

An Experimental Philosophy Manifesto

Joshua Knobe & Shaun Nichols

It used to be a commonplace that the discipline of philosophy was deeply concerned with questions about the human condition. Philosophers thought about human beings and how their minds worked. They took an interest in reason and passion, culture and innate ideas, the origins of people's moral and religious beliefs. On this traditional conception, it wasn't particularly important to keep philosophy clearly distinct from psychology, history, or political science. Philosophers were concerned, in a very general way, with questions about how everything fit together.

The new movement of *experimental philosophy* seeks a return to this traditional vision. Like philosophers of centuries past, we are concerned with questions about how human beings actually happen to be. We recognize that such an inquiry will involve us in the study of phenomena that are messy, contingent, and highly variable across times and places, but we do not see how that fact is supposed to make the inquiry any less genuinely philosophical. On the contrary, we think that many of the deepest questions of philosophy can only be properly addressed by immersing oneself in the messy, contingent, highly variable truths about how human beings really are.

But there is also an important respect in which experimental philosophers depart from this earlier tradition. Unlike the philosophers of centuries past, we think that a critical method for figuring out how human beings think is to go out and actually run systematic empirical studies. Hence, experimental philosophers proceed by conducting experimental investigations of the psychological processes underlying people's intuitions about central philosophical issues. Again and again, these investigations have challenged familiar assumptions, showing that people do not actually think about these issues in anything like the way philosophers had assumed.

Reactions to this movement have been largely polarized. Many find it an exciting new way to approach the basic philosophical concerns that attracted them to philosophy in the first place. But many others regard the movement as insidious—a specter haunting contemporary philosophy. We suspect that the subsequent cries for exorcism are often based on an incomplete understanding of the diverse ambitions of experimental philosophy. In this brief manifesto, we



4

aim to make clear the nature of experimental philosophy, as well as its continuity with traditional philosophy.

1. EXPERIMENTAL PHILOSOPHY AND CONCEPTUAL ANALYSIS

Experimental philosophers are certainly not the first to think that important philosophical lessons can be learned by looking carefully at ordinary people's intuitions about cases. This methodological approach has a long history within the research program sometimes known as 'conceptual analysis.' It may be helpful, then, to begin by discussing the ways in which experimental philosophy departs from this earlier program.

Of course, the most salient difference is just the fact that experimental philosophers conduct experiments and conceptual analysts do not. Thus, the conceptual analyst might write, "In this case, one would surely say ...," while the experimental philosopher would write, "In this case, 79% of subjects said...." But this is only the most superficial difference. Over time, experimental philosophers have developed a way of thinking about these issues that departs in truly substantial respects from the approaches familiar from conceptual analysis.

There is no single method of conceptual analysis, but typically a conceptual analysis attempts to identify precisely the meaning of a concept by breaking the concept into its essential components, components which themselves typically involve further concepts. In an attempt to determine the meaning of a philosophically important concept, one often considers whether the concept applies in various possible cases.

The aim of this project is to achieve ever greater levels of precision. Typically, one starts out with a nebulous sense of how to pick out the property in question. Perhaps something like this:

Knowledge seems to involve some kind of counterfactual relation between people's beliefs and actual facts.

But, over time, one hopes to arrive at a more precise analysis. For example:

S knows that *p* if and only if

- 1. p
- 2. S believes that p
- 3. Not- $p \rightarrow S$ does not believe p
- 4. $p \rightarrow S$ believes that p

This research program is, by all accounts, exceedingly difficult. The philosopher toils to put together his set of necessary and sufficient conditions for the concept of interest, let's say PENCIL. But then, when he presents his results, it inevitably happens that some guy in the back of the room gives an example of an object that meets all the conditions but isn't a pencil. This sends the philosopher back to his study to make some adjustments in his definition.

The program of conceptual analysis is a highly controversial one. Some believe that it is making considerable progress and will eventually converge on







correct analyses of certain important concepts; others feel that we have never succeeded in analyzing anything in terms of anything else and that this failure points to some intrinsic flaw in the assumptions that underlie the program itself. Regardless of how one feels about this controversy, it is important to understand how the aim of experimental philosophy differs from that of conceptual analysis.

As far as we know, no experimental philosopher has ever offered an analysis of one concept in terms of another. Instead, the aim is usually to provide an account of the factors that influence applications of a concept, and in particular, the *internal psychological processes* that underlie such applications. Progress here is measured not in terms of the precision with which one can characterize the actual patterns of people's intuitions but in terms of the degree to which one can achieve explanatory depth. Typically, one starts out with a fairly superficial characterization of certain patterns in people's intuitions. Maybe something like this:

People are more inclined to regard an agent as morally responsible when the case is described in vivid and concrete detail than they are when the case is described more abstractly.

The goal, however, is to provide some deeper explanation of why the intuitions come out this way. For example:

People are more inclined to regard an agent as morally responsible when they have a strong affective reaction to his or her transgression.

And ultimately, the hope is that one will be able to arrive at a more fundamental understanding of people's thinking in the relevant domain. Maybe something like this:

People's intuitions about moral responsibility are shaped by the interaction of two different systems—one that employs an abstract theory, another that relies more on immediate affective reactions.

But note that, even if we are able to construct a theory of this sort, we still may not be able to predict people's intuitions in all possible cases. Indeed, if our theory is that people's intuitions are shaped by their affective reactions to the case at hand, we would not be able to perfectly characterize the pattern of people's intuitions unless we could develop a complete theory of the nature of people's affective reactions.

In one sense, then, it seems that the task of experimental philosophy is considerably less demanding than that of conceptual analysis. As long as we can offer an account of the internal psychological processes that underlie our judgments, we do not also need to find necessary and sufficient conditions for the application of the concept in particular cases. Some philosophers think that this fact gives us reason for optimism. They think it amounts to trading an impossible task for one in which researchers are actually making substantial progress.





 \bigoplus

6

In another sense, though, the task of experimental philosophy is quite a bit more demanding than that of conceptual analysis as traditionally practiced. Experimental philosophers would not be content just to have an understanding of the patterns of intuition one finds on the surface. Indeed, even if we had a complete and perfectly accurate characterization of those patterns, we might feel that all of the truly deep questions still remained to be answered. What we really want to know is *why* people have the intuitions they do.

2. EXPERIMENTAL PHILOSOPHY AND PHILOSOPHICAL SIGNIFICANCE

With these considerations in the background, we can turn to an issue that might at first seem rather puzzling. The puzzle arises from a kind of gulf between the *evidence* that experimental philosophers are actually gathering and the *theories* that this evidence is alleged to support. In a typical experimental philosophy paper, the evidence being gathered is about the percentages of people who hold various sorts of intuitions, but the theories under discussion are not about people's intuitions but about substantive philosophical questions in epistemology, metaphysics, or ethics. It may appear, at least on first glance, that there must be some sleight of hand involved here. How on earth could information about the statistical distribution of intuitions ever give us reason to accept or reject a particular philosophical view?

The problem only becomes more acute when one thinks about how the approach could actually be applied in practice. Suppose, for example, that a philosopher has thought deeply about a particular case and, after sustained reflection, concluded that the agent in this case is morally responsible. And now suppose that experimental studies reveal that a majority of subjects (say, 63%) hold the opposite opinion. How could such a result possibly have any impact on her philosophical work? Is she supposed to change her mind just because she finds herself in the minority?

Of course she isn't. Philosophical inquiry has never been a popularity contest, and experimental philosophy is not about to turn it into one. If the experimental results are to have any meaningful impact here, it must be in some more indirect way. The mere fact that a certain percentage of subjects hold a particular view cannot on its own have a significant impact on our philosophical work. Instead, it must be that the statistical information is somehow helping us to gain access to some other fact and that this other fact—whatever it turns out to be—is what is really playing a role in philosophical inquiry.

Our aim in this section is to explain how this trick is supposed to work. The exposition here is somewhat complicated by the fact that different projects within experimental philosophy have used fundamentally different approaches. Hence, it is not possible to point to a single basic viewpoint and say: "This viewpoint lies at the heart of all contemporary work in experimental philosophy." The only way to present this material is to look separately at a number of different strands within the movement. Although experimental philosophy is a young movement, there are already more strands than we can adequately cover.







An Experimental Philosophy Manifesto

For instance, there has been interesting work on the meanings of words and on cultural universals that we will not be able to treat here. Instead, our focus will be on three strands that have proven especially influential.

1. Sources and Warrant

It is a commonplace that sometimes people acquire beliefs from untrustworthy sources. Some cultural sources—some books, some news media, some people—are manifestly unreliable. If the source of your belief that there is extraterrestrial life is the *National Enquirer*, then your belief lacks adequate justification.

But concerns about the sources of our beliefs are not limited to processes that take place outside of us; they can extend to processes inside the human psyche. Just as we might learn that a belief comes from an unreliable *external* source (e.g., an unreliable newspaper), we might learn that a belief is the result of an unreliable or distorting *internal* source (e.g., an unreliable cognitive process). This leads us to the first major goal of experimental philosophy. The goal is to determine what leads us to have the intuitions we do about free will, moral responsibility, the afterlife. The ultimate hope is that we can use this information to help determine whether the psychological sources of the beliefs undercut the warrant for the beliefs.

The basic approach here should be familiar from the history of philosophy. Just take a look at nineteenth-century philosophy of religion. At the time, there was a raging debate about whether people's religious beliefs were warranted, and a number of philosophers (Marx, Nietzsche, Feuerbach, etc.) contributed to this debate by offering specific hypotheses about the psychological sources of religious faith. These hypotheses led to an explosion of further discussion that proved enormously valuable for a broad variety of philosophical issues.

But then something strange happened. Although arguments of this basic type had traditionally been regarded as extremely important, they came to occupy a far less significant role in the distinctive form of philosophy that rose to prominence in the twentieth century. The rise of analytic philosophy led to a diminished interest in questions about, for example, the fundamental sources of religious faith and a heightened interest in more technical questions that could be addressed from the armchair. The shift here is a somewhat peculiar one. It is not that anyone actually offered arguments against the idea that it was worthwhile to understand the underlying sources of our beliefs; rather, this traditional form of inquiry seems simply to have fallen out of fashion. We regard this as a highly regrettable development. It seems to us that questions about the sources of our religious, moral, and metaphysical beliefs are deeply important questions and that there was never any good reason to stop pursuing them. Our aim now is to return to these questions, this time armed with the methods of contemporary cognitive science.

When experimental research is understood in this broader context, one can easily see how it might have important philosophical implications. It is not that the actual percentages themselves are supposed to directly impact our philosophical inquiries. Rather, the idea is that these experimental results can





have a kind of indirect impact. First we use the experimental results to develop a theory about the underlying psychological processes that generate people's intuitions; then we use our theory about the psychological processes to determine whether or not those intuitions are warranted.

Of course, this sort of question becomes especially pressing in cases where the intuitions are actually serving as evidence for a particular philosophical view. Thus, suppose we return to our hypothetical philosopher and her question about the nature of moral responsibility. She considers a particular case and finds herself inclined to think that the agent described in this case is morally responsible. But now there is often an additional question—can the intuition be *trusted?* Clearly, an intuition developed in a jealous rage is less trustworthy than one developed after calm and careful consideration. Thus, if our hypothetical philosopher discovers that her intuition about a case is driven by such distorting emotional reactions, this will and should affect how much she trusts the intuition.

Not only does it seem to us that empirical considerations can be relevant here; it seems to us just *obvious* that empirical considerations are relevant. Surely, the degree to which an intuition is warranted depends in part on the process that generated it, and surely the best way to figure out which processes generate which intuitions is to go out and gather empirical data. How else is one supposed to proceed?

But, unfortunately, what seems obvious to one philosopher often seems obviously mistaken to another. Instead of greeting these methodological remarks as simple truisms (which, we continue to think, is what they really are), many philosophers have reacted by offering various sorts of objections. We focus here on four of the most prominent.

The Expertise Objection, Version 1

8

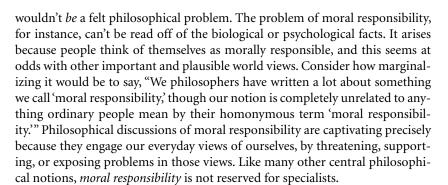
"Throughout the academy, we rely on experts to advance inquiry. It would be absurd for physicists or biologists to conduct surveys on folk intuitions about physics or biology. Rather, physicists and biologists specialize in their domains and advance the field by exploiting their specialized knowledge. The same is true of philosophy. Just as physicists don't consult folk physics, so philosophers needn't consult folk philosophy."

Reply: This view of academic specialization strikes us as entirely apt for some philosophical concerns. In some areas of philosophy, the disputes float free of commonsense intuitions. If we want to know whether the representational theory of mind is superior to connectionist alternatives, it would be ridiculous to think that we should invest our resources mulling over what the folk think about connectionism. That debate turns on facts about cognitive architecture, not facts about what people think about cognitive architecture. But in many other areas of philosophy, it's much harder to maintain that the disputes are so disconnected from commonsense intuitions. Indeed, for many standard philosophical problems—for example, problems concerning free will, personal identity, knowledge, and morality—if it weren't for commonsense intuitions, there









The Expertise Objection, Version 2

"It's true that we are concerned with questions about commonsense concepts. The point is just that philosophers can use those very concepts—the ordinary commonsense concepts that people employ every day—with a precision and subtlety that ordinary people can't quite achieve. For the philosophers are specially trained to draw fine distinctions and to think carefully; and philosophers bring these skills to bear on uncovering the true nature of our commonsense intuitions. As a result, philosophers have a much more tightly honed ability to arrive at unsullied intuitions about cases than the folk."

Reply: This version of the expertise objection argument brings up a number of fascinating issues, but we don't see how it even begins to serve as an objection to the practice of experimental philosophy. On the contrary, we would love to know more about the ways in which philosophers differ from ordinary folks, and it seems to us that the best way to find out would be to run some experiments. One could devise a series of questions and then give those questions both to philosophers and non-philosophers, checking to see how intuitions differed between the two groups. Although these experiments have not yet been conducted, we have a tentative guess about how the results would turn out. Specifically, our guess is that the overall pattern will be far more complex—and far more interesting—than anyone could have predicted from the armchair.

Furthermore, even if we discover important differences between the philosophers and the folk, it would hardly follow that data from the folk are irrelevant. Rather, the whole pattern of the data might tell us something important about the ultimate source of the philosophical problems. Philosophers are less prone to certain mistakes when processing thought experiments. On the other hand, the folk are less likely to have their intuitions biased by extensive philosophical training and theoretical affiliations. As a result, if problems like free will, moral responsibility, and personal identity flow from commonsense, then to understand these problems, it would be myopic to look only at the responses of philosophers. Rather, to understand the intuitions that are at the core of philosophical problems, one would surely want to look at different groups to see whether interesting patterns of similarity and difference emerge. The extant







work in experimental philosophy already suggests that such an investigation will reveal some very interesting patterns indeed.

The That's-Not-All-There-Is Objection

"You are simply missing the whole point of philosophy. Philosophy isn't just a matter of looking at people's intuitions and trying to understand how people think. Rather, when we are truly philosophizing, we need to subject people's intuitions to criticism, looking at arguments that might show that people's intuitions are actually mistaken in certain cases."

Reply: Here again, we think the point is well taken, but we can't see how it is supposed to be an objection to experimental philosophy. No one is suggesting that we boot out all of the moral philosophers and replace them with experimentalists, nor is anyone suggesting that we do away with any of the methods that have traditionally been used for figuring out whether people's intuitions truly are right or wrong. What we are proposing is just to add another tool to the philosopher's toolbox. That is, we are proposing another method (on top of all of the ones that already exist) for pursuing certain philosophical inquiries. Clearly, nothing in this proposal commits us to the preposterous idea that we should stop subjecting people's intuitions to philosophical scrutiny.

The You-Can't-Get-Something-for-Nothing Objection

"You'll never get anywhere if you just run a lot of experiments. Thus, suppose you are wondering about certain questions in moral philosophy. You might find that a particular psychological process tends to yield a particular type of intuition about those questions, but that knowledge won't do you any good unless you *already* have some information about either whether the process is reliable or whether the intuitions are correct. And how are you going to figure that out? Surely not just by running more experiments!"

Reply: We think that the key claim being made in this objection is right on target. If philosophers gave up all other forms of thought and just spent all of their time running experiments, it really is true that they would never get anywhere. But what we don't understand is how this claim is supposed to be an objection to the practice of experimental philosophy. After all, we are *not* going to give up all other forms of thought, and we therefore *do* have independent reasons to adopt certain beliefs. Once experimental philosophy is understood in this way as part of a broader philosophical inquiry, it shouldn't be hard to see how it could prove helpful.

The basic idea here is a straightforward one. Before we begin experimental work, we have certain beliefs both about which processes are reliable and about which answers are correct. We can then update these beliefs in light of the experimental data. Hence, when we learn that a particular process tends to generate certain types of answers, we can adjust our assessment of the process using our prior assessments of the answers. But the inference also goes in the other direction. We can use our prior beliefs about whether a given process is reliable to adjust our assessments of the answers it generates. Working back and







An Experimental Philosophy Manifesto

11

forth in this way, we gradually arrive at better assessments both of the processes and of the answers.

2. Diversity

People in different cultures have different beliefs about absolutely fundamental issues, and the recognition of this can be powerfully transforming. When Christian children learn that many people have very different religious beliefs, this can provoke a deep and disorienting existential crisis. For the discovery of religious diversity can prompt the thought that it's in some sense accidental that one happens to be raised in a Christian household rather than a Hindu household. This kind of arbitrariness can make the child wonder whether there's any reason to think that his religious beliefs are more likely to be right than those of the Hindu child. These matters are not peripheral—they strike to the heart of issues we care about most deeply.

The philosophical import of doxastic diversity is hardly restricted to child-hood. At the turn of the century, anthropologists provided a catalog of the striking cultural diversity in moral views. Some cultures, it turned out, thought that one is morally obligated to eat parts of one's deceased parents; other cultures thought it was permissible to rape women from an enemy tribe. Such diversity in moral norms was an important catalyst to philosophical reflections about the status of our moral norms, and this led to deep discussions in metaethics and normative ethics that persist to this day.

Experimental philosophy promises to make significant new contributions in this arena. Work in experimental philosophy suggests that there is diversity even in the most basic concepts we deploy in Western philosophy. For instance, basic ideas about what is required for knowledge are apparently different across cultures. This can generate a crisis akin to that of the child confronted with religious diversity. If I find out that my philosophical intuitions are a product of my cultural upbringing, then, since it's in some sense an accident that I had the cultural upbringing that I did, I am forced to wonder whether my intuitions are superior at tracking the nature of the world, the mind, and the good. These are manifestly philosophical questions. And to determine the answers, we need to know a great deal more about both our own intuitions and those of other cultures. In some cases, we might find that there are large swathes of universality in intuitions about philosophical cases. Where we do find diversity, then, we can ask more informed questions about the relative merits of these different ways of thinking about the world. And just as some Christian children come to think that there's no rational basis for preferring Christian to Hindu beliefs, we too might come to think that there's no rational basis for preferring Western philosophical notions to Eastern ones.

3. The Mind and Its Workings

Analytic philosophers have long been concerned with patterns in people's intuitions about cases, but the study of these patterns has been regarded merely as a means to an end. Hence, the philosopher might look at people's ordinary







intuitions about causation, but the true goal would not be to learn something about people and their intuitions. Instead, the goal would be to reach a better understanding of the true nature of causation, and people's intuitions would be considered relevant only insofar as they shed light on this other topic.

With the advent of experimental philosophy, this familiar approach is being turned on its head. More and more, philosophers are coming to feel that questions about how people ordinarily think have great philosophical significance in their own right. So, for example, it seems to us that there are important philosophical lessons to be gleaned from the study of people's intuitions about causation, but we do not think that the significance of these intuitions is exhausted by the evidence they might provide for one or another metaphysical theory. On the contrary, we think that the patterns to be found in people's intuitions point to important truths about how the mind works, and these truths—truths about people's minds, not about metaphysics—have great significance for traditional philosophical questions.

We are well aware that this approach is a controversial one, but we find it hard to say precisely where the controversy might lie. It seems unlikely that anyone would literally say, for example, "I know that some researchers are trying to investigate the most fundamental concepts that people use to understand their world, but that whole research program strikes me as a big mistake. In my view, these issues just aren't all that interesting or important." Nor does it seem plausible for a person to say, "I agree that we ought to be studying people's concepts and the way they think, but I don't think there is any need for experimental research here. These are the sorts of problems one can resolve entirely from the armchair." But if no one would make either of these claims, how exactly can the approach be controversial?

One complaint we sometimes hear is that philosophers should not be content merely to understand how people think, that they should also be engaged in an effort to figure out whether people's ordinary views are actually right or wrong. The thought here seems to be that, for instance, philosophers should be concerned not just with people's ordinary intuitions about causation but also with questions about what truly causes what. Clearly, this complaint rests on a confusion. No one is suggesting that philosophers should stop thinking about what really causes what. The suggestion is just that, whatever else we do, we should also be looking at people's intuitions about causation as a way of coming to a deeper understanding of how the human mind works. In other words, experimental philosophers are calling for a more pluralistic approach to philosophy. The philosopher on one end of the hall can be developing complex mathematical theories about the relevance of Bayesian inference to causal modeling, while the philosopher at the other end of the hall can be developing complex theories about how people's causal intuitions reveal some fundamental truth about human nature. If all goes well, the two philosophers will actually be able to help each other's projects advance.

As far as we can tell, the only legitimate controversy here is about whether this sort of inquiry can legitimately be considered philosophy. That is, someone





UNCORRECTED PROOF

An Experimental Philosophy Manifesto

13

might think that it is all well and good to launch an inquiry into basic questions about human nature but that such an inquiry should not take place in a philosophy department, should not be discussed in philosophy journals, should not be featured in the philosophy section of the bookstore, and so forth.

To this objection, we respond with what we have come to call the *quizzical stare*. The questions addressed in this research program strike us as so obviously philosophical that we find it a little bit difficult to know how to respond. To understand our confusion here, perhaps it would be helpful to think about the questions we ourselves have actually been investigating. One of us has been trying to figure out whether people's moral judgments are derived from reasoning, from emotion, or from some mixture of the two. The other has been trying to figure out whether the basic concepts people use to understand their world are similar to scientific concepts or whether science should be regarded as a radical departure from people's ordinary mode of understanding. To us at least, these questions seem to lie at the core of what is ordinarily regarded as philosophy.

Now, it is true that some philosophers have thought that questions about how the mind works lie outside the proper domain of philosophy, but this is a relatively recent development. Throughout almost all of the history of philosophy, questions about the workings of the mind were regarded as absolutely central. Philosophers wanted to know whether the mind was composed of distinct parts (reason, the passions, etc.) and how these parts might interact with each other. They wanted to know whether all knowledge came from experience or whether we were endowed by God with certain innate ideas. They wanted to know how exactly people come to make the moral judgments they do. The view that questions like these lie at the core of our discipline prevailed throughout most of the history of philosophy, and we therefore refer to it as the *traditional conception*.

In the early twentieth century, the rise of analytic philosophy led to a diminished interest in questions about how the mind works and a greater interest in more technical questions involving language and logic. Some of the more radical adherents of this new approach developed a particularly extreme view about how the discipline should proceed. They suggested that philosophers should not only begin to think more seriously about the new sorts of questions they had recently introduced but also stop thinking at all about more traditional questions regarding the workings of the mind. In other words, the suggestion was that the questions that had traditionally been taken to lie at the core of philosophy should now be regarded as falling outside the discipline altogether.

The result is a curious approach to undergraduate education. When students first enter the program, we tell them in reverential tones about how Plato posited a number of different parts of the mind and explained various phenomena in terms of conflict and cooperation between them. The most thoughtful and motivated students then find themselves thinking: "What an interesting idea! I wonder whether it's actually true. Let's see; I wonder what sorts of evidence







might be relevant here...." But then we are immediately supposed to put a stop to such thoughts: "No, no, you've got it all wrong. If you actually start trying to figure out whether the mind has different parts, you aren't doing philosophy at all. To truly be a philosopher, you've got to learn to leave those questions to someone else."

In our view, this is all a big mistake. There simply wasn't anything wrong with the traditional conception of philosophy. The traditional questions of philosophy—the questions that animated Plato, Aristotle, Spinoza, Hume, Nietzsche, and so many others—are just as profound and important today as they were when they were first posed. If experimental philosophy helps to bring our discipline back to these issues, we think that is cause for celebration.

3. CONCLUSION

We hope we've said enough to justify the initiation of the enterprise of experimental philosophy. But we don't think that such general considerations can provide any ultimate justification to sustain experimental philosophy. The real measure of a research program depends on whether the program generates exciting new discoveries. We invite you to read the papers and decide for yourself. For our part, we think that experimental philosophy has already begun to produce surprising and illuminating results. The thing to do now is just to cast off our methodological chains and go after the important questions with everything we've got.



