Religious Studies (2025), 1–17
doi:10.1017/S0034412525000150

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Experimenting with philosophy of religion: Lessons from two decades of experimental philosophy

Paul Rezkalla

Department of Philosophy, United States Air Force Academy, Colorado Springs, CO, USA
Email: paul.rezkalla@afacademy.af.edu

(Received 19 February 2025; accepted 20 February 2025)

Abstract

In this paper, I bring together several strands of criticism of experimental philosophy and draw out certain lessons for the nascent field of experimental philosophy of religion (XPoR). I argue that the negative/positive distinction conflates several underlying questions that conceptually come apart, thus undermining the framework such that XPoR need not try to accommodate the framework. I then argue that for certain topics of study in XPoR, the folk may actually be treated as a kind of 'expert' class, thus defending the utility of gauging folk beliefs on those issues. Lastly, I offer some reflections on the etic/emic distinction as it relates to the philosopher/folk divide with respect to topics in XPoR.

Keywords: experimental philosophy; philosophy of religion; philosophical methodology; intuition; expertise

'What is philosophy in Java is theatre in Bali'. – Clifford Geertz

Introduction

In this paper, I bring together several strands of criticism of experimental philosophy and draw out certain lessons for the nascent field of experimental philosophy of religion (XPoR). I begin by recounting the standard origin story of experimental philosophy as a negative research programme and challenge the negative/positive framework. I argue that the negative/positive distinction conflates several underlying questions that conceptually come apart, thus undermining the framework such that XPoR need not try to accommodate the framework. I then argue that for certain topics of study in XPoR, the folk may actually be treated as a kind of 'expert' class, thus defending the utility of gauging folk beliefs on those issues. However, I argue that if classical theism is true then the domain of expertise for the folk might be narrower. Lastly, I offer some reflections on the etic/emic distinction as it relates to the philosopher/folk divide with respect to topics in XPoR.

What is experimental philosophy?

Experimental philosophy is a big tent and, as a result, its boundaries are ill-defined, but the ambiguity is, plausibly, a feature and not a bug since it allows for novel interpretations of

the exhortation to take seriously empirical methods in conversation with standard philosophical practice (Knobe & Nichols 2008). However, the field's most narrow conceptions are marked by 'a commitment to using controlled and systematic experiments to explore people's intuitions and conceptual usage and to examine how the results of such experiments bear on traditional philosophical debates' (Nahmias and Nadelhoffer 2007). Broader construals go further and include all 'philosophical work that uses various empirical results, particularly from the cognitive sciences, in philosophical theorizing' (Rose and Danks 2013, p. 515). And even more broadly, theorists like Theodore Bach include the use of empirical tools to assess the trends of philosophical 'consensus' and even mark out the contours of the philosophical 'Overton Window'.

Irrespective of how we demarcate the outer bounds of experimental philosophical methodology, the field itself began in the early 2000s when philosophers like Edouard Machery, Steve Stich, Alex Weinberg, Shaun Nichols, Eddy Nahmias, Joshua Knobe, and others began employing questionnaire methods to gauge intuitions about philosophical issues like knowledge and moral responsibility across demographic lines. In 2001, Machery, Stich, and Weinberg ran a study to figure out how Western and Asian audiences compared in their assessment of traditional Gettier-style problems (Weinberg et al 2001). In 2003, Joshua Knobe's study showed that some people asymmetrically attribute intentionality to good and bad side effects – spawning a literature on what came to be called 'The Knobe Effect' (Knobe 2003). In 2004, Machery, Mallon, Nichols, and Stich found that Asian participants indicated a preference for descriptionist over causal-historical theories of references (Machery et al. 2004). Dozens more projects have since followed targeting mainly philosophical concepts like knowledge, moral responsibility/free will, and moral judgement culminating in a distinct research programme that's come to be known as 'the negative programme'.

The 'negative/positive' distinction

Despite its ominous sounding name, the negative programme's aims and methods merely emphasized and tried to capture the diversity of intuitions and viewpoints on philosophical issues across demographic lines – this has become especially salient now since the scientific study of human behaviour realized just how Western, Educated, Industrialized, Rich, and Democratic (or WEIRD, for short) it was (Henrich et al. 2010). As a result of this 'negative' emphasis in the early x-phi research programme, one common way of conceptually slicing up the experimental philosophy programme now is along the 'negative' and 'positive' axis. More formally, in 2006, Joshua Alexander and Jonathan Weinberg distinguished between the 'proper foundations' and 'restrictionist' programmes which roughly maps onto the positive/negative (respectively) programme distinction (although the proper foundationist/restrictionist distinction may have a slightly more restrictive usage – pun intended). They say that on the proper foundationist programme, 'empirical research should be conducted in order to determine what intuitions are generated in response to certain thought-experiments. The results of such research, it is proposed, can then be used as a proper evidentiary foundation for arguing for or against certain philosophical claims' (p. 62). And for the restriciontist programme, 'the problem with standard philosophical practice is that experimental evidence seems to point to the unsuitability of intuitions to serve as evidence at all' (p. 63). In other words, projects in the negative programme have been characterized by their attempt to complicate certain philosophical narratives through findings that cast doubt on the reliability of certain kinds of intuition or judgement. Projects in the positive programme are characterized by an optimism about folk intuition for philosophical progress – more specifically, surveying folk intuition can provide evidence

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

to help move past philosophical impasses in debates about, for example, free will, moral responsibility, and knowledge.

Negative programme

Experimental philosophy began as a 'negative' research programme in the early 2000s with folks like Edouard Machery, Steve Stich, Alex Weinberg, Shaun Nichols, Eddy Nahmias, and others showing that intuitions about moral responsibility and knowledge vary across demographic lines and are susceptible to manipulation by varying features of the issue's presentation. In 2001, Machery, Stich, and Weinberg ran a study to identify how Western and Asian audiences compared in their assessment of traditional Gettier-style problems. They presented the following vignette:

> Bob has a friend Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

And they asked participants whether Bob 'Really Knows' or 'Only Believes' that Jill drives an American car. Moreover, 71% of the Western participants indicated that Bob 'Only Believes' that Jill drives an American car, whereas 56% of East Asian and 61% of Indian participants indicated that Bob 'Really Knows' that Jill drives an American car. Similar versions of this study were later run and produced both similar and conflicting results (see Starmans and Friedman 2013; Nagel et al. 2013). In 2004, Machery, Mallon, Nichols, and Stich asked participants a version of Kripke's thought experiment about descriptionist versus causalist theories of reference and found that Western participants were almost twice as likely as East Asian participants to give the causalist judgement of the case. These and many other subsequent studies in the research programme tried to support the conclusion that intuitions about philosophical issues, especially controversial ones, varied significantly along demographic lines. This evidence is then taken up and marshalled against one version of what's taken to be the standard philosophical practice of using thought experiments to pump intuitions and then use those intuitions as evidence for or against certain philosophical positions.

More specifically, the negative programme's challenge against the reliability and stability of folk intuitions about philosophical issues comes from two distinct, but related, streams of evidence, namely evidence from intuition diversity and evidence from lack of sensitivity. The evidence from diversity targets the purported stability of intuitions across demographic lines. In short, people seem to disagree – and their disagreements seem to pretty reliably fall along predictable lines, or so, at least, allege the negative programmers. For example, a recent x-phi study showed that intuitions about whether evil is 'gratuitous' and therefore evidence against the existence of God seem to vary along axes of sex and level of education (McAllister et al. 2024). Arguments from sensitivity, on the other hand, try to undermine the utility of intuitions from the fact that the folk are just not sensitive to the relevant philosophical issues at play. So, it's not so much that intuitions diverge along demographic lines but that folk responses to philosophical questions are susceptible to a host of distortions and confounders and can be predictably manipulated in experimental settings. The core insight of the negative programme is that, if this story of what counts as standard philosophical practice is true, then the traditional philosophical method is in deep trouble given the fragility of the intuitions relied upon. However, some philosophers,

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

like Williamson (2016), have rightly pushed back on this conclusion by challenging this portrayal of the standard philosophical practice. And even thoroughgoing experimental philosophers like Edouard Machery have come to argue against this wooden, wielding of intuition as the bedrock of philosophical methodology. These will be discussed later.

Positive programme

Other philosophers have tried to establish a positive research programme for experimental philosophy that takes intuitions to be more stable and thus offer a promising line of evidence for philosophical theorizing – intuitions can 'provide a proper evidentiary foundation for certain philosophical claims and projects' (Alexander & Weinberg 2007, p. 61). Philosophers engaged in the positive programme are

> motivated to explore intuitions experimentally because they think that by doing so they can do a better job of conceptual analysis. They can avoid some of the idiosyncrasies, biases and performance errors that are likely to confront philosophers who attend only to their own intuitions and the intuitions of a few professional colleagues who read the same journals and who may have prior commitments to theories about the concepts under analysis. By collecting the intuitions of a substantial number of non-philosophers, Knobe maintains, we may discover important facts about ordinary concepts that have gone unnoticed by philosophers using more traditional methods of conceptual analysis. (Stich and Tobia 2016, p. 7)

Knobe's 2003 study is taken as a paradigmatic example of the positive programme. Participants were presented with the following two vignettes:

(1) The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new programme. It will help us increase profits, but it will also harm the environment'. The chairman of the board answered. 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new programme'. They started the new programme. Sure enough, the environment was harmed.
(2) The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new programme. It will help us increase profits, but it will also help the environment'. The chairman of the board answered. 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new programme'. They started the new programme. Sure enough, the environment was helped.

Most subjects (82%), for the first thought-experiment (in which the action had negative moral qualities) indicated having the intuition that the action was intentional. By contrast, most subjects (77%), for the second thought-experiment (in which the action had positive moral qualities) indicated having the intuition that the action was unintentional. Positive experimentalists take these intuitions as evidence against the philosophical claim that person S's action A is intentional just in case S intended to do A.

Problems with the negative/positive framework

What we have seen so far is the standard telling of the origin story of experimental philosophy. And, the negative/positive distinction, while it enabled the emergence of distinct

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

experimental philosophical programmes as the field was beginning to take shape and may have some current utility as a heuristic for easy identification of a project's aims, might ultimately be an unhelpful one, as I shall argue. The negative/positive framework seems to have taken its shape from the way the history of experimental philosophy unfolded, and the distinction is better seen as a loose divider between two kinds of projects that one can undertake within the broad conceptual space of experimental philosophy. However, there are good reasons to think that both the way the history of x-phi is told, and the negative/positive distinction is articulated, face serious difficulties. The reason for this is twofold. First, the aims of each research programme overlap and bleed into one another. For example, it's not clear that experimental philosophers like Josh Knobe saw their work as trying to contribute to a positive or negative programme or uncover the 'proper foundations' of certain philosophical issues. Perhaps, however, these labels are simply useful, post hoc, sorting devices to classify the distinct kinds of projects in x-phi. The problem, though, is that one can run an x-phi project without taking a position on the stability or role of intuitions, thus not designing or aiming for the experiment to get at certain conceptual targets of which she is unaware. And second, the distinction does not exhaust the possible conceptual positions one can hold regarding the issues at stake because the distinction conflates three separate issues on which experimental philosophers and their interlocuters can disagree – issues on which even those who share a camp may disagree. I will develop each of these objections in turn.

Programme overlap

While most theorists might concede that the negative/positive distinction is not a hard distinction, my objection goes further than this. The projects included under each research programme are, in themselves, neutral with respect to the aims of the distinct research programmes. No experimental paradigm, on its own, merits inclusion in one of the competing programmes – it's only when the experimenter's aims and assumptions are included that any sorting can begin to take place. For example, Knobe's 2003 study highlighting the Knobe effect could have been carried out with the aim of undermining a folk concept of intention rather than amassing an intuition base to improve philosophical theorizing about intentional action. And the 2001 study showing that East Asian populations are more inclined to attribute knowledge in Gettier cases could have been carried out with the aim of reclaiming JTB or deflationary conceptions of knowledge.

Additionally, in the aforementioned 2021 Rowe study, results suggested that intuitions diverged slightly and added factors affected the intuition elicited. For example, participants were asked about whether a fawn dying in a forest counted as an instance of gratuitous evil, and the assumption in the existing literature was that people would mostly share the traditional intuition of William Rowe. Participants were asked for their agreement on a scale from 1 to 15. The mean judgement for men was 6.7; the mean, for women was 7.1. So, women agreed with Rowe more than men. Additionally, more educated people agreed more with Rowe. The presence of pictures of the animal did not affect the intuition, but the addition of context to the vignette did. And those with experience hunting and butchering lessened agreement with Rowe. Now, is this project a negative one? Or a positive one? It really can be described either way. The experimental paradigm and results, on their own, are neutral with respect to the positive/negative distinction. The point is hopefully clear – what counts as a negative or positive project depends on how the results are related to the philosophical issues, not the experimental design or results, in themselves.

Additionally, is it the aims or the results that are supposed to determine whether a project is negative of positive? If it's the aim of the project, understood as the aim of the experimenters carrying out the project, then arguably no experimental philosopher, apart

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

from those who explicitly articulated the goals of their experiment as either undermining or bolstering the role of intuition in philosophical inquiry, has contributed to either the negative or positive programmes. It's not clear that most x-phi research is even carried out with the explicit aims of either the negative or positive programmes. For instance, what if these researchers are simply interested in uncovering the psychological processes underpinning human judgement making? What if the goals are to simply understand human psychology and behaviour better? If it's the experimenter's goal that determines whether an experiment is negative or positive, then most x-phi, which is, like Knobe's work for example, done without these explicit aims, is incorrectly characterized as negative or positive.

This assumption would also have the unsavoury implication that the same experiment can be classified as either negative or positive depending on the experimenter's aims – making it merely a label establishing programme inclusion on the basis of mere experimenter fiat. On the other hand, if it's the results that determine negative or positive programme inclusion, then this might be more plausible but highlights another problem with the negative/positive framework. To illustrate, if a study's results show that intuitions radically diverge across demographic lines, then this result might seem to land it in the negative programme. However, the negative programme does not merely take a stance on the empirical question of intuition stability; it also includes a philosophical commitment to the role of intuition in philosophical methodology. That is to say that these two commitments don't have to go together. They come apart. This is, I believe, the second and more important problem for the negative/positive framework in that it conflates several, different questions. The distinction relays more about the historical development of the field than it does about the conceptual layout of x-phi's purported aims and methods.

Conflating questions

As mentioned above, a second reason that the negative/positive distinction is unhelpful is that it papers over several, distinct but conceptually related issues, that philosophers can and have disagreed about, even with those in their own 'camps'. There are at least three separate issues at stake in the debates about experimental philosophy's goals, methods, and merits that get unhelpfully obscured by the course-grained negative/positive distinction:

1. The stability of intuitions across demographic lines
2. The role of intuitions in philosophical practice
3. The role of expertise (and who counts as an expert on what issues)

As discussed above, part of what differentiates the negative and positive programmes is the level of optimism concerning intuition stability across populations. There are two things to note here. First, this is an empirical question rather than a substantive philosophical question of methodology or metaphysical commitment. So, the answer is one that cannot be procured from the armchair. Second, this is only partly what differentiates the negative and positive programmes, so one's answer to the empirical question of intuition stability comes apart from the other commitments of the programmes, as I will show.

Now, the question of what roles intuitions should play in philosophical practice is a normative question, or at least not completely empirical. It is not something we can answer by doing x-phi, it's something we bring to bear on how we interpret the results of x-phi research. The further upstream, historical-linguistic-sociological question of how and to what extent intuitions are, in practice, used in the discipline of philosophy is partly an empirical question, but the question of whether philosophical practice should incorporate

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

and rely on intuition – and to what extent it should – are not merely empirical questions. And lastly, downstream from the affirmative answer to the first normative question regarding the use of intuition in philosophical practice comes another normative question regarding the 'who' of the intuitions to be used. This is sometimes referred to as the question of relevant philosophical expertise – whose intuitions are going to be relevant for answering and informing the philosophical questions that our discipline asks and tries to answer? This also is a normative question because whatever answer it gives assumes a conception of how groups and/or individuals are epistemically situated across the intuition landscape. In other words, when we decide to take seriously a given population's intuitions or to ignore another subset's intuitions, we are making normative claims about which sets of intuitions are good, useful, reliable, likely to advance the discussion, illuminative, etc. These are all normative concepts; thus, the question of relevant or useful expertise is not merely an empirical question. All of this is to say that these three distinct issues are papered over by the negative/positive distinction – three issues that, even if they might be related in certain ways or whose answers may come in prepackaged, theoretical-commitment bundles, do in fact come apart.

There are obviously plenty of other issues about which much ink has already been spilled such as the range of permissible data-gathering methods and how experimental philosophy differs from related fields like experimental psychology and empirically engaged philosophy, but I'm choosing these issues for four reasons. First, it seems that these three issues are most obviously obscured by the negative/positive sorting schema. Second, each enjoys a significant (and measurable) level of internal disagreement within each camp. Third, these come apart from each other and allow for novel combinations of potentially fecund frameworks not accommodatable on the current, binary, negative/positive framework. And last, pulling apart these questions may offer unique import for the emerging subdiscipline of XPoR.

The stability of intuitions across demographic lines

Let us now turn to each of these three separate issues, beginning with the stability of intuitions. Whether and to what extent intuitions are stable across demographic lines is an empirical question. The negative programme, spearheaded by experimental philosophers like Stich and Machery on this question, aims to defend the instability thesis – namely, that intuitions are not stable. This thesis is then marshalled as evidence in a broader argument against traditional philosophical practice. Note that this empirical premise is combined with an additional thesis that traditional philosophical practice is, in fact, driven by the exact sort of intuitions that the empirical thesis has deemed unreliable. And together, these assumptions drive the conclusion of the negative programme.

Now, it may have been easier to defend the instability thesis at the start of the negative programme when the aforementioned Gettier studies, for example, seemed to show that Gettier intuitions drastically varied between Western and East Asian populations. However, subsequent attempts to replicate have discovered a more complicated story (Kim and Yuan 2015; Machery et al. 2017). Or take the initial Knobe study on asymmetrical intent attributions – it doesn't follow from the fact that people reported attributing intentionality asymmetrically that therefore they do not believe an agent's intentionality is sufficient to make an action intentional. Much less does it undermine the philosophical position that an action is intentional just in case the agent intends it! The data have an alternate interpretation, namely these results provide evidence for a view on which each person has an intuition that favours compatibilism and another intuition that favours incompatibilism. What we have here is just the familiar Aristotelian point that there are tensions within

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

people's intuitions. But subsequent research suggests that this conflict is itself extremely stable (see Knobe 2021).

Additionally, in the aforementioned 2021 Rowe study, the two means of men's and women's support for Rowe's thesis were not exactly the same (6.7 and 7.1, respectively), but overall, the means are very similar, and the result is both men and women reject Rowe's intuition. This seems to hold for the other demographic groups surveyed, as well. In other words, we are not seeing evidence for the view that people's intuitions are unstable across either situations or demographic groups. We are instead of finding evidence for the view that people have conflicting intuitions, and that the conflict is itself stable. This is a conclusion that seems to be ruled out, all too quickly I might add, by proponents of the negative programme.

The above discussion of intuition stability was all too brief, but as it's an empirical question, it's one best left to the hands of capable experimentalists. The main thing is to recognize that this is an empirical question – one whose answer is not as neat as the negative programme purports and is complicated by a rapidly-changing experimental landscape.

The role of intuitions in philosophical practice

Now, as to the normative question of the utility of intuition for philosophical practice, the world of philosophy is itself divided between various camps that differ on whether philosophy as a discipline even descriptively operates fundamentally by eliciting intuitions as raw material for philosophical theorizing. More significant is the debate about whether philosophy should care about intuitions, and for what kinds of philosophical questions. And then even further downstream is the question of whose intuitions are relevant. This section will focus on the first two questions and argue that philosophy does not use intuitions in the way the negative programme alleges – and this is good news for several reasons, one of which is that it sidesteps the empirical question of intuition stability entirely!

'Intuition' is notoriously difficult to define but suffice it to say that what's at stake in the x- phi debates concerning the utility of intuition for philosophy is whether intuitions of the sort typically elicited from the folk through surveys are the building blocks of philosophical theorizing. Intuitions in this sense are closer to pretheoretical snap judgements. Now, philosophical intuitions can't really serve as pretheoretical snap judgements for the obvious reason that the philosopher is not uninitiated in the way that the folk are – at least in the kind of uninitiatedness that makes their intuitions pristine and free from the muddying of philosophical discourse (and thus useful for philosophical theorizing) according to proponents of the positive programme. So, if philosophers' intuitions are useful for philosophical theorizing, then they won't be useful in the same way that folk intuitions are supposed to be useful – perhaps even using the term 'intuition' to capture both is misleading. And in fact, this is precisely what sceptics of philosophical intuitions have argued (see Sosa 2007; Williamson 2016; Ludwig 2017).

However, before look at the case against intuition in philosophy, it's worth noting that the descriptive case is complicated. That is, the empirical story about the extent to which philosophers actually rely on intuition in their philosophizing, whether good or bad, is a complicated one. For example, there is evidence that over 50% of philosophers agree with the statement 'intuitions are useful for justifying philosophical claims' (Kuntz and Kuntz 2011). On the other hand, recent textual analyses of philosophical texts performed by Max Deutsch (2009, 2015) and Herman Capellen (2012) suggest that philosophers' psychological states (i.e., intuitions) do not play the evidential role that the restrictionists and positive experimental philosophers suppose. Even thoroughgoing experimental philosophers like Edouard Machery express skepticism that intuition plays the role described by those in the positive programme:

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

> It is thus unfortunate that experimental philosophers, including myself, have followed the philosophical tradition in describing the method of cases as eliciting intuitions, and have given the impression that their argument was directed at the alleged use of intuitions in philosophy. (2017, p. 178)

Intuition sceptics, as we can refer to them, don't believe the philosophical method, at least properly speaking, relies on intuitions in the way positive experimentalists take them to. Kirk Ludwig, for example, balks at the idea that philosophical theorizing rests on 'spontaneous judgements':

> Philosophers aim to arrive at a reflective judgment about a case and then to review it in the light of other judgments (their own and others) and more general theoretical considerations. They do not simply record their spontaneous judgments and take the third person stance toward them as neutral observations to be explained. […] we do not do this like hermits in the woods: we try out ideas and thought experiments on others, give and publish papers, take criticism, make revisions, try out new ideas generated in this process, and so on. (2017, p. 388)

And Ernest Sosa argues that philosophers don't, or at least shouldn't, care about analysing mere 'concepts' but the phenomena in the world that our concepts aim at – if this is right, then it's unclear how gauging intuitions about how concepts are deployed will be of serious use to the philosopher:

> It is often claimed that analytic philosophy appeals to armchair intuitions in the service of "conceptual analysis." But this is deplorably misleading. The use of intuitions in philosophy should not be tied exclusively to conceptual analysis. Consider some main subjects of prominent debate: utilitarian versus deontological theories in ethics, for example, or Rawls's theory of justice in social and political philosophy, or the externalism/internalism debate in epistemology; and many others could be cited to similar effect. These are not controversies about the conceptual analysis of some concept. They seem moreover to be disputes about something more objective than just a description or analysis of our individual or shared concepts of the relevant phenomena. Yet they have been properly conducted in terms of hypothetical examples, and intuitions about these examples. The questions involved are about rightness, or justice, or epistemic justification. Some such questions concern an ethical or epistemic subject matter, and not just our corresponding concepts. (2007, p. 100)

Theodore Bach (2022) defends a qualified intuition scepticism and articulates a three-level model of philosophical theorizing where 'snap judgements' provide only raw theoretical material to be refined dialectically:

> At the bottom level (Level 1) are snap judgments about category membership. These might be unbidden, immediate, not fully conscious judgments concerning the classification of features of the target case. Still, these judgments do not emerge from an epistemic vacuum. They are sprung from some schema, theory, or other cognitively stored knowledge representations (see, e.g., Kornblith 2007; Kahneman and Klein 2009), and in that respect they are theory-laden. At the next level, Level 2, are considered, reflective judgments. Here, one thinks carefully about reasons for making the classificatory judgment. One might uncover to some extent the reasons that drove

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

> the snap judgment. One might sift through alternative reasons that lead to contrary judgments. One considers arguments, implications, and explanatory values of competing classificatory judgments. And so on. At the last level (Level 3), the reflective procedures of Level 2 are extended through time as well as the social, expert community. One talks with other philosophers, receives feedback from commentators and peer reviewers, reads articles and arguments, and so on. Given that Level 1 judgments are theory-laden, the fruits of Level 2 and Level 3 judgments feed back causally into Level 1 judgments, affirming or changing them in various ways. This is like how if one receives decades of formal chess training, then one's snap judgments about position classification will change as a result.

Suppose the positive experimentalist rejects the special, narrow, privileged sense of intuition above and argues that that all of our judgements at some level bottom out in S1 intuition. If so, then the intuitionist faces a dilemma. Either she adopts the narrow sense of intuition whereby it's not clear that those snap, spontaneous judgements are directly relevant for philosophical theorizing, or she adopts a wide sense of intuition whereby:

> If intuitive judgments are the outputs of system 1 and reflective judgments of system 2, the point is that all system 2 thinking involves system 1 thinking. It is an illusion that reliance on intuitive judgments, characterized along anything like the lines sketched above, constitutes a distinctive method of armchair philosophy. In that sense of 'intuitive', all human thinking relies on intuitive judgments. (Willamson 2016, p. 26-27)

The three-level model shows the theory-ladenness of philosophical judgements – this can be good or bad. Philosophers can often miss obvious truths, but if the x-phi paradigm is to prove a useful corrective in these cases it would have to show that a given issue is one that would be immediately accessible to the folk such that the pre-reflective judgements would render philosophically respectable or productive answers.

Philosophers making claims about folk intuitions

However, even if intuition is not directly useful for philosophical theorizing, this doesn't prevent philosophers from making bold claims about folk intuitions, especially the intuitions they believe the folk have or ought to have. It's perhaps here that traditional x-phi methods can be illuminative. Take the following examples:

- 'Could an omnipotent, omniscient being have prevented the fawn's apparently pointless suffering? The answer is obvious' (Rowe 1979, p. 337).
- 'Is the fawn's suffering a pointless evil? Clearly, it certainly seems to us to be pointless' (Rowe 2006, p. 79).
- 'There appears to be near universal agreement' that the fawn's suffering will at least seem pointless to most people' (Rowe 2006, p. 79).
- 'Beginning students typically recoil at the compatibilist response to the problem of moral responsibility' (Pereboom 2016, p. xvi).
- '… we come to the table, nearly all of us, as pretheoretic incompatibilists' (Ekstrom 2002, p. 310).
- 'In my experience, most ordinary persons start out as natural incompatibilists.… Ordinary persons have to be talked out of this natural incompatibilism by the clever arguments of philosophers' (Kane 1999, p. 218).

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

- 'When ordinary people come to consciously recognize and understand that some action is contingent upon circumstances in an agent's past that are beyond that agent's control, they quickly lose a propensity to impute moral responsibility to the agent for that action' (Cover and O'Leary-Hawthorne, 1996 p. 50).

A couple of things can be said, here. We can first draw a distinction between descriptive claims where philosophers make reference to the beliefs of the general population and prescriptive claims where philosophers deploy normative conceptions of hypothetical ideal observers or even more simply claim that the general population should adopt a certain view. These two are often conflated with claims of the latter sort being taken as evidence that philosophers do indeed take seriously the intuitions and judgements of the general population as evidence for or against a certain philosophical position. For example, Stich and Tobia take even Socrates to have had this understanding of standard philosophical practice. For example, they cite this famous passage from Plato's *Republic* where Socrates seems to make reference to popular intuition in support of a claim about what justice permits:

> Well said, Cephalus, I replied: but as concerning justice, what is it? – to speak the truth and to pay your debts – no more than this? And even to this are there not exceptions? Suppose a friend when in his right mind has deposited arms with me and he asks for them when he is not in his right mind, ought I to give them back to him? No one would say that I ought or that I should be right in doing so, any more than they would say that I ought always to speak the truth to one who is in his condition.
>
> You are quite right, he replied.
>
> But then, I said, speaking the truth and paying your debts is not a correct definition of justice.
>
> Quite correct, Socrates. (Plato 1892, I, 131, 595)

However, this seems too quick. This could be a claim about folk psychology – that if you surveyed the people of Athens, you truly would not find anyone who would say that justice require you return the arms to the man who's not in his right mind. Or it's a normative claim – a claim that either (or both) people should render this verdict or that an ideal listener would render such a verdict. Stitch and Tobia defend intuition populism, namely that there is epistemic merit in the ubiquity of a belief or intuition in a population. Others like Williamson defend intuition elitism, namely the view that the philosopher's judgement should be privileged over the widespread intuitions of the folk.[1]

One reason in support of intuition elitism is the fact that philosophy is a technical subject, and we should expect that those well-versed in the technical practice of philosophy may plausibly have better instincts about philosophical questions and issues. However, philosophers are not interested in free will, moral responsibility, knowledge, love, and all the furniture of philosophical arsenal as mere technical concepts but as ordinary, everyday concepts. We want to understand what knowledge is – not what knowledge-as-a-technical-concept is. If this is the case, then it's less clear that intuition elitism is correct. Alexander and Weinberg take this as a reason in favour of intuition populism – the view that philosophers take their intuitions as evidence in virtue of the fact that the intuition is likely shared by the folk. In other words, everyday concepts require everyday intuitions. Philosophical intuitions are useful insofar as they are representative of how everyday people think about the everyday concepts that make their appearance in the philosophical context.

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

Nevertheless, this move is still too quick. While it's true that philosophers are deploying and theorizing about every day, rather than technical-versions-of-every day, concepts, it does not follow from this that philosophers are not better situated to form judgements about these everyday concepts. For example, it might be the case that the philosophical skillset equips its wielder to better see distinctions often missed in the course of everyday experience, disambiguate similar-but-distinct concepts, and ignore conceptually distorting 'noise'. These are all empirical claims, but the point is that the mere every-day-ness of the concepts that philosophers theorize about does not flatten the philosopher/folk distinction in support of intuition populism as Alexander and Weinberg contend. More support is needed.

In the words of David Papineau, 'we wouldn't expect physicists to throw up their hands in excitement just because somebody shows that different cultures have different views about the origin of the universe' (2011, p. 83–84). Should a physicist rethink her theory in light of folk intuitions? If not, then why should we think anything different about philosophy?

The folk as religious experts?

Perhaps 'philosophy' is too course-grained of a term here when debating the priority of one group's intuitions over another's. What if instead of 'philosophy' we took philosophical topics and questions on a case-by-case basis and asked of each who the relevant experts are. Even if one is an intuition elitist, it may turn out that for some philosophical topics the folk may be well-equipped, perhaps even better equipped than philosophers are, to philosophize. One reason to think that perhaps XPoR might be unique and immune to the kinds of criticisms raised by Williamson and others who privilege philosophers' judgements is that perhaps given the absolute pervasiveness of religious ideas and concepts in the lives of most ordinary individuals, perhaps the folk are peers or even experts in the religious concepts of interest to the XPoR research programme. This seems to align well with the Geertz-ian impulse that the local perspective is best understood by those for whom it is local. Similarly, we might think that for those for whom God, suffering, purpose, hell, heaven, etc. are daily features, they constitute a kind of expert base whose intuitions and judgements could, if gauged appropriately, be leveraged to advance our understanding of important concepts in the philosophy of religion. For example, in a recent study on how people think about gratuitous evil, an experimental paradigm reasonably judged that folk were well-enough acquainted with evil and suffering so as to comment on their nature and relationship, whatever it might be, to God. One need not acquire any special expertise to comment on the horrendousness or gratuitousness of evil or suffering. Indeed, perhaps the human experience is itself enough to bestow this kind of expertise. Eleonore Stump argues, for example, that certain kinds of evil and suffering can only be understood as meaningful from the 'inside' in a way that requires having gone through the experience and emerging on the other side to see whatever, if any, narrative coherence there might be to certain, traumatic series of events (see Stump 2010). Perhaps this kind of life experience constitutes a certain kind of expertise that situates one well to identify philosophical insights in the philosophy of religion.

On the other hand, it might turn out the exact opposite for other topics in the philosophy of religion. It may be that certain concepts and phenomena of interest to philosophers of religion are not what the folk encounter in their daily religious experience, at all. Philosophers of religion themselves argue about whether the 'God of the Philosophers' is the 'God of the Bible' and if XPoR is interested in the God of the Philosophers, then folk insight may not be as useful.

Jonathan Jong, for example, raises this kind of critique when evaluating the field of cognitive science of religion (CSR; Jong et al. 2015). CSR is a research programme that came

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

into its own at the turn of the century. It employs the tools of cognitive science to make sense of and articulate the origins of religion as a by-product of cognitive mechanisms that evolved for other purposes. The findings of CSR have been wielded by sceptics and believers alike for the purposes of undermining and supporting, respectively, the merits of religious belief. Sceptics like Daniel Dennett, Jesse Bering, Richard Dawkins, and Paul Bloom argue that since CSR shows how religious belief evolved as a by-product of our cognitive faculties, then religious belief is somehow less justified. Using the same evidence, believers like Justin Barrett arrive at the exact opposite conclusion – that the fact that religious belief is so ubiquitous and its origins found in our cognitive faculties is evidence that, or is at least consonant with, the conclusion that God exists and has designed humans with a predisposition to easily acquire religious belief.[2] We are 'born believers'.

However, Jong argues that a closer look at the evidence suggests that the 'God of CSR' and the 'God of the Bible', or even the 'God of the Philosophers' are not the same. It's unclear that the supernatural entities that CSR purports we have a tendency to believe in share enough features with the God of classical theism. The CSR story has humans overattributing agency to natural phenomena resulting in belief in and worship of local sky, rain, nature deities – divine trees that can grant wishes etc., and ancestral spirits:

> If they existed, the gods of the cognitive science of religion seem to be things in the world, effectively parts of the spatio-temporal universe; agents among other agents that interact with other things in the world in much the same way that we do. They are more or less like normal agents, except more powerful. (Jong 251)

If God is as the classical theistic tradition conceives of, then perhaps Jong's critique has some merit. And this would further complicate things for XPoR. For example, studies that attempt to gauge how folk attribute purpose to the world and the items in it, and from there draw conclusions about how the folk think about God and God's activity in the world may miss something if God's causation is not detectable in the way other secondary forms of causation are detectable. The classical theistic picture according to Jong conceives of God as:

> … not an object in the world, alongside other objects, God is not a cause among other causes, not one additional causal agent alongside you and me and billiard balls and other things that exert measurable forces in the world. God does not exert measurable forces in the world. Rather, God is always in everything causing everything to exist…It is impossible to point at a particular bush that God rustles, a particular gust of wind in which God is present; God rustles all bushes and is present in all gusts of wind. Thus, insofar as the cognitive mechanisms described above – considered by some to be a "god faculty" – detect God in this bush but not that, this gust of wind but not that, then it is not a particularly good detector of God, who is everywhere always acting in all things, causing them to be. (Jong 256)

Whatever the case may be, it's important to recognize that the identification of experts is a normative exercise. The answer to who we should poll is a normative one – that is, we cannot identify experts merely formally, we must rely on some substantive understanding of who the real 'knowers' are likely to be. And this question is a normative one – it requires the ability to reliably identify those who have knowledge or are likely to have it.

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

The etic and the emic

Not only might the folk not be not better than philosophers at solving technical, philosophical problems, even asking the folk for their input on technical, philosophical problems and accurately translating between the academic-folk divide also poses challenges. I won't spend too much space rehearsing problems with the reliability of self-reporting (Kauppinen 2007; Cullen 2010; Woolfolk 2013) and ecological validity (Neisser 1976; Machery 2016). Additionally, ordering effects are so powerful that Schwitzgebel and Cushman found that even philosophers with PhDs are susceptible to them 'marginally higher than the comparison groups'. They go on to observe that

> [i]t is particularly striking that philosophical expertise did not reduce order effects for cases intended to target the doctrine of double effect, the action-omission distinction, and the principle of moral luck, given that these philosophical principles are widely discussed in terms of hypothetical scenario comparisons very much like those we presented to our participants.… Aggregating across all three principles, we found a significant order effect on philosophers' endorsements of general moral principles that was three times larger than the corresponding, non-significant effect for non-philosophers. (Schwitzgabel & Cushman 2012, p. 148-149)

While these criticisms are important, they are not insurmountable. XPoR needs to take seriously the psychological pitfalls of experimental designs. In this final section, I'd like to offer some reflections on the etic/emic distinction as it pertains to the folk/philosopher divide both in general philosophy and as a way forward for XPoR. In 1954, Kenneth Pike articulated the etic/emic distinction as a way to capture what went wrong in the failure of certain ethnographic strategies – ethnographic approaches that failed to understand people groups 'from the inside' (Pike 1954). As Clifford Geertz came to later point out, this distinction is captured by, or at least virtually synonymous with, other families of distinctions which we regularly deploy – these include 'inside' versus 'outside', or 'first person' versus 'third person' descriptions; 'phenomenological' versus 'objectivist', or 'cognitive' versus 'behavioural' theories … 'experience-near' and 'experience-distant' (Geertz 1974, p. 28). This criticism forced scientists of human behaviour to think more carefully about question of interpretation and translation, not just across linguistic divides, but across religious, cultural, ethnic, and even expert/non-expert divides. The basic thrust of the criticism is the idea that observers can only acquire so much knowledge 'from the outside' so to speak – or at least, there might be assumptions, ways of speaking, and even logics internal to behavioural paradigms that may be missed from the outside perspective of the observer.

All of the debates rehearsed above can be understood as particular instantiations of a perennial problem in the scientific study of human behaviour – the problem of articulating an 'inside' account with 'outside' constructs, frameworks, and terms. XPoR can perhaps glean insights from theoretical and practical advances in the anthropological and ethnographic debates of the latter 20th century. Specifically, the distinction articulated as the etic/emic distinction articulated initially by Kenneth Pike and later defended by Clifford Geertz. According to Pike, the scientific etics, 'the application of our theories in analysing others' behaviour and institutions' differs from the emics, which is 'the interpretation of others' worlds as they appear to them' (Jardine 2004, p. 261). Geertz channels Pike's insight into a reminder for those working on the scientific study of human behaviour that the activity we're engaged in is, at least partially, a normative one – one that indispensably makes use of the paradigms we inhabit:

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

> In finished anthropological writings, including those collected here, this fact–that what we call our data are really our own constructions of other people's constructions of what they and their compatriots are up to-is obscured because most of what we need to comprehend a particular event, ritual, custom, idea, or whatever is insinuated as background information before the thing itself is directly examined. (Even to reveal that this little drama took place in the highlands of central Morocco in 1912-and was recounted there in 1968- is to determine much of our understanding of it.) There is nothing particularly wrong with this, and it is in any case inevitable. But it does lead to a view of anthropological research as rather more of an observational and rather less of an interpretive activity than it really is. Right down at the factual base, the hard rock, insofar as there is any, of the whole enterprise, we are already explicating: and worse, explicating explications. (Geertz 1973, p. 9)

And good news for x-phi: the field has already demonstrated its willingness to learn from the errors of mainstream psychology and take on-board critiques from philosophy and cultural anthropology. For example, researchers participating in the 'XPhi Replicability Project' (Cova et al. 2021) provide evidence that suggests there is greater replication of x-phi results than there is of results in social psychology generally (see also Machery 2017; Colombo et al. 2018; Stuart et al. 2019). X-phi has also admirably sought to avoid drawing conclusions from WEIRD samples, but the etic/emic critique goes perhaps even further and requires even more reflection about how we even assemble our constructs and experimental paradigms, as is illustrated by this anecdote from Joseph Henrich's work on the Matsigenka and Mapuche in southern Chile:

> We looked for a shared underlying mental model of why one would not eat these marine species during pregnancy or breastfeeding—a causal model or set of reasoned principles. Unlike the highly consistent answers on what not to eat and when, women's responses to our why questions were all over the map. Many women simply said they did not know and clearly thought it was an odd question. Others said it was "custom." Some did suggest that the consumption of at least some of the species might result in harmful effects to the fetus, but what precisely would happen to the fetus varied greatly, though a nontrivial segment of the women explained that babies would be born with rough skin if sharks were eaten and smelly joints if morays were eaten. Unlike most of our interview questions on this topic, the answers here had the flavor of post-hoc rationalization: "Since I'm being asked for a reason, there must be a reason, so I'll think one up now."
>
> Of course, it's not particularly difficult to get similar responses from educated Westerners, but there remains a striking difference: educated Westerners are trained their entire lives to think that behaviors must be underpinned by explicable and declarable reasons, so we are more likely to have them at the ready and feel more obligated to supply "good" reasons upon request. Saying "it's our custom" is not considered a good reason. The pressure for an acceptable, clear, and explicit reason for doing things is merely a social norm common in Western populations, which creates the illusion (among Westerners) that humans generally do things based on explicit causal models and clear reasons. They often do not. (Henrich 2016, p. 101–102)

The main anthropological insight invites deeper reflection on how we can translate between paradigms, namely the paradigms of the academic philosopher and the folk.

https://doi.org/10.1017/S0034412525000150 Published online by Cambridge University Press

Moreover, the exhortations from Geertz and the anecdotes from Henrich caution us against academic hubris that would have us flatten subtleties of association and interpretation across the epistemic divides we seek to traverse. And lastly, we should be mindful of the normativity inherent in our judgement calls of who is and is not a useful source of information in our experimentation.

Notes

1. I'm borrowing these terms from Weinberg and Alexander.
2. Kelly James Clark & Justin Barrett, 'Reformed Epistemology and the Cognitive Science of Religion,' in *Faith and Philosophy* 27 (2010), 174–189. Kelly James Clark & Justin Barrett, 'Reidian religious epistemology and the cognitive science of religion', in *Journal of the American Academy of Religion* 79 (2011), 639–675. Joshua Thurow, 'Does Cognitive Science Show Belief in God to Be Irrational? The Epistemic Consequences of the Cognitive Science of Religion', in *International Journal for Philosophy of Religion* 74 (2013), 77–98.

References

Alexander J and Weinberg JM (2007) Analytic epistemology and experimental philosophy. *Philosophy Compass* **2**(1), 56–80.

Bach T (2022) On the limitations and criticism of experimental philosophy. In Bauer, Alexander Max and Kornmesser, Stephan. (ed.) *The Compact Compendium of Experimental Philosophy*, pp. 101.

Cappelen Herman (2012) *Philosophy without Intuitions*. Oxford: Oxford Academic, (online edn) doi: 10.1093/acprof:oso/9780199644865.001.0001

Colombo M, Duev G, Nuijten MB and Sprenger J (2018) Statistical reporting inconsistencies in experimental philosophy. *PLoS One* **13**(4), e0194360. doi:10.1371/journal.pone.0194360.

Cova F, Strickland B, Abatista A, Allard A, Andow J, Attie M, Beebe J, Berniūnas R, Boudesseul J, Colombo M, Cushman F, Diaz R, van Dongen, N'Djaye Nikolai N, Dranseika V, Earp BD, Torres AG, Hannikainen I, Hernández-Conde JV, Wenjia H, Jaquet F, Khalifa K, Kim H, Kneer M, Knobe J, Kurthy M, Lantian A, Liao S-Y, Machery E, Moerenhout T, Mott C, Phelan M, Phillips J, Rambharose N, Reuter K, Romero F, Sousa P, Sprenger J, Thalabard E, Tobia K, Viciana H, Wilkenfeld D, and Zhou X (2021) Estimating the reproducibility of experimental philosophy. *Review of Philosophy and Psychology* **12**(1), 9–44.

Cover JA, and O'Leary-Hawthorne J (1996) Free agency and materialism. In Jordan J and Howard-Snyder D (eds), *Faith, Freedom and Rationality*. Lanham, MD: Roman and Littlefield, pp. 50.

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

Deutsch M (2009) Experimental philosophy and the theory of reference. *Mind & Language* **24**(4), 445–466.

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

Ekstrom L (2002) Libertarianism and Frankfurt-style cases. In Kane R (ed), *The Oxford Handbook of Free Will*. Oxford University Press: New York.

Geertz C (1973) Thick description: Toward an interpretive theory of culture. In *The Interpretation of Cultures: Selected Essays* Clifford Geertz. New York, NY: Basic Books, 3–30.

Geertz C (1974) "From the native's point of view": On the nature of anthropological understanding. *Bulletin of the American Academy of Arts and Sciences* **28**(1), 26–45.

Henrich J (2016) *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, NJ: Princeton University Press.

Henrich J, Heine SJ and Norenzayan A (2010) The weirdest people in the world? *Behavioral and Brain Sciences* **33**(2–3), 61–83.

Jardine N (2004) Etics and emics (not to mention anemics and emetics) in the history of the sciences. *History of Science* **42**(3), 261–278.

Jong J, Kavanagh C and Visala A (2015) Born idolaters: The limits of the philosophical implications of the cognitive science of religion. *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie* **57**(2), 244–266.

Kahneman D and Klein G (2009) Conditions for intuitive expertise: A failure to disagree. *American Psychologist* **64**(6), 515.

Kane R (1999) Responsibility, Luck, and Chance. *Journal of Philosophy* **96**(5), 217–240.

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

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

Knobe J (2003) Intentional action and side effects in ordinary language. *Analysis* **63**(3), 190–194.
Knobe J (2021) Philosophical intuitions are surprisingly stable across both demographic groups and situations. *Filozofia Nauki* **29**(2), 11–76.
Knobe J *et al.* (2008) An experimental philosophy manifesto. *Experimental Philosophy* Joshua Knobe & Shaun Nichols. 3–14. Oxford, UK: Oxford University Press.
Kornblith Hilary (2007) Naturalism and intuitions. In: *Grazer Philosophische Studien*, **74**.
Kuntz JR and Kuntz JRC (2011) Surveying philosophers about philosophical intuition. *Review of Philosophy and Psychology* **2**, 643–665.
Ludwig K (2017) Thought Experiments and Experimental Philosophy. M.T.Stuart et al. (eds), *The Routledge Companion to Thought Experiments*. New York & London: Routledge. 385–405.
Machery E (2016) Experimental philosophy of science. Justin Sytsma, Wesley Buckwalter. *Blackwell Companion to Experimental Philosophy*, 473–490. Malden, MA: Wiley Blackwell.
Machery E (2017) *Philosophy within Its Proper Bounds*. Oxford, UK: Oxford University Press.
Machery E, et al. (2004) Semantics, cross-cultural style. *Cognition* **92**(3), B1–B12.
Machery E, Stich S, Rose D, et al. (2017) Gettier across cultures 1. *Noûs* **51**(3), 645–664.
McAllister B, Church IM, Rezkalla P, and Nguyen L (2024) Empirical challenges to the evidential problem of evil. In Nichols S and Knobe J (eds), *Oxford Studies in Experimental Philosophy*, Volume 5, Oxford University Press 185-220.
Nadelhoffer T and Nahmias E (2007) The past and future of experimental philosophy. *Philosophical Explorations* **10**(2), 123–149.
Nagel J, Mar R and San Juan V (2013) Authentic Gettier cases: A reply to Starmans and Friedman. *Cognition* **129**(3), 666–669.
Neisser U (1976) *Cognition and Reality*. W. H. Freeman and Co: San Francisco.
Papineau D (2011) The Philosophical Insignificance of A Priori Knowledge. Michael J. Shaffer & Michael L. Veber (eds.). *What Place for the a Priori?* 61–83. Chicago, IL: Open Court.
Pereboom D (2001) *Living without Free Will*. Cambridge, UK: Cambridge University Press.
Pike KL (1954) *Language in relation to a unified theory of the structure of human behavior. Mouton.* The Hague, Netherlands: Mouton.
Plato (1892) *The Dialogues of Plato: Republic. Timaeus. Critias.* Vol. 3. New York, NY: Macmillan.
Rose D and Danks D (2013) In defense of a broad conception of experimental philosophy. *Metaphilosophy* **44**(4), 512–532.
Rowe WL (1979) The problem of evil and some varieties of atheism. *American Philosophical Quarterly* **16**(4), 335–341.
Rowe WL (2006) Friendly atheism, skeptical theism, and the problem of evil. *International Journal for Philosophy of Religion* **59**(2), 79–92.
Schwitzgebel E and Cushman F (2012) Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language* **27**(2), 135–153.
Sosa E (2007) Experimental philosophy and philosophical intuition. *Philosophical Studies* **132**, 99–107.
Starmans C and Friedman O (2013) Taking 'know' for an answer: A reply to Nagel, San Juan, and Mar. *Cognition* **129**(3), 662–665.
Stich S and Tobia KP (2016) Experimental philosophy and the philosophical tradition. Wesley Buckwalter & Justin Sytsma (eds.) *Blackwell Companion to Experimental Philosophy* 3–21 Malden, MA: Wiley Blackwell.
Stuart MT, Colaço D and Machery E (2019) P-curving x-phi: Does experimental philosophy have evidential value? *Analysis* **79**(4), 669–684.
Stump E (2010) *Wandering in Darkness: Narrative and the Problem of Suffering*. Oxford, UK: Oxford University Press.
Weinberg JM, Nichols S and Stich S (2001) Normativity and epistemic intuitions. *Philosophical Topics* **29**(1/2), 429–460.
Williamson T (2016) Philosophical criticisms of experimental philosophy. *Blackwell Companion to Experimental Philosophy* 22–36 Malden, MA: Wiley Blackwell.
Woolfolk RL (2013) Experimental philosophy: A methodological critique. *Metaphilosophy* **44**(1–2), 79–87.

Cite this article: Rezkalla P (2025) Experimenting with philosophy of religion: Lessons from two decades of experimental philosophy. *Religious Studies*, 1–17. https://doi.org/10.1017/S0034412525000150