

## *What Is Experimental Philosophy?*

Stephen STICH\*

The term “experimental philosophy” has no standard or widely agreed upon definition. As I prefer to use the term, it has a broad extension and very fuzzy boundaries: Experimental philosophy is empirical work undertaken with the goal of contributing to a philosophical debate, though of course that may not be the only goal. Sometimes people doing experimental philosophy conduct experiments, but sometimes they don’t. Philosophically motivated ethnography, like Richard Brandt’s pioneering study of Hopi ethics, certainly counts as experimental philosophy, on my interpretation of the term (Brandt 1954). Indeed, though I am sometimes described as one of the pioneers of the experimental philosophy movement – or, more ominously, as “the Godfather” – I think Brandt was the real pioneer of contemporary experimental philosophy. Many experimental philosophers are philosophers by training and professional affiliation, but some of the best work in experimental philosophy has been done by people who do not have advanced degrees in philosophy and do not teach in philosophy departments. The work on altruism by the social psychologist Daniel Batson is, in my view, far and away the best example of experimental philosophy to date (Batson 1991, 2011).

In recent years, the term “experimental philosophy” has often been used in a much more restricted way. On that more restricted interpretation, which I will adopt for the remainder of this paper, experimental philosophy is the empirical investigation of philosophical intuitions, the factors that affect them, and the psychological and neurological mechanisms that underlie them. This characterization of experimental philosophy immediately raises a pair of questions:

1. What are philosophical intuitions?
2. Why do experimental philosophers want to study them using the methods of empirical science?

My goal in the remainder of this paper will be to explore answers to these questions.

I’ll begin with the first question, around which a lively controversy has erupted, with different philosophers defending quite different accounts (Bealer 1998; Goldman 2007; Ludwig 2007; Pust 2000; Sosa 2007; Williamson 2004; for a useful overview, see

---

\* Rutgers University

Alexander 2012). I think that the best way to approach this question is to focus on paradigm cases – uncontroversial examples of the appeal to intuition in philosophical argument<sup>1</sup>. Throughout the history of Western philosophy, episodes like the following have played an important role in philosophical argument. A philosopher describes a situation, sometimes real but more often imaginary, and asks whether some of the people or objects or events in the situation described have some philosophically interesting property or relation:

- Is the action described *morally wrong*?
- Does the person described *know* that she will not win the lottery?
- When the speaker in the story uses the word ‘water’ does the word *refer* to H<sub>2</sub>O?
- Does the “Chinese Room” exhibit real *intentionality*?

When things go well, both the philosopher and his audience will agree on an answer, with little or no conscious reflection, and they will take the answer to be *obvious*. The answer will then be used as evidence for or against some philosophical thesis. The mental states that underlie episodes of this sort are paradigm cases of philosophical intuitions.

Examples of this strategy of philosophical argument are common in the writings of Plato and Aristotle. Here is a well-known passage from *The Republic* in which Socrates uses the strategy in a conversation about the nature of justice.

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 given 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.

---

<sup>1</sup> Perhaps it would be better to say “largely uncontroversial examples,” since in philosophy almost nothing is uncontroversial. In a recent book, Cappelen (2012) maintains that, while many philosophers claim that they use intuition as an important source of evidence, they are mistaken about their own practice. For a convincing critique of Cappelen, see Wysocki (in preparation), who argues that Cappelen’s argument depends on an implausibly demanding account of when a philosopher is relying on intuition.

Quite correct, Socrates. (Plato, 1882, I, 131, p. 595; italics & boldface added.)

In the italicized sentence, Socrates sets out the imaginary situation and poses a question about justice. In the following sentence (boldface), he reports his own intuition and confidently asserts what contemporary philosophers typically assume, viz. that everyone who was confronted with the question would share his intuition. Cephalus reports the same intuition and agrees that the intuition would be shared by everyone. Then Socrates argues that the intuition shows that the account of justice that Cephalus has offered is mistaken.

Lots of examples like this can be found in more recent philosophy; they are particularly abundant in many areas of contemporary “analytic” philosophy.

- In metaphysics, debates about personal identity still invoke intuitions about Locke’s famous example of the prince and the cobbler, along with a wide variety of more recent cases in which brains are transplanted, memories and whole bodies are duplicated, and people use *Star Trek* teleporters.
- In ethics, intuitions about wayward trolleys, organ harvesting, Roman circus, inquiring murderers, children drowning in bathtubs, violinists whose survival requires being connected to someone else for nine months, and a host of other cases fill the literature.
- In discussions of free will, philosophers often invoke intuitions about people locked in prison cells, people with brain implants, people who dislike their own desires and a variety of other cases.
- In epistemology, appeal to intuitions about Gettier cases, lottery cases, fake barn cases, and stakes cases abound.
- In the philosophy of language, philosophers rely on intuitions about sorites cases, Twin Earth cases, Gödel cases and arthritis cases, among many others.
- In the philosophy of mind, intuitions about inverted spectrums, zombies, Mary cases and Chinese Room cases are widely invoked.

This list is, of course, far from complete. It would be an easy task to add dozens of additional examples.

What do the intuitions invoked in these examples have in common? As noted earlier, when things go well, people who are asked about these cases find themselves almost immediately disposed to offer an answer, though they are not consciously aware of engaging in any reasoning that leads them to that answer. I am inclined to think that this is *all* that these cases have in common. Thus I endorse a broadly inclusive account of philosophical intuition. On this point, I agree with Timothy Williamson, who maintains that more restrictive accounts of philosophical intuition will not reflect the way the term ‘intuition’ is invoked in contemporary philosophy.

Although we could decide to restrict the term ‘intuition’ to states with some list of psychological or epistemological features, such a stipulation would not explain the more promiscuous role the term plays in the practice of philosophy. This emerges more clearly in appeals to intuition in disputes over actual cases. (Williamson 2007, 218)<sup>2</sup>

Of course, it *could* turn out that most or all of the mental states that philosophers have called ‘philosophical intuitions’ share interesting psychological properties that can’t be detected without careful empirical work. But in an important recent paper, Jennifer Nado (2014) argues that this is not the case. Nado reviews a growing body of scientific evidence suggesting that “the mental states which are generally assumed to fall under the category of ‘intuition’ likely comprise a highly heterogeneous group; from the point of view of psychology or of neuroscience, in fact, ‘intuitions’ appear to be generated by several fundamentally different sorts of mental processes.” If Nado is right, and I am inclined to think she is, then one of the debates that will be center stage in the pages to follow, which focuses on the reliability of philosophical intuitions and the appropriateness of using philosophical intuitions as evidence in support of philosophical theories, will need to be recast in an important way. Rather than challenging, or defending, the epistemic status of most or all philosophical intuitions, we should be exploring the epistemic status of different sorts of philosophical intuitions that, according to our best psychology and neuroscience, are produced by different mental mechanisms.

Let me turn now to the second question posed earlier: *Why do experimental philosophers want to study philosophical intuitions using the methods of empirical science?* Broadly speaking, I think experimental philosophers can be divided into two clusters, and the people in these clusters answer this question in quite different ways. One group, which was, and continues to be, inspired by the groundbreaking work of Joshua Knobe (2003, 2005, 2010), answers the question by noting that, in one guise or another, conceptual analysis has always played a central role in philosophy, and that intuitions are typically invoked as an important source of evidence for (or against) a proposed conceptual analysis. Why are intuitions useful as evidence in conceptual analysis? Many philosophers would, I think, endorse something like the answer proposed by Alvin Goldman.

---

<sup>2</sup> Most philosophers who propose more restricted accounts of philosophical intuition are not concerned that their characterization of intuition excludes mental states that some philosophers have called ‘intuitions’. Rather than trying to capture philosophical usage, typically those who offer restricted accounts of intuition are trying to *defend* the use of intuitions as evidence in philosophy. In constructing their definitions of ‘intuition’ they try to exclude cases that they don’t think *should* be used as evidence, even if many philosophers would call those cases ‘intuitions’.

It's part of the nature of concepts . . . that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept. If the content of someone's concept F implies that F does (doesn't) apply to example x, then that person is disposed to intuit that F applies (doesn't apply) to x when the issue is raised in his mind. . . . [P]ossessing a concept makes one disposed to have pro-intuitions toward correct applications and con-intuitions toward incorrect applications – correct, that is, relative to the contents of the concept as it exists in the subject's head. (Goldman 2007, 14–15)

This account is quite similar to the standard account that Chomskian linguists offer for the use of *linguistic* intuitions in confirming or disconfirming a theory about the grammar of a person's "I-language" – the language whose grammar is actually represented in the person's mind (Chomsky 1986). Linguistic intuitions about a sentence, Chomskians maintain, typically reflect what the speaker's mentally represented grammar entails about the sentence. So, for example, if the grammar entails that the sentence is grammatical, speakers will be disposed to intuit that the sentence is grammatical, and if the grammar entails that the sentence is not grammatical, then speakers will be disposed to intuit that the sentence is not grammatical. However, as Chomsky and his followers have long noted, a speaker's linguistic intuitions are not an infallible source of information about the grammar of the speaker's I-language. Failures of attention, limits of short term memory, and a variety of other factors can produce what Chomsky calls "performance errors" leading to linguistic intuitions that do not reflect the speaker's underlying grammatical competence. Much the same can happen with philosophical intuitions. As Goldman notes, there are a number of ways in which "intuitions can go wrong." For example, the person having the intuition may have a mistaken belief about some detail of the example, or she may "lose track of some features of the example while mentally computing the applicability of F to it." For our purposes, however, the most interesting source of performance errors that Goldman mentions is that the person having the intuition "might have a false theory about her concept of F, and this theory may intrude when forming an application intuition" – i.e. an intuition about whether the concept applies to an example specified in a thought experiment. (Ibid) As Goldman goes on to note, this is a danger to which philosophers are particularly vulnerable, since they have often have well elaborated theories about the concepts they are attempting to analyze.

Knobe, and many experimental philosophers who have followed in his footsteps, in what has been dubbed "the Positive Program" of experimental philosophy, are motivated to explore intuitions experimentally primarily 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.

Knobe's own groundbreaking discovery of "the side-effect effect" provides an excellent example of the way in which experimental philosophy can lead to unexpected discoveries about philosophically important concepts. Prior to Knobe's work, there was a substantial philosophical literature aimed at analyzing the concept of intentional action. But the philosophers who contributed to that literature assumed that *intentional action* was a purely descriptive concept, and that whether or not the outcome of an action was intentional would depend entirely on the psychological states that led to the action. Knobe suspected that non-philosophers' intuitions about foreseen side-effects of actions would also be affected by the moral valence of the side-effect, though philosophers might not have these intuitions because of the influence of philosophical theories in this area. To test his hypothesis, Knobe (2003) presented subjects with scenarios like the following:

The vice-president of a company went to the chairman of the board and said, "We are thinking of starting a new program. 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 program."

They started the new program. Sure enough, the environment was harmed.

Another group of subjects was presented with the identical text, except that 'harm', 'harming' and 'harmed' were systematically replaced by 'help,' 'helping' and 'helped'. In the harm case, participants were asked how much blame the chairman deserved (on a scale from 0–6) and whether he intentionally harmed the environment. In the help case, participants were asked how much praise the chairman deserved (on a scale from 0–6) and whether he intentionally helped the environment. The results, which have been replicated many times, were quite striking. In the harm case, 82% said the chairman brought about the side-effect intentionally. But in the help case, 77% said the chairman did *not* bring about the side-effect intentionally. This remarkable finding has led to an on-going debate about what factors are responsible for this effect (Adams & Steadman 2004; Nichols & Ulatowski 2007; Alicke 2008; Machery 2008). Though not everyone agrees, Knobe and many others think that the finding reflects a previously unsuspected feature of the ordinary concept of intentional ac-

tion – a feature that had not been noticed by philosophers who relied on their own intuitions<sup>3</sup>.

Before turning to another account of why experimental philosophers want to use the methods of empirical science to study philosophical intuitions, I think it is worth noting that a largely parallel “experimental movement” has also emerged in linguistics. Concerned that reliance on the intuitions of professional linguists might obscure dialect differences or enshrine performance errors that may be caused by linguists’ theoretical commitments, a growing group of linguists have begun to do experiments in which they survey the linguistic intuitions of ordinary speakers. In a number of cases, the findings have posed an important challenge to earlier work based entirely on linguists’ intuitions<sup>4</sup>.

The second answer to the question: *Why do experimental philosophers want to study philosophical intuitions using the methods of empirical science?* is best set out against the backdrop of a different account the philosophical goals that many philosophers are pursuing. As noted earlier, the analysis of concepts has long played an important role in philosophy. And in the middle years of the 20<sup>th</sup> century, some philosophers, influenced by Logical Positivism and “ordinary language philosophy,” held the view that this was the main job of philosophy. But many philosophers, both historical and contemporary, would reject this view. According to Ernest Sosa, for example,

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

---

<sup>3</sup> An alternative hypothesis is that the side – effect effect is itself a performance error, and that the intuitions of subjects in these experiments do not tell us what their concept of intentional action actually says about some of these cases. See Alexander et al. (2010) for a discussion of some of the problems encountered in the attempt to decide between these hypotheses.

<sup>4</sup> For further discussion and some examples, see Machery & Stich (2012).

our corresponding concepts. (Sosa 2007, p. 100)

Sosa is surely right that many analytic philosophers would reject the idea that philosophy is principally concerned with the analysis of concepts. Rather, they would insist, many epistemologists are concerned with *the nature of knowledge – what knowledge is* – not with some person’s or group’s *concept of knowledge*, that many moral and political philosophers are concerned with *the nature of justice – what justice is* – not just with some person’s or group’s *concept of justice*, and so on for many other branches of philosophical inquiry.

In order to use intuitions about hypothetical examples to study these phenomena, philosophers typically assume that the contents of philosophical intuitions are likely to be true. Thus, for example, if we have the intuition that the protagonist in a typical Gettier case does not know the specified proposition, *p*, then it is likely that a person in that situation does not know that *p*, and any theory about the nature of knowledge that entails the protagonist does know that *p* is challenged. Similarly, if our intuition in a “Magistrate and the Mob” case is that it is morally wrong for the magistrate to knowingly find the innocent man guilty in order to prevent a violent riot, then it probably is morally wrong, and a moral theory that entails that it is wrong would be supported.

In the previous paragraph, I followed the ubiquitous practice of talking about “our” intuitions and what follows if “we” have a specified intuition, without ever saying who “we” are. The tacit assumption behind this practice seems to be that everyone (or almost everyone) will have the same intuition. As I noted earlier, this is an assumption that goes all the way back to Plato. Socrates insists that “*no one* would say that I ought [to return the weapons] 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” (Plato 1992; italics added). But it is an assumption about which I have long been suspicious. (See, for example, Stich 1988 & 1990.) Starting in the final decade of the 20<sup>th</sup> century, the work of Richard Nisbett and other cultural psychologists demonstrated that there are important differences in the perception, memory and cognition of people in different cultures (Nisbett 2003; Henrich et al. 2010; Heine 2011). In light of these findings, I wondered, why should we blithely assume that people in different cultures would share the same philosophical intuitions? The assumption seems to be making a bold empirical claim which is susceptible to empirical exploration. So in the final years of the last century, some of my former students and I decided to test the assumption. Many philosophers found the results of our first few studies quite unsettling. In some cases, it seemed, people with different cultural backgrounds had significantly different intuitions about standard philosophical thought experiments (Weinberg et al. 2001). If this is right, then it poses an important challenge to the use of intuitions that Sosa sketches in the passage quoted



above. For, as Sosa himself noted, if one group of people have the intuition that (say) the protagonist in a Gettier case *does* know that *p*, and another group have the intuition that the protagonist *does not* know that *p*, then they can't both be right. So without a well-supported "theory of error" that explains why the people or cultures who disagree with us are mistaken, the finding of cultural or individual differences in philosophical intuition makes the assumption that "our" intuitions are very likely to be true more than a bit problematic. (Sosa 2007)

In the dozen years since the appearance of the first papers reporting cultural variation in philosophical intuition, there have been a number of studies suggesting that intuitions vary across other demographic groups divided by age or personality or gender<sup>5</sup>. Other studies found that the order in which questions are asked, or minor details of the wording, or the physical circumstances in which the intuitions are solicited can all have an effect on the intuitions that people report<sup>6</sup>. All of this is bad news for the standard philosophical assumption that the contents of people's intuitions are very likely to be true. As Walter Sinnott-Armstrong has noted, in discussing order effects in moral intuitions,

The truth about what is morally right or wrong in the cases did not vary with [the order in which they were presented]. Hence moral [intuitions] fail to track the truth and are unreliable insofar as they are subject to such order effects. (Sinnott-Armstrong 2008, p. 67)

This challenge to a method of inquiry that has played an important role in philosophy from Plato's time onward has not, of course, gone unanswered. Some critics have raised questions about the design of experiments which claim to find demographic differences in philosophical intuitions, or to find troublesome features like order effects (Sosa 2007). Others have claimed that when design flaws are mitigated, some of the worrisome findings cannot be replicated (Nagel et al. 2013). It will, I think, require many more experimental studies, and better designed studies, before we have an accurate picture of whether and where philosophical intuitions are susceptible to problematic variation. What is clear, though, is that one of the main reasons why philosophers want to study philosophical intuitions empirically is that these studies are a central component in the debate over the viability of one of philosophy's oldest and most often used methods.

---

<sup>5</sup> For age see Colaço et al. (2014); for personality see Feltz & Cokely (2009); for gender see Buckwalter & Stich (2013).

<sup>6</sup> For order effects and wording effects (sometime also called "framing effects") see Sinnott-Armstrong (2008). For the effects of physical circumstances, see Tobia et al. (2013).

### References

- Adams, F. & Steadman, A. (2004). “Intentional Action in Ordinary Language: Core concept or pragmatic understanding?” *Analysis*, 64.
- Alexander, J. (2012). *Experimental Philosophy: An Introduction*, Cambridge, UK: Polity Press.
- Alexander, J., Mallon, R., & Weinberg, J. (2010). “Accentuate the Negative,” *Review of Philosophy and Psychology*, 1.
- Alicke, M. (2008). “Blaming Badly,” *Journal of Cognition and Culture*, 8.
- Batson, C. D. (1991). *The Altruism Question: Toward a Social – Psychological Answer*, Hillsdale, NJ: Erlbaum Associates.
- Batson, C. D. (2011). *Altruism in Humans*, New York: Oxford University Press.
- Bealer, G. (1998). “Intuition and the Autonomy of Philosophy,” in M. DePaul & W. Ramsey (eds.), *Rethinking Intuition*, Lanham, MD: Rowman & Littlefield.
- Brandt, R. (1954). *Hopi Ethics: A Theoretical Analysis*, Chicago: University of Chicago Press.
- Buckwalter, W. & Stich, S. (2013). “Gender and Philosophical Intuition,” in J. Knobe & S. Nichols (eds.), *Experimental Philosophy*, vol. 2, New York: Oxford University Press.
- Cappelen, H. (2012). *Philosophy Without Intuitions*, Oxford: Oxford University Press.
- Chomsky, Noam (1986). *Knowledge of Language*, New York: Praeger.
- Colaço, D., Buckwalter, W., and Stich, S. (2014). Epistemic Intuitions in Fake – Barn Thought Experiments. *Episteme*, 11, 2, 199–212.
- Feltz, A. & Cokely, E. (2009). “Do Judgments About Freedom and Responsibility Depend on Who You Are? Personality Differences in Intuitions About Compatibilism and Incompatibilism,” *Consciousness and Cognition*, 18.
- Goldman, A. (2007). “Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status,” *Grazer Philosophische Studien*, 74.
- Heine, S. (2011). *Cultural Psychology* (2nd ed.). New York: Norton.
- Henrich, J., Heine, S. & Norenzayan, A. (2010). “The Weirdest People in the World?” *Behavioral and Brain Sciences*, 33.
- Knobe, J. (2003). “Intentional Action and Side Effects in Ordinary Language,” *Analysis*, 63.
- Knobe, J. (2005). “Theory of Mind and Moral Cognition: Exploring the Connections,” *Trends in Cognitive Sciences*, 9.
- Knobe, J. (2010). “Person as Scientist, Person as Moralist,” *Behavioral and Brain Sciences*, 33.
- Ludwig, K. (2007). “The Epistemology of Thought Experiments: First Person vs. Third Person Approaches,” *Midwest Studies in Philosophy*, 31.
- Machery, E. & Stich, S. (2012). “The Role of Experiment,” in G. Russell & D. Graff Fara (eds.), *The Routledge Companion to the Philosophy of Language*, New York: Routledge.
- Machery, E. (2008). “The Folk Concept of Intentional Action: Philosophical and Experimental Issues,” *Mind & Language*, 23.
- Nado, J. (2014). “Why Intuition?” *Philosophy and Phenomenological Research*, 89(1): 15–41.

- Nagel, J., San Juan, V. & Mar, R. (2013). "Lay Denial of Knowledge for Justified True Beliefs," *Cognition*, 129(3): 652–61.
- Nichols, S. & Ulatowski, J. (2007). "Intuitions and Individual Differences: The Knobe Effect Revisited," *Mind & Language*, 22.
- Nisbett, R. (2003). *The Geography of Thought: How Asians and Westerners Think Differently . . . and Why*, New York: The Free Press.
- Plato (1882). *The Dialogues of Plato*, translated by B. Jowett, New York: Random House.
- Pust, J. (2000). *Intuitions as Evidence*, New York: Garland Publishing.
- Sinnott–Armstrong, W. (2008). "Framing Moral Intuitions," in W. Sinnott – Armstrong (ed.), *Moral Psychology*, vol. 2, *The Cognitive Science of Morality: Intuition and Diversity*, Cambridge, MA: MIT Press.
- Sosa, E. (2007). "Experimental Philosophy and Philosophical Intuition," *Philosophical Studies*, Vol. 132, pp. 99–107.
- Stich, S. (1988). "Reflective Equilibrium, Analytic Epistemology and the Problem of Cognitive Diversity," *Synthese*, 74.
- Stich, S. (1990). *The Fragmentation of Reason*, Cambridge, MA: MIT Press.
- Tobia, K., Stich, S. & Chapman, G. (2013). "Cleanliness is Next to Morality, Even for Philosophers," *Journal of Consciousness Studies*, 20 (11 & 12), 2013, 195–204.
- Weinberg, J., Nichols, S. & Stich, S. (2001). "Normativity and Epistemic Intuitions," *Philosophical Topics*, 29.
- Williamson, T. (2004). "Philosophical 'Intuitions' and Skepticism About Judgment," *Dialectica*, 58.
- Williamson, T. (2007). *The Philosophy of Philosophy*, Oxford: Blackwell Publishing.
- Wysocki, T. (in preparation). "Intuitions Aren't on the Rocks."

(Received 2013.8.31; Revised 2014.10.10; Accepted 2014.10.29)