Americans[8]). Of the 9 articles that set to examine variation (Figure 3), only one examined the role of culture and language (although the participants were students in the USA); two examined the influence of personality, one the influence of disgust sensitivity, five the influence of philosophical training, and one the influence of various memory disorders. While this was not their research focus, a few articles also examined whether some demographic factors (typically gender) influence participants' answers, usually to report finding no effect. Finally, few studies describe their sample of participants with much detail. In some studies, the descriptions are incomplete, and in a few studies no detail is provided (Figure 2). This careless attitude toward the demographic characteristics of their samples also reveals experimental philosophers' assumption that little variation is to expected across groups.

This study confirms that experimental philosophers rarely examine whether their results vary across populations. Why don't they? Most obviously, collecting data across cultures, languages, ages, generations, and even socioeconomic groups is fraught with unique challenges. Among others, stimuli and measures must be meaningful and understood similarly

[6] "Appear" because some studies do not describe their samples at all or describe them in an incomplete manner.

[7] Online pools of participants were used and most papers do not clarify whether the participants were limited to those with a USA location or to native speakers of English.

[8] Seven studies were done in France, and three were done in Australia.

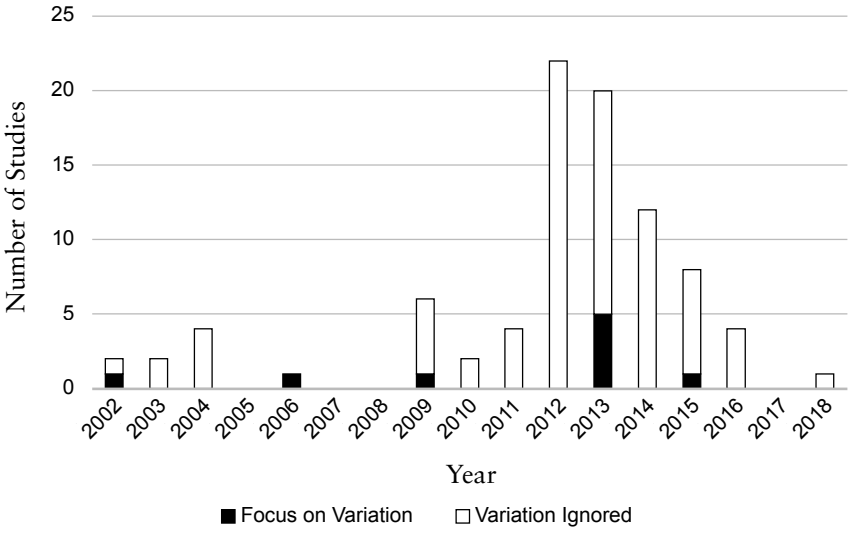

FIGURE 3. Distribution of Studies Examining Variation Per Year for the 88 Studies in the Sample

across all groups (e.g., across cultures), verbal stimuli must be accurately translated when several languages are involved, and participants can be hard to find when they are not students in Western societies or when they are not drawn from online pools of participants such as Amazon Turk or Prolific. Doing comparative experimental philosophy is thus slow and expensive, which explains in part why the empirical record is so incomplete. There are technical solutions to some of these challenges. Translations can be backtranslated,[9] the comparability of measures across groups can be established by examining their measurement invariance, and controls can be added to assess whether stimuli are understood similarly. Some of these solutions are simple and have been widely used by experimental philosophers (e.g., the use of control questions); others are simple, but have been less frequently used (e.g., backtranslation); yet others are technically challenging. Measurement invariance, for instance, has only recently become a well-recognized concern in psychology (although the topic has a longer history); establishing measurement invariance requires some statistical sophistication; and its significance has been, and remains, contested.

Despite experimentalists' best efforts, some measure of uncertainty is also bound to color the interpretation of variation across populations, which might have deterred experimental philosophers from conducting this kind of

[9] Backtranslation occurs when a translation into a language L2 (e.g., French) from a language L1 (e.g., English) is translated from L2 to L1 in order to assess whether the original translation was accurate.

research. There is often room to suspect that differences across populations are due to some of the challenges mentioned above. Critics of some of the most well-known findings in experimental philosophy have done just that. Sosa speculated that the cultural differences reported in experimental philosophy could be due to how participants fill in the scenarios they are asked to read (2009, 107):

> When we read fiction we import a great deal that is not explicit in the text. We import a lot that is normally presupposed about the physical and social structure of the situation as we follow the author's lead in our own imaginative construction. And the same seems plausibly true about the hypothetical cases presented to our [Weinberg, Nichols, and Stich's] subjects. Given that these subjects are sufficiently different culturally and socio-economically, they may because of this import different assumptions as they follow in their own imaginative construction the lead of the author of the examples, and this may result in their filling the crucial C [i.e., the epistemic situation of the character of a scenario] differently. Perhaps, for example, subjects who differ enough culturally or socio-economically will import different background beliefs as to the trustworthiness of American corporations or zoos, or different background assumptions about how likely it is that an American who has long owned an American car will continue to own a car and indeed an American car. For some if not all of the examples, I can't myself feel sure that C stays constant across the cultural or socio-economic divide. But if C varies across the divide, then the subjects may not after all disagree about the very same content.

Critics of the cross-cultural results reported by Machery et al. (2004) have also appealed to differences in how participants understand the question they are asked to answer, raising the possibility that people who appear to answer differently might in fact understand this question differently. Deutsch (2009, 455), for instance, writes the following about Mallon et al. 2009:

> The fact that the vignette question can be interpreted as either (Q1), which asks for the speaker's reference of John's uses of "Gödel," or (Q2), which asks for the semantic reference of those uses, casts severe doubt on Mallon et al.'s claim that the polls' results show that there are cross-cultural differences in referential intuitions. Given the ambiguity of the vignette question, it may be that some of their respondents were answering (Q1), while some were answering (Q2). If so, Mallon et al. cannot claim that

> their results show that Western and East Asian intuitions about the Gödel case conflict. They have no right, even, to another claim of theirs, which is that significant minorities in the Western and East Asian groups have intuitions that conflict with the majorities in those groups.

Of course, this kind of challenge to the interpretation of variation in experimental studies is not limited to experimental philosophy, and cross-cultural work in the behavioral sciences is sometimes criticized in similar terms. Henrich et al. (2005) famously reported that people behave differently in behavioral economics games such as the dictator and ultimatum games, but critics responded that this variation is possibly due to different populations interpreting the experimental situation through different local frames (e.g., Cronk 2007).

Difficulties are genuine, and they can understandably discourage experimental philosophers to engage in comparative experimental philosophy. The Geography of Philosophy Project has tried to address some of the challenges involved in studying variation across populations. Participants were recruited from Peru, Ecuador, Morocco, South Africa, India, China, Japan, South Korea, Slovakia, Ukraine, Russia, Croatia, the USA, Canada, and Germany; in many of these countries, the studies were run in different languages; we recruited participants from large-scale and small-scale societies; and we did our best to recruit from student and non-student populations. Verbal materials were translated and backtranslated; the cultural appropriateness of the stimuli and measures was qualitatively assessed; the measurement invariance of some of the measures used is under examination. However, it has turned out to be difficult to address all the challenges to our full satisfaction. Among other challenges, despite our best efforts to sample from non-student populations, many of our samples involve students, because those are the easiest to obtain for academics in non-Western countries as they were in the USA before online pool of participants became widely available.

Technical difficulties aren't the only source of experimental philosophers' neglect of the possible variation of their experimental results. Experimental philosophers, I suspect, believe that philosophically relevant empirical findings are unlikely to vary substantially. Such a belief might have been encouraged by the fact that prominent results, including some reported by Weinberg et al. (2001), have failed to replicate (Seyedsayamdost 2015; Kim and Yuan 2015), although some of their results appear to be robust (Sękowski et al. Forthcoming). The systematic replication audit of experimental philosophy led by Florian Cova found that overall experimental philosophy replicates well, but that studies focusing on variation are less likely to replicate than the rest of experimental philosophy (Cova et al. 2021, 29):

> For our sample at least, it does appear that content-based studies have a higher replication rate when compared to context-based and demographic-based studies.

However, Cova et al. examined only four studies with a comparative focus: Machery et al. 2004, Nadelhoffer et al. 2009, Grau and Pury 2014, and Sytsma and Machery 2010 (the first three failed to replicate). This is a small sample, and it would be irresponsible to generalize on its basis. What's more, the three failed replications fail to show that comparative experimental philosophy is less likely to replicate. Nadelhoffer et al. 2009 was itself a failed replication of an earlier finding reporting that extraversion predicts people's attitude toward the relation between determinism and free will (Feltz and Cokely 2009). Thus, the failed replication of Nadelhoffer et al. 2009 in fact confirms the variation reported by Feltz and Cokely instead of showing that there is no variation. The other two failed replications have been undermined by follow-up work. The failed replication of Machery et al. 2004 by Van Dongen et al. (https://osf.io/qdekc/) appears to be an outlier (see below on Machery et al. 2004), and Van Dongen et al. themselves were then able to replicate Machery et al. 2004. Grau and Pury (2014) presented evidence that judgments about reference and judgments about love are correlated in a predicted manner: people who make Kripkean judgments about the reference of proper names tend to believe that when one genuinely loves someone, this love is independent of what the object of love is like and would persevere if the object of love were to change. The failed replication of Grau and Pury 2014 fails to challenge this finding since its power is very low, and since the original effect is robust in sufficiently powered studies (Machery et al. 2020).

Other papers have found a surprising amount of convergence across populations, which has perhaps led some experimental philosophers to expect more convergence than variation and has perhaps discouraged them from examining possible variation. For instance, Machery and colleagues (2017a, 2017b) report that people in more than 15 countries share the so-called Gettier intuition at least for some ways of asking people to assign knowledge: people distinguish having a justified true belief in a proposition and knowing it.

It is, however, unclear whether experimental philosophers can expect a priori convergence or variation across populations. Many psychological and behavioral phenomena that were originally observed with Western participants have turned out not to be universal (Henrich et al. 2010). Similarly, some phenomena in experimental philosophy that originally appeared to be universal seem to vary across populations. A good example is the well-known side-effect effect (Knobe 2003): the intentional nature of a foreseen side effect depends on its valence. While the effect has been replicated many times with speakers of languages other than English (e.g., Hindi; Burra and Knobe 2006), in two rural cultures (Samoa and Vanuatu)

people were more likely to judge the good side effect than the bad side effect to be intentional if the protagonist had a high status (Robbins et al. 2017).

Finally, some experimental philosophers might believe that the phenomena they are after are likely to be characteristic of human nature and possibly be innate (Knobe 2019, 33; Phillips et al. 2021), and thus that examining whether these findings generalize across populations isn't a serious concern. The evidence for these claims is often underwhelming. In this vein, Phillips et al. have argued that the capacity to ascribe knowledge is a "basic" capacity (their term) that is shared by apes and human beings. To their credit, they review some important comparative work supporting this claim, going beyond what is usually done by experimental philosophers. The non-comparative literature they also review consists either of developmental studies or of linguistic studies of knowledge ascription. Most of the studies reporting linguistic data that Phillips et al. consider were conducted in English with American participants, one of more than 6,500 languages currently spoken (Machery et al. 2021). The problem here is that it isn't clear that the primate studies, the infant studies, and the linguistic studies are all about the same construct: it isn't clear that what primate studies operationalize as knowledge ascription (and similarly what developmental studies operationalize as knowledge ascription) corresponds to what people mean when they use "to know that" in English and, a fortiori, to its standard translations in the thousands of languages ignored by Phillips and colleagues. So, it is unclear what should be expected about concepts expressed by "knowledge" and its standard translations on the basis of the comparative and developmental work.

In summary, comparative experimental philosophy is undoubtedly challenging, and it is difficult to address all these challenges even with the best intentions. However, factors other than the difficulty of comparative experimental philosophy have probably contributed to experimental philosophers' failure to take into account the possible variation of their findings, although neither replicability concerns nor findings of invariance across some populations justify this failure. To conclude this discussion, it is worth emphasizing what is ultimately at stake: is it enough for experimental philosophers to rely on convenience samples of students or online participants? Is "Mturk experimental philosophy" good enough, or is comparative experimental philosophy necessary?

2 Resolving the Invariance Controversy

Where does the evidence stand? While some results in comparative experimental philosophy have not been robust and while some phenomena do not appear to vary across the populations that have been examined (e.g., Sarkissian et al. 2010, Machery et al. 2017a, 2017b), some striking

examples of variation are extremely robust. Machery et al. (2004) examined whether East Asians and Americans make similar judgments about the reference of proper names in response to Gödel cases (Kripke 1980, 83–84) and Jonah cases (Kripke 1980, 67). In a Gödel case, a proper name is associated by a speaker or a community of speakers with a description that is not satisfied by the original bearer of this name, but by someone else. Thus, Kripke's Gödel case reads as follows (1980, 83–84):

> Suppose that Gödel was not in fact the author of [Gödel's] theorem. A man called "Schmidt" . . . actually did the work in question. His friend Gödel somehow got hold of the manuscript and it was thereafter attributed to Gödel. On the [descriptivist] view in question, then, when our ordinary man uses the name 'Gödel,' he really means to refer to Schmidt, because Schmidt is the unique person satisfying the description "the man who discovered the incompleteness of arithmetic." . . . But it seems we are not. We simply are not.

In a Jonah case, a proper name is associated by a speaker or a community of speakers with a description that is not satisfied by anyone, including by the original bearer of this name. Data were collected in Hong Kong and in the USA, using vignettes in English. As predicted, Chinese people were more likely than Americans to make descriptivist judgments (i.e., judgments in line with the descriptivist theory of reference) in response to the Gödel case; in fact, a majority of Chinese made descriptivist judgments, while a majority of Americans made causal-historical judgments (i.e., judgments in line with Kripke's causal-historical theory of reference). Figure 4 shows that this finding has been replicated many times, including with children (Li et al. 2018). In addition, van Dongen et al. (2021) meta-analyzed the literature on judgments about reference inspired by Machery et al. 2004 and concluded as follows (763):

> [O]ur meta-analysis supports the hypothesis that cross-cultural factors affect semantic intuitions about proper names. . . . Neither do specific analysis tools aimed at detecting publication bias or QRPs (e.g., funnel plots, p-curves) provide evidence of systematic suppression of negative results.

Literature reviews show that variation is not limited to a few phenomena. Chapter 2 of *Philosophy Within Its Proper Bounds* reviewed a large part of the relevant experimental-philosophical literature, and found much evidence for variation. I will not repeat this literature review here (Machery et al. 2017a). Alfano et al. (2022) further review additional evidence for variation of moral judgment across populations in their revised entry on experimental moral psychology. To single out a few highlights, men

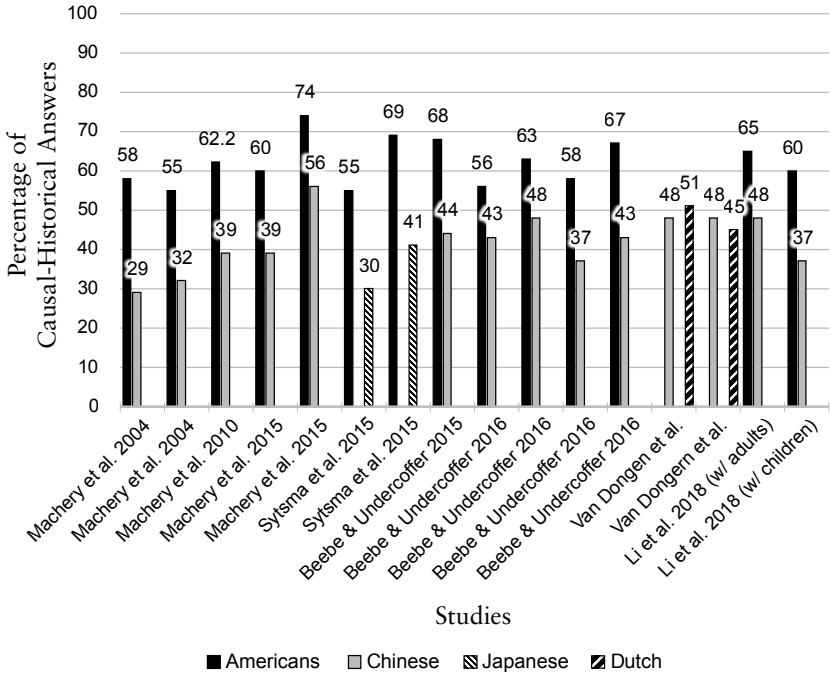

FIGURE 4. Proportion of Causal-Historical Answers in Response to a Gödel Case

and women respond differently to moral dilemmas that investigate the permissibility of causing harm to prevent a greater harm (e.g., trolley cases). In a meta-analysis of 40 studies, men were more likely to find it permissible to cause harm to prevent a greater harm (Friesdorf et al. 2015). Hannikainen et al. (2018) have shown that Millennials are more likely to judge it permissible to act in the footbridge case than Gen Xers and Boomers. Judgments related to free will, control, blame, and punishment also vary across cultures (Hannikainen et al. 2019). In most cultures, people deny free will and control (and thus blame and punish less) when the agent's action is described as antecedently caused; by contrast, they assign free will and control when the action originates from the agent's own will even if she could not have done otherwise (a situation illustrated by Frankfurt cases). East Asians, however, differ in treating these two kinds of situations similarly; if the surrounding circumstances undercut the agent's capacity to have done otherwise, they tend to deny her free will and control.

Most telling of all, Stich and Machery (Forthcoming) have put together a list of 100 studies by 205 different researchers comprising a total sample size of over 40 million participants showing that philosophical judgments vary across some group or other. These 100 studies cover a large range of

philosophical topics, including morality, epistemology, semantics, and free will. They also provide evidence for many different types of demographic variation: culture, gender, philosophical training, and other variables influence judgments of philosophical interest. A finding that is of particular metaphilosophical significance (Machery 2017, 90–125) is the finding of divergence between philosophers and non-philosophers. A case at hand is the idea that knowledge ascription is influenced by how costly it would be to have a false belief (what philosophers usually call "the stakes"): the higher the stakes, the higher the standards the ascribee must meet to be properly said to have knowledge. While many professional epistemologists have claimed that knowledge ascription is sensitive to the stakes, Rose et al. (2019) found no evidence whatsoever of a stakes effect on knowledge attribution in 3530 participants from 19 countries on 5 continents, speaking 15 languages.

Knobe has insisted that for all the evidence of variation it remains *surprising* that some of the effects observed by experimental philosophers appear not to be influenced by demographic factors. As Weinberg and Alexander (Forthcoming) have however argued, it is not clear how to interpret Knobe's claim about how surprising the lack of demographic influence is; what's more, plausible interpretations of this claim do not justify it. I agree with Weinberg and Alexander's discussion, but my concern here is different. Knobe does not tell us why surprisingness matters. While I do not have space to elaborate in this article, suffice it to say that the relation between surprisingness and scientific importance is far from straightforward. Scientific communities reward original findings, plausibly in order to incentivize scientists to explore widely the space of scientific possibilities, but the recent replication crisis has shown that this incentive can have perverse side effects. In any case, that original research is incentivized to maximize the prospect of discovery does not mean that original findings are scientifically more important. In a Bayesian context, surprising findings have greater evidential value, but assessing the surprisingness of evidence (P(E|B)) objectively is difficult given its dependence on the catch-all hypothesis (P(E|~T&B)), and overall surprisingness does not matter when two hypotheses are compared (Salmon 1990). Finally, historically, surprisingness and scientific importance differ (Stigler 1955; Johnson et al. 2019).

Before moving on to the philosophical significance of the variation in judgments of philosophical interest, I would like to highlight another form of variation that has been largely overlooked: heterogeneity. As I use the term here, "heterogeneity" refers to the fact that the impact of a manipulated cause on a dependent variable varies in size and even direction because of differences among experimental units (e.g., experimental participants) (Bolger et al. 2019). The difference between heterogeneity and the variation discussed earlier (across cultures, genders, etc.) is that the latter takes place between populations that are distinguished by well-established variables

(e.g., gender), the former within a given population. Traditional statistical analysis in the behavioral sciences (e.g., by means of ANOVA) typically examine only the impact on manipulations on means, and relegate differences among people to the error term, but behavioral scientists have recently grown more sensitive to the fact that manipulations affect participants to a different degree and that such differences needed to be modeled statistically either in a frequentist or a Bayesian context (Gelman 2015; Bolger et al. 2019).

Experimental philosophers have followed the traditional practices in the behavioral sciences and rarely analyze and comment on the diversity of answers among their participants. Diversity is however present in all experimental-philosophy studies. In the research spurred by Machery et al. (2004), on average 40% of American participants make descriptivist judgments, and 35% of Asian participants make causal-historical judgments. Myers-Schulz and Schwitzgebel (2013) report that there is substantial disagreement regarding the "entailment thesis," according to which knowing that p entails believing that p, with a substantial percentage appearing to endorse it and a substantial percentage appearing to reject it. Even the side-effect effect is not found among all Western participants. No doubt, part of this variation is due to measurement error: people answering randomly, misreporting their real opinion, misunderstanding the question, etc. But part of it is probably due to genuine heterogeneity.

From a philosophical point of view, heterogeneity would seem to matter as much as variation across established populations, as Sytsma and Livengood (2011) noticed in their influential discussion of Machery et al. 2004. What is philosophically relevant (more on why in the remainder of this paper) is the fact of variation (provided it is not due to mere measurement error), not whether it is predicted by some well-established variable or other. It is surprising that Knobe overlooks heterogeneity in his discussion of the demographic effects that have been central to experimental philosophy. For one thing, while admittedly heterogeneity is not incompatible with the claim that some of experimental philosophers' findings might be innate (not every human being must possess an "innate" trait, as illustrated by blindness and deafness), it calls this claim into question (whatever "innate" means, see Griffiths and Machery 2008).

3 Why Variation Matters to Philosophy

Let's assume that many judgments of philosophical interest vary across and within populations. Why would it matter to philosophy? Machery (Machery 2017, 90–125) develops an answer in detail. Here I want to highlight the gist of this answer, which the details could overshadow.

There are broadly two different ways of looking at variation in philosophical judgments: people could genuinely disagree with one another, or they could in fact be speaking about different, perhaps subtly different,

things.[10] Both situations are familiar. When a speaker says, "Cats are wonderful," and her interlocutor replies, "Cats are awful," the two interlocutors typically assume that both are talking about the same animals, but disagree about them if there is enough overlap between what both are willing to say using "cat." When the overlap is more limited, particularly for assertions that one of them takes to be important for what the referents of a given word are (however importance is understood), interlocutors could assume that they are talking about different things. This latter situation is easy to deal with for ambiguous words such as "bank" (i.e., words whose meanings are unrelated) or for polysemous words (i.e., words whose meanings are related) whose plurality of meanings is conventionalized. Navigating such a situation is more difficult for words whose plurality of meaning is less obvious, but we do sometimes conclude that our interlocutors are not using a given word to refer to exactly the same things as we do. This is the case, among others, for words whose extensions are partly unspecified and evolving, for instance words for musical subgenres: two interlocutors could appear to disagree about whether some song is a timba song, and they could come to realize that they do not really refer to the same class of songs with the word "timba." The disagreement would then be purely verbal.

The same distinction applies to philosophical disagreements across and within populations. Two interlocutors might appear to disagree about a philosophical matter, one of them saying, say, "justice . . . ," the other one denying her assertion. The assertions could be about an abstract matter (e.g., "stealing is unjust") or about a concrete action ("what this character did in this story is unjust"). They could be expressed in the same language (e.g., in English) or one could involve the standard translation of the relevant words (e.g., "la justice . . ."). In all these cases, the two interlocutors could either genuinely disagree or merely appear to disagree, while talking about something different, perhaps slightly different (justice and justice*).

One might think that two people who appear to disagree about a philosophical matter while sharing a common language (e.g., both could be speakers of English) are more likely to disagree genuinely than verbally, but the example of musical subgenres above suggests that verbal disagreements do happen among speakers of the same language: a given language often includes many dialects, and of course speakers have their own idiolects. The extension of many words within a given language is contested and evolving; this is also true of words expressing concepts of philosophical interest such as "cause" (Macleod 2019). On the other hand, while words and their standard translations can have a somewhat different extension

[10] This distinction does not require a distinction between analytic and synthetic truths (Machery 2017, 30–35).

(e.g., "frasca" in Spanish and "jar" in English, see Ameel et al. 2009), they need not.

In either case, apparent disagreement constitutes a form of second-order evidence (e.g., Feldman 2005; Christensen 2010): not evidence that bears on the truth of what I believe, but evidence that bears on how well my belief was formed. The significance of this second form of evidence varies depending on how we interpret cases of apparent disagreement. If people genuinely disagree, the second-order evidence can challenge their *justification* for holding the beliefs they hold. The philosophical literature on second-order evidence usually discusses this situation. If people's disagreement is merely verbal, the second-order evidence does not challenge the justification for holding the beliefs we hold; it challenges the *value* of holding these beliefs. It raises the question of why we are having beliefs about a given topic (rather than another).

Let's elaborate on each point in turn, beginning with the case of genuine disagreement. Disagreement of course does not in itself challenge the justification one might have for one's belief since the disagreeing parties could simply have different bodies of evidence. However, when they have access to the same evidence and are equally good reasoners, disagreement seems to be telling us something about the way we have formed our beliefs, although epistemologists disagree about how drastically this kind of disagreement should shake our confidence. When we find out that a given topic elicits a *massive* disagreement among people equipped with the same evidence and the same reasoning capacities (e.g., think about the disagreements about wine quality among experts or about the quality of journal submissions during peer review), most epistemologists concur that we should become uncertain about our beliefs, arguably because we learn that we are not very good at reasoning about the relevant subject matters or that the evidence is too impoverished.

Apparent disagreements of course do not bear on our confidence in the same way since the disagreeing parties aren't talking about the same thing. Instead of raising questions about the way we have formed our beliefs—an epistemic matter—they raise questions about why we have formed our beliefs about this rather than about that—a practical matter: why should we care about theorizing about this rather than that, say about justice rather than justice*—viz. what the party that appears to be disagreeing with me when she makes an assertion about justice happens to be talking about? One might object we don't usually wonder whether we should be using the word as our interlocutor does, when we discover than we use a given word (e.g., "timba") somewhat differently from she does. But the reason is that nothing is at stake in those cases. It often does not matter how, e.g., musical subgenres are distinguished from one another. By contrast, when something is at stake in how distinctions are drawn, and when a linguistic issue is contested, we do care about how words are to be used. The Supreme Court in the USA debated "causation"

(Macleod 2019) and "house" (Totenberg 2013[11]); people angrily disagree about how to use "woman," etc. In philosophy, it could matter whether we are theorizing about, say, justice or justice* because theorizing about justice*, not justice, might be key to answering the theoretical and practical questions we are ultimately interested in. Discovering that others form beliefs about properties that though related are distinct from what we form beliefs about, should lead philosophers to worry about the value of theorizing about what they are currently theorizing about.

So, variation in judgments of philosophical interest suggests either that we should be much less certain about what we believe or that we might not be thinking about what really matters. Comparative experimental philosophy is thus an exercise in philosophical humility. It can teach us that we are overconfident or that we might be chasing wrong leads. What's more, it might help us address these problems. It might lead us to calibrate properly our confidence, or it might reveal new topics for our philosophical theorizing, similar to, but at the same time importantly different from what we have already been theorizing about. With respect to the last point, we might discover that our focus might be slightly, or less than slightly, off.

4 Two Objections

Some philosophers have, however, argued that while perhaps interesting for anthropology or psychology the empirical fact of variation is of no philosophical significance for philosophy (e.g., Cappelen 2012; Deutsch 2015). In this final section, I reply to two objections.

The first objection consists in denying that the judgments made by philosophers and their disagreeing interlocutors have the same epistemic standing. Exactly as disagreement carries no epistemic weight when one of the disagreeing parties is in a better epistemic position, either because she has more evidence or because she uses the available evidence more aptly, philosophers should not revise their beliefs when they discover that others (genuinely) disagree with them because they are in a better epistemic standing than the disagreeing parties. If disagreement is merely verbal, philosophers can insist that *their* concerns are more likely to be the relevant ones that those of the disagreeing parties, exactly as a psychiatrist's concerns about the source of a neurosis are more likely to be on target than her patient's. Objections of this ilk are known as the expertise defense in the metaphilosophical literature (e.g., Williamson 2011; Machery 2015, Machery 2017, 158–169; Nado 2014, 2015; Schindler and Saint-Germier Forthcoming). It goes without saying that philosophers have genuine expertise in some domains: among others, they know more about the history of philosophy, and they have acquired distinct skills, from close, accurate reading to clarity and precision in argumentation. On the other

[11] I owe this example to Dejan Makovec.

hand, philosophers appear to suffer from the biases that impact lay people's judgments (Machery 2012; Horvath and Wiegmann 2016, Forthcoming; Wiegmann et al. 2020): order effects, framing effects, etc., influence some of their judgments (particularly, the judgments about thought experiments) to the same extent as lay people's.

The first issue then is to determine the bounds of philosophical expertise (Egler and Ross 2020). The issue is empirical, and while some aspects of the issue do not call for empirical studies (e.g., we don't need evidence that most philosophers know more about the history of philosophy than non-philosophers), others do. For over ten years experimental philosophers have investigated the matter, but their work remains largely limited to judgments about thought experiments (but see Livengood et al. 2010; Byrd 2021, Forthcoming). The second issue is to determine which disagreement can be ignored in light of the confirmed facts of philosophical expertise. Obviously, a disagreement between lay people and philosophers about the history of philosophy would usually carry little epistemic weight. I have argued that disagreements about thought experiments cannot be ignored because philosophers do not appear to have a higher epistemic standing in this respect (Machery 2017).

I now turn to the second objection (Deutsch 2015, 2020; Horvath Forthcoming; for discussion, see Colaço and Machery 2017; Machery 2020). Philosophical judgments, including those about thought experiments, are the conclusions of arguments, and studies in experimental philosophy that report variation are irrelevant. What matters is the strength of the arguments. There are many issues with this position. For one, the claim that the judgments about thought experiments are the conclusions of arguments is underspecified. Its interpretations range from trivially true—the judgments of thought experiments are made inductively on the basis of the information given in the thought experiments or they can be reconstructed as arguments—to trivially false—these judgments are the conclusions of explicit arguments (philosophers often do not present explicitly their judgments as the conclusions of arguments). What's more, exactly as evidence of unreliability (e.g., variation among experts in how good journal submissions are or about how good wine bottles are) should undermine our confidence in our capacity to determine good and bad reasons, evidence of variation shows that we are not very good at distinguishing good and bad reasons. In this respect, describing judgments as products of arguments makes little difference to the point made in the previous section.

A critic could respond as follows. Describing philosophers' judgments about thought experiments as the conclusions of arguments undermines the relevance of experimental philosophers' findings about variation because it is unclear whether the judgments studied by experimental philosophers are genuinely the conclusions of arguments. However, at this point, the vagueness of the appeal to arguments matters. If the judgments about thought experiments count as conclusions of arguments merely because

they are based on the facts stipulated by the thought experiments (e.g., the runaway trolley will kill five people if nothing is done), then there is no reason to doubt that in that sense the usual participants of experimental philosophical studies answer on the basis of arguments. (In fact, they are able to justify their answers, and asking them to do so explicitly makes no difference to their judgments, see Kneer et al. 2021.) If the judgments about thought experiments count as conclusions of arguments because they are explicitly deduced from abstract principles, then there is no reason to believe that philosophers' judgments are conclusions of arguments.

5 Conclusion

Variation in judgments of philosophical interest is real across and within populations; it is largely ignored by traditional philosophers and, more surprisingly, by experimental philosophers; recent arguments highlighting the invariance of philosophical judgments contribute to this pattern of overlooking variation; and variation matters for philosophy. It is time for the philosophical community to take it more seriously.

Edouard Machery
University of Pittsburgh
University of Johannesburg
E-mail: machery@pitt.edu

References:

Alfano, Mark, Edouard Machery, Alexandra Plakias, and Don Loeb. 2022. "Experimental Moral Philosophy." In *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), edited by Edward N. Zalta and Uri Nodelman. https://plato.stanford.edu/archives/fall2022/entries/experimental-moral/.

Ameel, Eef, Barbara C. Malt, Gert Storms, and Fons Van Assche. 2009. "Semantic Convergence in the Bilingual Lexicon." *Journal of Memory and Language* 60 (2): 270–290. https://doi.org/10.1016/j.jml.2008.10.001.

Bolger, Niall, Katherine S. Zee, Maya Rossignac-Milon, and Ran R. Hassin. 2019. "Causal Processes in Psychology Are Heterogeneous." *Journal of Experimental Psychology: General* 148 (4): 601–618. https://doi.org/10.1037/xge0000558.

Burra, Arudra and Joshua Knobe. 2006. "The Folk Concepts of Intention and Intentional Action: A Cross-Cultural Study." *Journal of Cognition and Culture* 6 (1/2): 113–132. https://doi.org/10.1163/156853706776931222.

Byrd, Nick. 2021. "Reflective Reasoning and Philosophy." *Philosophy Compass* 16 (11): e12786. https://doi.org/10.1111/phc3.12786.

Byrd, Nick. Forthcoming. "Great Minds Do Not Think Alike: Philosophers' Views Predicted by Reflection, Education, Personality, and Other Demographic Differences." *Review of Philosophy and Psychology*.

Acknowledgments I would like to thank Helen De Cruz for her invitation to deliver the Wade Memorial Lecture at the 2021 Res Philosophica Conference as well as Helen De Cruz and Eric Schwitzgebel for comments on a previous version of this article. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Cappelen, Herman. 2012. *Philosophy without Intuitions*. Oxford: Oxford University Press.
Christensen, David. 2010. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81 (1): 185–215. https://doi.org/10.1111/j.1933-1592.2010.00366.x.
Colaço, David and Edoaurd Machery. 2017. "The Intuitive Is a Red Herring." *Inquiry* 60 (4): 403–419. https://doi.org/10.1080/0020174X.2016.1220638.
Cova, Florian, Brent Strickland, Angela Abatista, Aurélien Allard, James Andow, Mario Attie, James Beebe, et al. 2021. "Estimating the Reproducibility of Experimental Philosophy." *Review of Philosophy and Psychology* 12 (1): 9–44. https://doi.org/10.1007/s13164-018-0400-9.
Cronk, Lee. 2007. "The Influence of Cultural Framing on Play in the Trust Game: A Maasai Example." *Evolution and Human Behavior* 28 (5): 352–358. https://doi.org/10.1016/j.evolhumbehav.2007.05.006.
Deutsch, Max E. 2009. "Experimental Philosophy and the Theory of Reference." *Mind and Language* 24 (4): 445–466. https://doi.org/10.1111/j.1468-0017.2009.01370.x.
Deutsch, Max E. 2015. *The Myth of the Intuitive: Experimental Philosophy and Philosophical Method*. Cambridge, MA: MIT Press.
Deutsch, Max E. 2020. "The Method of Cases Unbound." *Analysis* 80 (4): 758–771. https://doi.org/10.1093/analys/anaa052.
Egler, Miguel and Lewis D. Ross. 2020. "Philosophical Expertise under the Microscope." *Synthese* 197 (3): 1077–1098. https://doi.org/10.1007/s11229-018-1757-0.
Feldman, Richard. 2005. "Respecting the Evidence." *Philosophical Perspectives* 19: 95–119. https://doi.org/10.1111/j.1520-8583.2005.00055.x.
Feltz, Adam and Edward T. Cokely. 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. https://doi.org/10.1016/j.concog.2008.08.001.
Friesdorf, Rebecca, Paul Conway, and Bertram Gawronski. 2015. "Gender Differences in Responses to Moral Dilemmas: A Process Dissociation Analysis." *Personality and Social Psychology Bulletin* 41 (5): 696–713. https://doi.org/10.1177/0146167215575731.
Gelman, Andrew. 2015. "The Connection between Varying Treatment Effects and the Crisis of Unreplicable Research: A Bayesian Perspective." *Journal of Management* 41 (2): 632–643. https://doi.org/10.1177/0149206314525208.
Grau, Christopher and Cynthia L. Pury. 2014. "Attitudes towards Reference and Replaceability." *Review of Philosophy and Psychology* 5 (2): 155–168. https://doi.org/10.1007/s13164-013-0162-3.
Griffiths, Paul E. and Edouard Machery. 2008. "Innateness, Canalization, and 'Biologicizing the Mind'." *Philosophical Psychology* 21 (3): 397–414. https://doi.org/10.1080/09515080802201146.
Hannikainen, Ivar R., Edouard Machery, and Fiery A. Cushman. 2018. "Is Utilitarian Sacrifice Becoming More Morally Permissible?" *Cognition* 170 (1): 95–101. https://doi.org/10.1016/j.cognition.2017.09.013.
Hannikainen, Ivar R., Edouard Machery, David Rose, Stephen Stich, Christopher Y. Olivola, Paulo Sousa, and Florian Cova, et al. 2019. "For Whom Does Determinism Undermine Moral Responsibility? Surveying the Conditions for Free Will across Cultures." *Frontiers in Psychology* 10: 2428. https://doi.org/10.3389/fpsyg.2019.02428.
Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2/3): 61–83. https://doi.org/10.1017/S0140525X0999152X.
Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, Richard McElreath, et al. 2005. "'Economic Man' in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies." *Behavioral and Brain Sciences* 28 (6): 795–815. https://doi.org/10.1017/S0140525X05000142.
Horvath, Joachim. Forthcoming. "Mischaracterization Reconsidered." *Inquiry*.

Horvath, Joachim and Alex Wiegmann. 2016. "Intuitive Expertise and Intuitions about Knowledge." *Philosophical Studies* 173 (10): 2701–2726. https://doi.org/10.1007/s11098-016-0627-1.

Horvath, Joachim and Alex Wiegmann. Forthcoming. "Intuitive Expertise in Moral Judgments." *Australasian Journal of Philosophy*.

Johnson, Samuel, Amanda Royka, Peter McNally, and Frank C. Keil. 2019. "When Is Science Considered Interesting and Important?" In *Proceedings of the 41st Annual Conference of the Cognitive Science Society*, 1970–1976. Curran Associates.

Kim, Minsun and Yuan Yuan. 2015. "No Cross-Cultural Differences in the Gettier Car Case Intuition: A Replication Study of Weinberg et al. 2001." *Episteme* 12 (3): 355–361. https://doi.org/10.1017/epi.2015.17.

Kiper, Jordan, Stephen Stich, H. Clark Barrett, and Edouard Machery. 2022. "Experimental Philosophy." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo, 61–74. London: Routledge.

Kneer, Markus, David Colaço, Joshua Alexander, and Edouard Machery. 2021. "On Second Thought: Reflections on the Reflection Defense." *Oxford Studies in Experimental Philosophy* 4: 257–296.

Knobe, Joshua. 2003. "Intentional Action and Side Effects in Ordinary Language." *Analysis* 63 (3): 190–194. https://doi.org/10.1093/analys/63.3.190.

Knobe, Joshua. 2019. "Philosophical Intuitions Are Surprisingly Robust across Demographic Differences." *Epistemology and the Philosophy of Science* 56 (2): 29–36. https://doi.org/10.5840/eps201956225.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Li, Jincai, Longen Liu, Elizabeth Chalmers, and Jesse Snedeker. 2018. "What Is in a Name: The Development of Cross-Cultural Differences in Referential Intuitions." *Cognition* 171 (2): 108–111. https://doi.org/10.1016/j.cognition.2017.10.022.

Livengood, Jonathan, Justin Sytsma, Adam Feltz, Richard Scheines, and Edouard Machery. 2010. "Philosophical Temperament." *Philosophical Psychology* 23 (3): 313–330. https://doi.org/10.1080/09515089.2010.490941.

Machery, Edouard. 2012. "Expertise and Intuitions about Reference." *Theoria* 27 (1): 37–54.

Machery, Edouard. 2015. "The Illusion of Expertise." In *Experimental Philosophy, Rationalism and Naturalism: Rethinking Philosophical Method*, edited by Eugen Fischer and John Collins, 188–203. Lodon: Routledge.

Machery, Edouard. 2017. *Philosophy within Its Proper Bounds*. Oxford: Oxford University Press.

Machery, Edouard. 2020. "Response to Alexander and Weinberg, Baz and Deutsch." *Analysis* 80 (4): 771–788. https://doi.org/10.1093/analys/anaa058.

Machery, Edouard, H. Clark Barrett, and Stephen P. Stich. 2021. "No Way around Cross-Cultural and Cross-Linguistic Epistemology." *Behavioral and Brain Sciences* 44: e160. https://doi.org/10.1017/S0140525X20001831.

Machery, Edouard, Christopher Grau, and Cynthia L. Pury. 2020. "Love and Power: Grau and Pury (2014) as a Case Study in the Challenges of X-phi Replication." *Review of Philosophy and Psychology* 11 (4): 995–1011. https://doi.org/10.1007/s13164-020-00465-x.

Machery, Edouard, Ron Mallon, Shaun Nichols, and Stephen P. Stich. 2004. "Semantics, Cross-Cultural Style." *Cognition* 92 (3): B1–12. https://doi.org/10.1016/j.cognition.2003.10.003.

Machery, Edouard, Stephen Stich, David Rose, Amita Chatterjee, Kaori Karasawa, Noel Struchiner, Smita Sirker, Naoki Usui, and Takaaki Hashimoto. 2017a. "Gettier across Cultures." *Noûs* 51 (3): 645–664. https://doi.org/10.1111/nous.12110.

Machery, Edouard, Stephen Stich, David Rose, Mario Alai, Adriano Angelucci, Renatas Berniūnas, Emma E. Buchtel, et al. 2017b. "The Gettier Intuition from South America to Asia." *Journal of Indian Council of Philosophical Research* 34 (3): 517–541. https://doi.org/10.1007/s40961-017-0113-y.

Macleod, James A. 2019. "Ordinary Causation: A Study in Experimental Statutory Interpretation." *Indiana Law Journal* 94 (3): 957–1029.
Mallon, Ron, Edouard Machery, Shaun Nichols, and Stephen P. Stich. 2009. "Against Arguments from Reference." *Philosophy and Phenomenological Research* 79 (2): 332–356. https://doi.org/10.1111/j.1933-1592.2009.00281.x.
Myers-Schulz, Blake and Eric Schwitzgebel. 2013. "Knowing that P without believing that P." *Noûs* 47 (2): 371–384. https://doi.org/10.1111/nous.12022.
Nadelhoffer, Thomas, Trevor Kvaran, and Eddy Nahmias. 2009. "Temperament and Intuition: A Commentary on Feltz and Cokely." *Consciousness and Cognition* 18 (1): 351–355. https://doi.org/10.1016/j.concog.2008.11.006.
Nado, Jennifer. 2014. "Philosophical Expertise." *Philosophy Compass* 9 (9): 631–641. https://doi.org/10.1111/phc3.12154.
Nado, Jennifer. 2015. "Philosophical Expertise and Scientific Expertise." *Philosophical Psychology* 28 (7): 1026–1044. https://doi.org/10.1080/09515089.2014.961186.
Phillips, Jonathan, Wesley Buckwalter, Fiery Cushman, Ori Friedman, Alia Martin, John Turri, Laurie Santos, and Joshua Knobe. 2021. "Knowledge before Belief." *Behavioral and Brain Sciences* 44: E140. https://doi.org/10.1017/S0140525X20000618.
Robbins, Erin, Jason Shepard, and Philippe Rochat. 2017. "Variations in Judgments of Intentional Action and Moral Evaluation across Eight Cultures." *Cognition* 164: 22–30. https://doi.org/10.1016/j.cognition.2017.02.012.
Rose, David, Edouard Machery, Stephen Stich, Mario Alai, Adriano Angelucci, Renatas Berniūnas, and Emma E. Buchtel, et al. 2019. "Nothing at Stake in Knowledge." *Noûs* 53 (1): 224–247. https://doi.org/10.1111/nous.12211.
Salmon, Wesley. 1990. "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes." In *Scientific Theories*, edited by C. W. Savage, 175–204. Minneapolis: University of Minnesota Press.
Sarkissian, Hagop, Amita Chatterjee, Felipe De Brigard, Joshua Knobe, Shaun Nichols, and Smita Sirker. 2010. "Is Belief in Free Will a Cultural Universal?" *Mind and Language* 25 (3): 346–358. https://doi.org/10.1111/j.1468-0017.2010.01393.x.
Schindler, Samuel and Pierre Saint-Germier. Forthcoming. "Philosophical Expertise Put to the Test." *Australasian Journal of Philosophy*.
Sękowski, Krzysztof, Adrian Ziółkowski, and Maciej Tarnowski. Forthcoming. "Western Skeptic vs Indian Realist. Cross-Cultural Differences in Zebra Case Intuitions." *Review of Philosophy and Psychology*.
Seyedsayamdost, Hamid. 2015. "On Normativity and Epistemic Intuitions: Failure of Replication." *Episteme* 12 (1): 95–116. https://doi.org/10.1017/epi.2014.27.
Sosa, Ernest. 2009. "A Defense of the Use of Intuitions in Philosophy." In *Stich and His Critics*, edited by Dominic Murphy and Michael Bishop, 101–112. London: Blackwell.
Stich, Stephen P. and Edouard Machery. Forthcoming. "Demographic Differences in Philosophical Intuition: A Reply to Joshua Knobe." *Review of Philosophy and Psychology*.
Stigler, George J. 1955. "The Nature and Role of Originality in Scientific Progress." *Economica* 22 (88): 293–302. https://doi.org/10.2307/2551184.
Stuart, Michael T., David Colaço, and Edouard Machery. 2019. "P-curving X-phi: Does Experimental Philosophy Have Evidential Value?" *Analysis* 79 (4): 669–684. https://doi.org/10.1093/analys/anz007.
Sytsma, Justin and Jonathan Livengood. 2011. "A New Perspective concerning Experiments on Semantic Intuitions." *Australasian Journal of Philosophy* 89 (2): 315–332. https://doi.org/10.1080/00048401003639832.
Sytsma, Justin and Edouard Machery. 2010. "Two Conceptions of Subjective Experience." *Philosophical Studies* 151 (2): 299–327. https://doi.org/10.1007/s11098-009-9439-x.
Totenberg, Nina. 2013. "Supreme court: Floating home still a man's castle." *NPR online*. January 15, 2013. https://www.npr.org/2013/01/15/169452244/supreme-court-rules-that-houseboats-are-houses-not-boats.

van Dongen, Noah, Matteo Colombo, Felipe Romero, and Jan Sprenger. 2021. "Intuitions about the Reference of Proper Names: A Meta-Analysis." *Review of Philosophy and Psychology* 12 (4): 745–774. https://doi.org/10.1007/s13164-020-00503-8.

Weinberg, Jonathan M. and Joshua Alexander. Forthcoming. "Practice Makes Perfect: On Minding Methodology When Mooting Metaphilosophy." *Oxford Studies in Experimental Philosophy*.

Weinberg, Jonathan M., Shaun Nichols, and Stephen P. Stich. 2001. "Normativity and Epistemic Intuitions." *Philosophical Topics* 29 (1/2): 429–460. https://doi.org/10.5840/philtopics2001291/217.

Wiegmann, Alex, Joachim Horvath, and Karina Meyer. 2020. "Intuitive Expertise and Irrelevant Options." *Oxford Studies in Experimental Philosophy* 3: 275–310.

Williamson, Timothy. 2011. "Philosophical Expertise and the Burden of Proof." *Metaphilosophy* 42 (3): 215–229. https://doi.org/10.1111/j.1467-9973.2011.01685.x.