Interview on Experimental Philosophy with Joshua Knobe

Pendaran Roberts*1, Joshua Knobe2

1Department of Philosophy, University of Warwick; 2 Program in Cognitive Science & Department of Philosophy, Yale University
*Correspondence: Pendaran@mac.com

Abstract This conversation piece contains an interview with Joshua Knobe. It provides a useful introduction to what experimental philosophy is and the interdisciplinary collaborations it encourages. Pendaran Roberts and Joshua Knobe collaboratively developed this conversation piece via email. Joshua Knobe is a renowned experimental philosopher, who works on a range of philosophical issues, including philosophy of mind, action and ethics. He is a professor in the Program in Cognitive Science and the Department of Philosophy at Yale University. He is most known for what is now called the ‘Knobe effect’.

Keywords: experimental philosophy, interdisciplinary research, teaching of philosophy, job market

Introduction

Experimental philosophy is an exciting new way that philosophers are engaging with philosophical questions. Joshua Knobe is here to tell us more about experimental philosophy. This new way of engaging with philosophical questions differs from the ways most modern philosophers engage with these questions. These ways can be loosely characterised by the analytic and continental methods. Analytic philosophy is characterised by an almost mathematical style of writing that emphasises precision (the writing style bears a resemblance to that of the sciences), whereas continental philosophy is characterised by a loser style. A thorough analysis of the distinction is outside our scope. What’s important to recognise is that both types of philosophy, at least as of our time, largely shy away from empirical methods. The questions with which they deal, many of them might say, are not empirical questions.

In contrast, experimental philosophers design experiments to directly engage with philosophically relevant questions. For example, whereas an analytic or continental philosopher may just claim that such and such is
intuitive from the armchair, the experimental philosopher will hold that it is important to run an experiment to see if the folk agree. Thus, experimental philosophy can be seen as questioning philosophers who see empirical methods as irrelevant to philosophical inquiry.

The above was just a brief description of experimental philosophy. Those outside of the philosophical tradition, and many in it, have a difficult time understanding exactly what experimental philosophy is and how it is relevant to philosophy and the broader academic world. Joshua Knobe is here to help us to understand these issues. He is a well-known experimental philosopher, appointed as a Professor in both the Program in Cognitive Science and the Department of Philosophy at Yale. He was previously Assistant Professor at the University of North Carolina, Chapel Hill. His BA, which he received in 1996, is from Stanford University. Knobe received his PhD from Princeton University in 2006.

Knobe is best known for his work on intentional action (2003). He found that people ascribe intentions to others in an asymmetrical fashion. In his experiment, people were asked to imagine a corporate executive motivated entirely by profit who embarks on a policy he knows will harm the environment. Is this harm intentional? Most people say, ‘yes.’ However, when considering a similar case in which the side effects of the executive’s actions are beneficial for the environment, most say, ‘no.’ This study spawned a series of papers that further investigated this, as it is now known, ‘Knobe effect’ (Knobe, 2006; Pettit & Knobe, 2009).

**Interview**

*Pendaran Roberts (PR):* Josh, I would like to start this interview by asking you to explain in rough terms, what exactly experimental philosophy is.

*Joshua Knobe (JK):* Experimental philosophy is an interdisciplinary field at the intersection of philosophy and psychology. Very roughly, the field aims to make progress on the kinds of questions traditionally associated with philosophy using the kinds of methods traditionally associated with psychology.

*PR: Some people argue that experimental philosophy is misguided. These people believe that philosophy is not concerned with what ordinary people think, so psychological methods are irrelevant. Do you agree with this objection? Why or why not?*

*JK: When people say things like this, they are presumably drawing on a conception of philosophy that became popular only relatively recently in the history of the discipline. In particular, they seem to be thinking
specifically about the kind of philosophy that became dominant with the rise of the analytic tradition in the twentieth century.

When it comes to the issues that have been most central within this tradition, there is indeed a difficult question as to whether or not it is at all relevant to know anything about how human beings actually think or feel. Some philosophers argue that empirical facts about people’s ordinary intuitions are relevant to these issues; others argue that such facts are completely irrelevant. This debate gets us into difficult questions in philosophical methodology, questions that are very much worthy of further exploration.

However, it is also worth emphasising that analytic philosophy is a relatively recent development. In many earlier periods in the history of philosophy, people had a much broader understanding of the purview of the discipline. Thus, if we focus on the kinds of issues that were taken up in earlier periods, it may begin to seem completely obvious that facts about how human beings think and feel are philosophically relevant. These facts are not relevant just because they might bear indirectly on topics in metaphysics, epistemology or conceptual analysis. They are relevant in and of themselves, just because philosophy is centrally concerned with questions about human beings.

For a simple example, take Spinoza’s *Ethics*. Spinoza makes enormously important contributions to questions in metaphysics, and philosophers could reasonably debate whether facts about human psychology could ever shed any light on those sorts of question. However, the majority of the text is not about metaphysics. Most of it is taken up with questions concerning human psychology – indeed, with some of the very same questions that experimental philosophers are investigating today. Spinoza does not seem to think that these psychological questions are relevant only because they might somehow shed light on more metaphysical issues. Rather, he is directly concerned with philosophical questions about human life and the human mind.

If I may be forgiven a brief autobiographical digression, it was this more traditional sort of work that drew me into philosophy in the first place. I was reading Spinoza, Nietzsche, Hume, Kierkegaard, Aristotle, and I completely fell in love. Then, when I went to graduate school, I was shocked to discover that the culture of philosophy had switched around in such a way that many of the core themes in the work of these thinkers were regarded as falling outside the scope of the discipline. (One professor told me explicitly that most of Hume’s *Treatise of Human Nature* didn’t count as philosophy.) Anyway, my sense is that the pendulum is now swinging back. A lot of the papers being published in
philosophy journals these days seem closer in spirit to Hume than they do to the stuff people were doing in the 1970s.

Okay, with all this in mind, consider again the question about the philosophical relevance of experimental philosophy. Experimental philosophers have learned a lot about how people understand morality, free will, knowledge, and the self. One might see this work as an attempt to rigorously investigate people’s ordinary intuitions, and one might then ask whether it bears on the kinds of issues most closely associated with analytic philosophy (metaphysics, epistemology, etc.). This is a good question and one we should continue to investigate. However, it would be a big mistake to suppose that experimental work could only be philosophically relevant if it turned out to be relevant to issues like these. Regardless of whether the findings shed light on the issues most closely associated with analytic philosophy, they are clearly telling us something important about how people think and feel, and they are therefore relevant in a very direct way to the sorts of issues explored in more traditional philosophy.

PR: But traditionally philosophy was a much broader discipline than it is today in that many issues with which it was concerned are now recognised as being scientific questions. If experimental philosophy is not relevant to the concerns of philosophy today, what right does it have to call itself ‘experimental philosophy.’ We don’t call psychology and/or neuroscience ‘experimental philosophy of mind.’ I mean we could, but that would be confusing. Is the name ‘experimental philosophy’ confusing then? If so, maybe some change of how we categorise things is called for?

JK: To begin with, I should emphasise that I am not advocating any kind of change; I am just defending the status quo. Right now, there are a lot of people widely recognised as ‘philosophers’ working on empirical questions about the human mind. Some of them do experiments, others engage in more purely theoretical research. My view is that the system presently in place is very well-justified and should not be changed.

To see why, it might be helpful to think in a more general way about the value of interdisciplinarity. Take the area at the intersection of psychology and economics. There are people in both disciplines working on questions about consumer decision-making. In many cases, it really seems that these different researchers are working on precisely the same questions. So suppose someone asked: ‘Why do we have people working on the same question in two different disciplines? Why can’t we just pick one discipline and have all work on this question be conducted there?’

In this case, the answer seems straightforward. It is not that there is some bright line between psychological questions and economic
questions. Rather, it is that people from these different disciplines are coming out of different intellectual traditions. The psychologists know more psychology; the economists know more economics. Thus, the field as a whole can make more progress with these different disciplines working together than either discipline could have if it tried to do everything by itself.

The interdisciplinary study of mind that has flourished at the intersection of psychology and philosophy should be understood in precisely the same way. It is not that the philosophers are working on some completely separate type of question. Rather, it is that the philosophers are able to make progress on certain questions by drawing on a distinctively philosophical tradition. The key to their contribution comes from the fact that they are *philosophers*. They have spent enormous amounts of time thinking carefully about possible world semantics, Aristotle, existentialism, free will, the veil of ignorance... all the things that seem so fundamental to the philosophical tradition. If you pick up a paper written by one of these philosophers, you can usually sense immediately that there is something palpably philosophical about it. What makes it so philosophical is simply that it is quite clearly coming out of the tradition of philosophy.

PR: Many say that there are, most generally speaking, two kinds of experimental philosophy: positive and negative. The positive project seeks to contribute to philosophical debates without questioning its usual methods, for example, by investigating people’s intuitions about philosophically relevant things. The negative project is concerned with undermining the methods of philosophy, by undermining the usefulness of thought experiments. Do you agree with this way of taxonomising the discipline? Why or why not?

JK: You are completely right to say that people often describe the field in this way, but it strikes me as a wildly inaccurate characterisation of what experimental philosophers actually do.

In the very early years of experimental philosophy, there were a couple of highly influential papers pursuing what has come to be called the ‘negative programme.’ These papers argued that the existence of cross-cultural differences and other such effects provided reason to reject traditional philosophical methodology. Because those papers were published so early on, they did a lot to shape people’s perception of what the field of experimental philosophy was all about. However, the field has evolved considerably over the past fifteen years. A quantitative analysis of more recent experimental philosophy studies shows that the negative programme accounts for approximately 1.3% of the studies being conducted in recent years (*Knobe, forthcoming*).
Now, I don’t at all mean to disparage the contributions of that 1.3%, but if we are trying to divide the entire field of experimental philosophy into parts, it seems a little silly to say that this kind of work should count by itself as one of the two major parts. Instead, I would divide up the field in a different way.

One strand is concerned primarily with questions that come out of the analytic tradition, broadly construed. Research in this strand proceeds through systematic studies of people’s intuitions, but it is not really concerned with facts about human psychology per se. Rather, its primary concern is with the topics that the intuitions are about. Thus, this work is interested in moral intuitions insofar as they bear on questions in moral philosophy, epistemic intuitions insofar as they bear on questions in epistemology, and so forth. (The negative programme is best understood as one sub-strand of this first strand.)

A second strand aims to engage not so much with the analytic tradition as with the broader intellectual world. Research in this second strand turns to systematic psychological studies because it is concerned with philosophical questions that are themselves about human life and human psychology. To get a sense for the character of this second strand, consider the work of Shaun Nichols and his collaborators. They have a paper about whether Buddhism helps to allay the fear of death (Garfield, Nichols, Rai & Strohminger, 2015), one about the role of disgust in the evolution of etiquette norms (Nichols, 2002), one about the factors that have shaped religious creeds (Nichols, 2004). This work might not be very tightly related to the principal themes of twentieth century analytic philosophy, but it does connect quite closely to a more traditional understanding of what philosophy is all about.

In general, philosophers like me – those who conduct experimental studies on a regular basis – haven’t done such a great job of articulating the core ideas behind either of these strands. We have tended to focus in on specific philosophical questions, rather than trying to describe at a more abstract level how experimental philosophy works. As a result, my sense is that philosophers in each separate area have a pretty good understanding of the experimental work that is being done in their specific area but that most philosophers don’t have a very accurate understanding of what experimental philosophy has been up to more generally. I very much regret the fact that I have done so little myself to help clarify these issues.

PR: The second strand doesn’t seem to me to be experimental philosophy at all; it seems to be psychology. Do you find that a lot of philosophers feel that way? Would you object to my considering the second strand to be psychology, at least fundamentally?
JK: Good question! Since a huge proportion of all experimental philosophy falls into that second strand, it is definitely very important to get clear about this issue.

First off, it should be emphasised that analogous issues arise for just about any area of philosophy that pursues interdisciplinary research. Some philosophers spend a lot of their time proving theorems, and one could ask whether the research they do should just be labelled ‘mathematics.’ Similarly, some philosophers focus on interpreting texts from Greek and Latin antiquity, and one could ask whether the research they do should be labelled ‘classics.’ Much the same could be said of work in everything from philosophy of language to philosophy of physics.

In thinking about this question, it might be helpful to distinguish the genuine intellectual issues from the purely administrative issues. For example, difficult questions arise about how to interpret certain passages in the works of Seneca, but these questions do not come conveniently labelled as either ‘philosophy questions’ or ‘classics questions.’ Thus, when we are really trying to get at the truth about how to interpret these passages, the best approach is probably to ignore these disciplinary distinctions and simply go after the relevant questions with all the methods at our disposal. However, because the university is divided up into separate departments, there will sometimes be a purely administrative pressure to step back for a moment from this quest for truth and think instead about which kind of research belongs in which department. This is certainly a reasonable question to ask, but one shouldn’t make the mistake of thinking that it sheds any light on the deeper intellectual issues.

The situation in experimental philosophy is much the same. For example, much of my own work is concerned with questions about the basic concepts people use to make sense of the world. So I am interested in people’s concepts of causation, of intentional action, of happiness, and so forth. One possible view would be that these concepts are best understood as being at least broadly similar to scientific concepts. In other words, one might think that people’s ordinary, common-sense understanding of the world is best understood as being something like a scientific theory. Alternatively, one might argue that people’s ordinary concepts of causation, intentional action, and so forth are infused with value judgments in a way that makes them deeply different from more scientific notions. The attempt to resolve this question obviously gets us into difficult psychological issues about the nature of people’s ordinary concepts, but at the same time, it remains a recognisably philosophical question. This seems to be one of those cases in which it just isn’t very
helpful from an intellectual standpoint to focus on disciplinary distinctions.

In any case, it seems vitally important not to confuse the interdisciplinary question that experimental philosophers are actually investigating with some kind of distinctively philosophical question that is in broadly the same neighbourhood. To take one obvious example, some philosophers are interested in metaphysical questions about intentional action – questions about the nature of the actual property of being intentional – and one might think that facts about how people use the word ‘intentional’ are somehow indirectly relevant to those metaphysical questions. This is certainly a legitimate research programme, but it is not the research programme that most of the experimental philosophy in this area is pursuing. To see what is important about the research that is actually being conducted in experimental philosophy, you have to be interested not only in these metaphysical questions but also in questions that are directly concerned with human beings and the way they make sense of the world.

PR: It is sometimes said that although people speak as if they endorse interdisciplinary research, the reality is that this research is not reinforced. For example, the research often cannot be published in the mainstream journals for one’s field, but less mainstream journals are not respected as much by one’s peers. Do you agree that interdisciplinary research is not reinforced adequately by the system in which we work?

JK: I agree that philosophers doing interdisciplinary research often feel unsupported or even marginalised. Yet, when you take a look at the actual composition of our field, it may begin to seem a little bit puzzling that people should feel this way.

A little while ago, I actually did a quantitative study of papers in philosophy journals that engage with questions about the mind (Knobe, 2015). First, I looked at a sample of highly cited papers about the mind from the end of the twentieth century. Of those, 62% were purely a priori, not discussing any results of empirical studies. Then I looked at a sample of highly cited papers from the past five years. Of those, only 12% were purely a priori. All of the rest discussed results from empirical studies.

Now, I could easily imagine how people who are still working in purely a priori philosophy of mind might feel unsupported or marginalised. I can empathise with this feeling and can understand how it might be difficult to pursue that sort of research in today’s philosophical environment. What seems puzzling, however, is that it is often the very people doing more empirical work on the mind who seem to feel most marginalised.
How can they feel this way, one might ask, when the whole field is so clearly going in their direction?

A similar pattern can be observed in philosophy of language. There has been a huge surge of work in formal semantics that is deeply influenced by empirical linguistics. One might therefore expect that those philosophers of language who want to completely ignore all of this empirical research would feel a bit disconnected from the mainstream. Yet, oddly enough, I often see precisely the opposite. It is the philosophers who are most closely connected with empirical linguistics who most often describe themselves as marginalised.

What we are witnessing here is a deeply puzzling sociological phenomenon, one in which the very people who seem clearly to be controlling the directions of their fields have been made to feel peripheral to those fields. I don’t feel that I have a very good understanding of how this has happened. It certainly doesn’t seem to be a matter of more tangible rewards, such as publications or jobs. (Philosophers pursuing interdisciplinary research have done fantastically well on those dimensions.) Rather, it seems to be a matter of a more nebulous sense people have that certain research programmes constitute the ‘core’ of our discipline.

PR: The job market in philosophy is notoriously poor, and PhDs even from top programmes can struggle to attain permanent employment. Do you see experimental philosophy as being an employable research area for one’s PhD? Why or why not?

JK: In general, experimental philosophers have probably done just about as well on the job market as philosophers of other types, but it has to be said that the philosophy job market is extraordinarily unpredictable. Some experimental philosophers have gotten excellent jobs; others have had trouble getting any position at all. My sense is that it is more or less impossible to predict these outcomes in advance.

In my view, there is actually something liberating about the sheer unpredictability of the job market. If we had a clear understanding of what kind of research would be most successful from a purely careerist perspective, students would experience a real temptation to sacrifice their intellectual integrity. Instead of just trying to find the truth about the questions they were exploring, they would be tempted to do whatever would be most likely to get them a job. As it stands, though, I don’t think there is anything much to be gained from this more mercenary approach. No one has any real idea what is most likely to get you a job, so the best strategy is just to focus on doing what you genuinely believe to be good philosophy.
PR: What do you see as your biggest contribution(s) to experimental philosophy?

JK: I wish I knew. I am not sure whether anything I've done will turn out to be an important contribution and, if so, which thing it will be. So let me just say a few words about the things I happen to be working on right now.

I am interested in the way people seem to see certain entities as having what we might call essences. For example, suppose you are thinking about an academic paper. You might think, ‘Certain aspects of the paper truly constitute its essence, while others are getting away from what the paper is all about. In fact, if the author simply deleted a few sections entirely, the paper as a whole would more fully express what it was truly about all along.’

People also seem to take up this approach when thinking about human beings. Someone who really knows you well might say, ‘These aspects of him reflect what he is really all about. Then, as for those other aspects, if he could only just get rid of them, he would be more fully expressing the person that he truly was all along.’ This latter notion is what philosophers have traditionally referred to as the ‘true self.’

In our experimental studies, we find that people tend to pick out the aspects of an entity that they believe to be good and regard those aspects in particular as constituting its essence. So people tend to pick out the aspects of your paper that they believe to be good and regard those aspects as most essential to the paper (De Freitas, Tobia, Newman & Knobe, 2016). Similarly, people tend to pick out the aspects of you that they believe to be good and treat those aspects as belonging to your true self (Newman, Bloom & Knobe, 2013; Newman, De Freitas & Knobe, 2015).

Perhaps this framework can shed some light on the question we were discussing earlier about how to understand the purview of philosophy. An opponent of experimental philosophy might think that the true essence of philosophy was most fully expressed in, say, twentieth century analytic metaphysics and that experimental philosophy is getting away from the essence of the discipline. People on the opposite side might respond that the true essence of philosophy comes through most clearly in the work from earlier periods (Aristotle, Spinoza, Hume). They might think that the more narrow conception of philosophy one finds in certain twentieth century figures is just some weird aberration and that experimental philosophy is getting back to what the discipline of philosophy was truly all about all along. In a case like this, how can one possibly decide which opinion is correct?
If the approach we have been exploring is broadly on the right track, it seems that the debate is not one that can be resolved just by gathering more factual information about the history of philosophy. Rather, it is ultimately a debate that comes down to certain value judgments. The only way to resolve it is to confront those value judgments directly, discussing questions about which aspects of our discipline are truly the good ones.

**PR:** We can talk about which aspects of a paper are its good aspects only if we have a prior understanding of what a paper is. Similarly, we can talk about what the good aspects of philosophy are only if we have a prior understanding of what philosophy is. So, it seems that the debate cannot ultimately be one that comes down to value judgements. There has to be a prior understanding of the discipline and not just anything goes. Do you agree? If so, what do you think would be the right way of understanding the discipline of philosophy?

**JK:** This is another really good question. I certainly wouldn’t pretend that I know the full answer, but maybe it will be helpful for me to say just a little more about the approach we’ve been developing in our work on this topic thus far (e.g., Knobe, Prasada & Newman, 2013).

The core idea of this approach is that people associate a category with certain features and then use those features to pick out a deeper essence. For example, suppose you are thinking about the category scientist. You might start out by associating this category with certain relatively superficial features (running experiments, analysing data, developing hypotheses, etc.). But then, as you think about these features, you might begin to feel that they are not just an arbitrary list. It might begin to seem that each of these features is actually best understood as a way of realising some deeper essence (say, the impartial quest for empirical truth). Importantly, even though you identified the essence entirely by looking at the features, you might end up concluding that some people who do have all of the features nonetheless lack the deeper essence. Thus, you could end up deciding that certain individuals who spend most of their time running experiments and analysing data actually fail to embody the true essence of being a scientist.

None of our studies have been directly concerned with the concept philosophy, but it would certainly be possible to adopt the same approach there. To even get the process started, one needs to begin by associating this concept with certain features. However, the fact that we need to start out with a view about these features does not mean that the process is inert. We can use those features to identify a deeper essence, and we can then determine that certain things that do display the features nonetheless fail to embody the essence. Conversely, we
might look at certain works that conspicuously fail to display all of the obvious features, and we might then claim that these works nonetheless embody the essence picked out by those features.

**PR:** What do you think is the best way to introduce undergraduate students to experimental philosophy and/or interdisciplinary work more generally?

JK: The way I teach experimental philosophy is actually very different from the way I teach other kinds of philosophy. When I am teaching undergraduate courses in moral philosophy, I don’t think of the people taking the course as moral philosophers but rather as students. That is, my aim is not for the undergraduates taking the course to make real contributions to moral philosophy; it is just for them to learn something about the field. (Perhaps they will make a contribution later, if they continue on in philosophy.)

Experimental philosophy is not like that. Because the field is so new, it is an area in which undergraduate students can themselves make genuinely important contributions. Here at Yale, we have a system in which undergraduate students can complete a senior thesis under the supervision of a faculty member. So far, two of the senior thesis students who did experimental philosophy projects have had their work published in philosophy journals (*Falkenstien, 2013; Kim & Yuan, 2015*). So if you are an undergraduate student reading this, please don’t think of yourself only as a student. You might be a student, but you can also be a philosopher!

**PR:** Thank you very much Josh for taking your time to do this interview. I think we managed to produce some interesting discussion. Do you have any last thoughts you would like to share with our readers?

JK: Absolutely. Our conversation has mostly been concerned with questions about what the field of experimental philosophy is like right now – what methods experimental philosophers are using, what aims they are pursuing, what results they have obtained. This is certainly an important topic, but I definitely wouldn’t want people to come away with the sense that the field is destined to continue on in precisely the same way it has been going these past few years. On the contrary, there is every reason to expect that a new generation of experimental philosophers will move things in directions that those of us who first created the field could never have anticipated.
PR: One last thing, for those readers who are interested in reading some experimental philosophy, could you point us towards a few papers that would be a good introduction to the field?

JK: If you just want a general introduction to experimental philosophy, probably the best place to start would be with one of the anthologies that bring together a bunch of different papers from leading figures in the field (e.g., Knobe & Nichols, 2008; Sytsma & Buckwalter, 2016).

But one of the things I find most exciting about recent work in experimental philosophy is that it isn’t just continuing to explore the more well-established lines of inquiry you would find in those anthologies. Instead, recent work has been taking up surprising new issues, many of which are almost entirely unrelated to the questions explored within earlier experimental philosophy research. For example, there have been a series of really amazing papers on issues at the intersection of philosophy and linguistics, using experimental methods to address quite technical issues about epistemic modals (Khoo, 2015), conditionals (Cariani & Rips, forthcoming), gradable adjectives (Liao, McNally & Meskin, forthcoming), and generics (Leslie & Gelman, 2012). Similarly, I’ve been excited to see the flowering of new work on issues in metaphysics, such as mereology (Rose & Schaffer, forthcoming) and personal identity (Tobia, 2015). Continuing with that latter theme, I really loved your recent experimental paper on color (Roberts, Andow & Schmidtke, 2014).

Ultimately, I guess the main thing I would want to convey is that it’s not as though there is some one core thing that everyone interested in experimental philosophy needs to know. Rather, experimental philosophy is a methodology that can be applied to all sorts of different questions. So if you want to learn more, the best approach is just to start reading papers about whichever topic is of most interest to you.
References


Rose, D. and J. Schaffer (forthcoming), ‘Folk mereology is teleological.’ *Noûs*.


**To cite this article:**
