

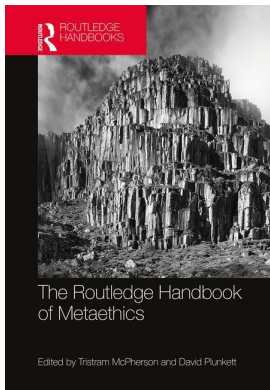
This article was downloaded by: 10.3.97.143

On: 01 Apr 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Metaethics

Tristram McPherson, David Plunkett

Experimental Philosophy and Moral Theory

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315213217.ch39>

Chandra Sripada

Published online on: 28 Aug 2017

How to cite :- Chandra Sripada. 28 Aug 2017, *Experimental Philosophy and Moral Theory from: The Routledge Handbook of Metaethics* Routledge

Accessed on: 01 Apr 2023

<https://www.routledgehandbooks.com/doi/10.4324/9781315213217.ch39>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Experimental Philosophy and Moral Theory

Chandra Sripada

In the course of philosophical debates, philosophers routinely appeal to intuitions about cases as a way of supporting their positions and challenging those of their opponents. Experimental philosophy is a relatively new field that uses the methods of psychology (and other sciences of mind, brain, and behavior) to study and scrutinize these intuitive judgments, with the ultimate goal of making progress in these philosophical debates. In this chapter, I distinguish between two important experimental philosophy projects, the sociological project and the psychological project, that are particularly relevant for moral theory. These two projects aren't exhaustive, as there are a number of other research programs that are part of experimental philosophy. Nonetheless, examining these two projects will give us a good sense of how experimental philosophy can potentially help us make progress in long-standing debates in moral philosophy.

THE SOCIOLOGICAL PROJECT

At critical junctures of philosophical arguments, one routinely encounters a family of assertions made by theorists as a way of supporting certain key claims. Where p is the claim at issue, theorists say such things as: “we intuitively judge that” p or that p is “common sense,” “a platitude of the folk,” “pre-reflectively endorsed,” “part of our ordinary understanding,” and other similar statements. One might raise doubts about whether the fact that we intuitively assent to a claim or regard it as part of common sense *ought* to count in its favor. Let us, however, put this serious issue aside. The main aim of the sociological project is to use experimental methods, especially survey methods, to test the preceding assertions by philosophers, in particular the contention that *we* intuitively endorse certain claims or that these claims form part of a *widely shared* folk common-sense understanding (see Knobe and Nichols 2008, where what I am calling the sociological project is closely related to what they call the “diversity” project).

One reason there is a need to undertake systematic testing, it is argued, is because philosophers constitute a self-selected and highly unrepresentative group. As Stephen Stich and Jonathan Weinberg put the point, why think “that the intuitions of high socio-economic status males ... who have advanced degrees in philosophy, and whose cultural background is Western European can serve as a basis for generalizations about the intuitions of ‘the folk’” (2001: 642)? There is thus a need to check philosophers’ assertions about what is part of the folk understanding with careful studies of what is in fact part of the folk understanding.

In moral philosophy, debates about moral objectivity and motivational internalism have been a fruitful area for investigation by experimental philosophers pursuing the sociological project; let us examine some of the work on these topics.

Moral Objectivity

Many philosophers say the folk are committed to the objectivity of morality or that it is implicit or “presupposed” in ordinary moral discourse and practice. Here are two examples:

[W]e seem to think moral questions have correct answers; that the correct answers are made correct by the objective moral facts; that moral facts are wholly determined by circumstances; and that, by engaging in moral conversation and argument, we can determine what these objective moral facts determined by the circumstances are.

(Smith 1994: 6)

...[W]e often see ourselves as engaged in a search for the truth about who is in the right, or where our obligations lie. We can well explain the point and persistence of moral disagreement by attributing to agents the presupposition that there is a right answer awaiting discovery.

(Shafer-Landau 2003: 23)

In the psychological literature, most studies have examined the *content* of people’s morals, the first-order principles they endorse, and the justifications they offer for their moral views. The metaethical question of moral objectivity has, in contrast, been widely neglected.

One exception is perhaps the large body of work on the so-called moral/conventional distinction that finds that children reliably distinguish moral wrongs and social/conventional wrongs (Turiel 1983; Smetana 1981; Nucci 1981). This work offers some suggestive evidence that the folk have objectivist leanings, but drawing conclusions from this work is challenging for a number of reasons. Most importantly, the theorists who did these studies weren’t directly interested in the issue of moral objectivity, and so did not ask about it specifically. Additionally, this work remains highly controversial, with a number of theorists, including philosophers, offering systematic critiques of its theoretical assumptions, methods, and findings (see for example Kelly et al. 2007).

Geoffrey Goodwin and John Darley, two psychologists, performed an interesting early study that directly investigated folk views on the objectivity of morality (in this case, the

relevant folk were undergraduates at Princeton University) (Goodwin and Darley 2008). They constructed a list of 26 statements that they classified into four types: factual, ethical, convention, and taste. In the first part of the experiment, participants indicated whether each statement is true, false, or a matter of opinion. In the second part, the experimenters told the participants that others who had previously rated the statements had disagreed with them on a number of specific items. They were then presented with questions intended to assess whether one party is mistaken or whether it is possible that neither is mistaken. A statement was assessed as having greater objectivity if participants viewed it as true or false (rather than being a matter of opinion) and endorsed that in cases of disagreement, one of the parties is mistaken. According to this measure, Goodwin and Darley found that participants rated ethical statements as being nearly as objective as factual statements and clearly more objective than convention and taste statements.

Goodwin and Darley's results extended previous work that also suggested the folk have objectivist leanings. For example, in a pioneering early study, Shaun Nichols found most participants (78 percent) rejected statements designed to capture a non-objectivist meta-ethical perspective (e.g., "There is no objective fact, independent of what different people think, about whether it was wrong for Frank to hit Bill or Lisa to shove Nancy. These actions were 'wrong for Ted' and maybe 'wrong for me', but they aren't *objectively wrong* independent of what people think about them.") (Nichols 2004). Similarly, Wainryb and colleagues (2004) found initial evidence that in cases of moral disagreement, children tended to say only one party is right, rather than that both are right, potentially suggesting that these children reject certain non-objectivist views that allow both parties to be correct.

In more recent investigations, Sarkissian and colleagues (2011) challenged the objectivist implications of the preceding body of work. They observed that prior work tended to present participants with moral disagreements in which it could be assumed the disagreeing parties come from the same sociocultural group. For example, in the Goodwin and Darley study previously discussed, participants were told the individuals disagreeing with them were other participants studied by the experimenters, and thus it could be readily inferred that these individuals were Princeton undergraduates. This is important because certain kinds of relativist views say the truth of a moral judgment depends on the culture of the person making the judgment. These relativist views would thus say that—just like objectivist views—if two individuals drawn from the same cultural group disagree about a moral matter, then one must be making a mistake.

Based on these observations, Sarkissian and colleagues performed studies designed to better distinguish the hypothesis that the folk are genuinely non-relativist realists versus the hypothesis that they are relativists who think the moral status of an action depends on the cultural standards of the person making the judgment.

In part of the study, all participants were given a statement:

Horace finds his youngest child extremely unattractive and therefore kills him.

Participants were next told about a disagreement about the statement between one of their classmates, who thinks the act is not permissible, and another party who thinks the act is permissible. Participants were assigned to one of three conditions. In the Same-Culture condition, they were told the other party is Sam, "a fairly ordinary student at their own college who enjoyed watching college football and hanging out with friends."

In the Other-Culture condition, the other party is a member of the Mamilon tribe, a tribe which “lives in the Amazon rainforests and has preserved a traditional warrior culture, with quite different values from those of the people in the surrounding society.” In the Extraterrestrial condition, the other party is a member of a race of extraterrestrial beings called the Pentars. Participants were told that “Pentars have a very different sort of psychology from human beings, that they are not at all interested in friendship or love and that their main goal is simply to increase the total number of equilateral pentagons in the universe.” Participants in all three conditions were then asked to rate their agreement on a 7-point scale (1=strongly disagree and 7=strongly agree) with the claim: “Since your classmate and (Sam, the Mamilon, the Pentar) have different judgments about this case, at least one of them must be wrong.”

Results for this question for the three conditions are shown below:

Same Culture: 5.4

Other Culture: 4.4

Extraterrestrial: 3.2

These results show participants tend to provide an “objectivist” response when the two parties in the disagreement are drawn from the same culture. But as cultural distance increases, participants are less likely to agree that at least one of the parties in the disagreement is making a mistake. This pattern is exactly what is predicted if the folk are cultural relativists about morality.

Sarkissian and colleagues report the results of several other studies that also support the idea that the folk have moral relativist leanings. For example, they replicated the results of the study reported above in an independent Singaporean sample, as well as a separate American sample, where participants were allowed to rate all three scenarios (i.e., Same Culture, Other Culture, Extraterrestrial). Finally, they showed that the tendency to give answers consistent with relativism is specific to the moral domain and does not extend to straightforward matters of fact, e.g., whether Napoleon used horses or helicopters in battle.

A potential weakness of moral relativism is that it can be hard to see why we should call a “faultless” disagreement—one in which neither party is making a mistake—a disagreement at all. In a follow-up to the Sarkissian et al. study, Justin Khoo and Joshua Knobe address this worry. Drawing on experimental findings that propose contextualist “non-exclusionary” semantics for moral statements, they show how parties who aren’t disagreeing about the facts can nonetheless be disagreeing (Khoo and Knobe 2016).

Sarkissian and colleagues’ findings have at least two notable implications. The first is for experimentalists. Their study highlights the need to use experimental designs that better pull apart the predictions of realist versus relativist views, something that hadn’t always been accomplished in previous studies. The second implication is for philosophers. Sarkissian and colleagues’ findings call into doubt the claims of philosophers cited earlier that the folk are moral objectivists, and this point is discussed further below.

Motivational Internalism

Another area in metaethics where philosophers have frequently made appeals to common sense concerns the issue of motivational internalism. There are many internalist theses in

metaethics (see David Faraci and Tristram McPherson's chapter "Ethical Judgment and Motivation" and Errol Lord and David Plunkett's chapter "Reasons Internalism" for discussion). Motivational internalism (hereafter "internalism") posits a necessary connection between moral judgments and motivation. According to this thesis, if a person makes a moral judgment that she should *A*, then necessarily the person has at least some degree of motivation in favor of *A*-ing. Most theorists think this bare statement of the view needs to be qualified in various ways, though there is no consensus on how. The two most common qualifications proposed are that the person must be *rational* and *psychologically healthy*, but a number of others have been put forward (see Björklund et al. 2012 for a review).

Some philosophers have claimed that internalism is part of the ordinary, shared understanding of moral judgment.

It seems to be a conceptual truth that to regard something as good is to feel a pull towards promoting or choosing it, or towards wanting other people to feel the pull towards promoting or choosing it.

(Blackburn 1984: 188)

Other philosophers have denied that common sense supports internalism (e.g. Brink 1986) and some have even suggested that externalism, i.e., the denial of internalism, is supported by common sense (see, for example, Svavarsdottir 1999).

Against the background of this dispute, a body of experimental work has developed over the last decade or so assessing whether folk opinion supports internalism. In an early and influential study, Nichols (2002) presented participants with the following vignette.

John is a psychopathic criminal. He is an adult of normal intelligence, but he has no emotional reaction to hurting other people. John has hurt and indeed killed other people when he has wanted to steal their money. He says that he knows that hurting others is wrong, but that he just doesn't care if he does things that are wrong.

Participants were then asked whether "John really understands that hurting others is morally wrong?" Results showed that 85 percent of the subjects answered "Yes" and 15 percent answered "No." Nichols took these results to provide some initial support for externalism.

What about the additional qualifications that internalists often insist on? Nichols' study seems best positioned to address the rationality qualification as John is described as an adult with normal intelligence. In describing John as a psychopath, however, it is possible that participants infer he has a serious mental disorder. This would violate the qualification that some internalists insist on that the relevant agent be psychologically healthy.

Strandberg and Björkland (2013) performed a study that probed in more detail whether the folk endorse internalism, directly comparing a scenario in which the agent is psychologically healthy from scenarios in which she is psychologically impaired. In their study, all participants were assigned to one of five groups. The Simple group read the following story.

Anna is watching a TV program about a famine in Sudan. In the TV program, it is shown how the starving are suffering and desperately looking for food. At the

same time, Anna is not motivated at all, not to any extent, to give any money to those who are starving.

Participants in the Normal Functioning group next read the following additional information.

Anna is mentally healthy and functions normally. For example, she is not depressed, apathetic, emotionally disturbed, psychopathic, or the like.

Participants in the other three groups read information describing a person with severe apathy, depression, and psychopathy, respectively. Finally, all participants received the following question: Could it be the case that Anna thinks she is morally required to give some of her money to the starving even if she is not motivated at all to do so?

Results showed the following percentages of people who answered affirmatively to this question:

Simple: 75.6 percent

Normal Functioning: 79.0 percent

Apathy: 60.0 percent

Depression: 79.4 percent

Psychopathy: 42.3 percent

The results for the first four conditions suggest the folk have externalist leanings. Moreover, the fact that participants gave a similar percentage of affirmative responses to the Simple condition compared to the Depression condition provides evidence against versions of internalism that require the agent to be psychologically healthy.

Why did participants tend to answer “No” more frequently in the Psychopathy condition? Strandberg and Björkland propose that their participants tended to think of psychopaths as persons who do not hold moral judgments at all. That is, they propose that participants think it is not possible that a psychopath ever thinks she is morally required to perform an action. If this is right, the fact that participants tended to answer “No” in this condition does not provide any support for the internalist.

One potential problem with the studies by Nichol and by Strandberg and Björkland is that they asked whether the character in the vignette *really understands* moral claims (Nichols) or whether the character could *think* something is morally required and remain unmotivated (Strandberg and Björkland). Björnsson and colleagues (2015) note that internalism is a thesis that concerns the necessary upshots of moral *judgment*, understood as a species of *belief*, rather than the upshots of understanding and thinking.

Based on this observation, Björnsson and colleagues constructed scenarios that manipulated the character’s attitude to the relevant proposition, i.e., whether the character understands, thinks, or believes that something is morally wrong. Their vignettes are far more detailed than typical experimental philosophy studies and the results of their careful study resist easy summary, and so I will focus on one aspect. They presented participants with a detailed vignette about a character who makes correct classifications of actions as being morally right and morally wrong, but she remains entirely unmotivated to act accordingly. Participants were then randomly assigned to three conditions. In each

condition, they were asked, respectively, whether the character “understands,” “herself thinks,” or “believes” that the relevant action is morally wrong. Results for these three conditions are shown below:

Understands: 76 percent
 Herself thinks: 49 percent
 Believes: 46 percent

Similar to Nichols’ study discussed earlier, they found the majority of participants (76 percent) in the Understands condition responded affirmatively. Notice, however, that results differed substantially in the other two conditions. In the Believes condition in particular, which Björnsson and colleagues contend (plausibly) is most relevant for the philosophical thesis of internalism, participants were essentially evenly split. These results suggest the folk are quite divided about internalism and don’t appear to provide much support for either internalism or externalism.

Implications of the Sociological Project for Philosophical Debates

Let us suppose that years in the future, substantial experimental philosophy research on the preceding topics has been completed and scholarly consensuses on key positions emerge. What impact might these findings have on philosophical debates such as the debates about moral objectivity and internalism?

Recall earlier it was pointed out that philosophers frequently assert that certain claims are intuitive or are commonsense as a way of supporting their theories. Experimental philosophy findings, specifically those coming from the sociological project, provide empirical tests of such assertions. Where the empirical findings disagree with such assertions, a (purported) source of support for the relevant theories is thereby challenged. Where the empirical findings agree with such assertions, a (purported) source of support for the relevant theories is thereby corroborated.

Critics of experimental philosophy sometimes say it tries to settle philosophical matters by surveying the masses; it trades polls for substantive philosophical arguments. This critique misses the mark, however, at least when the project is pursued carefully, in the way being advocated for here. Theorists pursuing the sociological project in experimental philosophy should not argue (and as far as I know, they mostly haven’t) that their work provides evidence *directly* for or against philosophical views. Rather, their aim should be to scrutinize one kind of (purported) evidence put forward for those views, i.e., appeals to intuitions and commonsense judgments that are claimed to be shared by the folk.

THE PSYCHOLOGICAL PROJECT

The sociological project looks at the distribution of intuitions and commonsense judgments within and across societies. The psychological project in experimental philosophy, in contrast, investigates psychological structure. In particular, it studies the mentally represented information structures that give rise to intuitive judgments.

To get started in understanding this project, consider the following case proposed by Edmund Gettier that is widely viewed as a counterexample to the “justified true belief” analysis of knowledge:

Suppose that Smith and Jones have applied for a certain job. And suppose that Smith has strong evidence for the following conjunctive proposition:

(d) Jones is the man who will get the job, and Jones has ten coins in his pocket.

Smith’s evidence for (d) might be that the president of the company assured him that Jones would in the end be selected, and that he, Smith, had counted the coins in Jones’s pocket ten minutes ago. Proposition (d) entails:

(e) The man who will get the job has ten coins in his pocket.

Let us suppose that Smith sees the entailment from (d) to (e), and accepts (e) on the grounds of (d), for which he has strong evidence. In this case, Smith is clearly justified in believing that (e) is true.

But imagine, further, that unknown to Smith, he himself, not Jones, will get the job. And, also, unknown to Smith, he himself has ten coins in his pocket.

(Gettier 1963: 122)

Though Smith has a justified true belief in (e), most people intuitively judge that he does not know (e). (Note: While early studies suggested some cross-cultural variation in intuitive judgments about Gettier cases [Weinberg et al. 2001], more recent studies performed by some of the same authors, among others, found these judgments were similar across cultural groups [Machery et al. 2015].) Notice further that it is notoriously hard to say why we make this judgment, that is, on what basis it was made. If asked, we would typically say it just seems to us that Smith does not know (e) or that it strikes us that way, but we can’t provide more information about how we ourselves came to form this judgment. Call this feature—the lack of conscious awareness of the basis of the intuitive judgment—*source opacity*.

There are interesting parallels between these observations about intuitions in the Gettier case and intuitions about the grammaticality of sentences, as can be seen from the following example (based on Anderson and Lightfoot 2000). Everyone knows that for the following sentence:

(1a) Kim is happy

you can reduce the “is” by changing the sentence to:

(1b) Kim’s happy

But in the following sentence, the “is” that follows Tim cannot be reduced:

(2a) Kim is happier than Tim is

Reducing this second “is” results in the following ungrammatical sentence:

(2b) *Kim’s happier than Tim’s

When people are asked what underlying principles they are applying in accepting (1b) but rejecting (2b), they are typically flummoxed. They are clearly applying systematic principles, but they cannot state their contents.

The actual principles being applied, which we know from linguistic and psycholinguistic investigations, turn out to be surprisingly complex. One of the principles says, very roughly: An “is” can be reduced except when it is on the left side of an extracted phrase (a phrase is “extracted” when it remains in the underlying syntactic description of the sentence but is absent in the surface representation). In (2a), the second “is” sits to the left of an extracted phrase, as prior to extraction, the sentence would have read: “Kim is happier than Tim is *happy*.” In (1a), the “is” does *not* sit to the left of an extracted phrase, and thus the change to (1b) is allowed.

This principle is a good example of what cognitive scientists call *tacit information*. The principle is applied in producing a judgment, but the content of the principle itself is not something that we can readily introspect or articulate. It is likely that judgments across various domains (grammar, physics, biological kinds, social cognition, etc.) are associated with distinct, sizable bodies of tacit information. I will use the term *tacit information structure* to refer to the collection of mentally represented packets of information that play an inferential role in underwriting a certain well-defined class of judgments. Importantly, this usage allows the units of information within a structure (which I will refer to as “principles”) to be encoded in a variety of different formats, for example, sentential representations, map-like representations, distributed representations, etc.

In the case of the Gettier intuition as well as other philosophically relevant intuitions, there have to be information structures that are utilized in generating the relevant intuitions. The alternative, it would seem, would be to say that the intuitive judgment arises by magic. Furthermore, it is plausible that the relevant information structures are, like in the case of grammar, tacit. When presented with hypothetical cases, we find ourselves making intuitive judgments, but we cannot readily articulate the contents of the information structures that are the basis for these judgments.

A primary aim of the psychological project in experimental philosophy is to uncover the tacit information structures that are utilized in generating certain philosophically relevant intuitions. It is claimed that by scrutinizing the contents of these structures, we can better assess the evidential value of the relevant intuitions. (Note: There is vigorous ongoing debate about whether philosophers actually do rely on intuitions as a source of evidence for their views. See Cappelen 2012 and Williamson 2007 for the anti-intuition position and Nagel 2012, and Sripada in preparation for responses.)

To provide a concrete illustration of the psychological project and how it can yield findings that can potentially advance philosophical debates, let us examine in some detail recent psychological investigations into “trolley problems.” These cases originated with Philippa Foot (1967) and Judith Jarvis Thomson (1985), and they have been extensively discussed by other philosophers. Joshua Greene was an early leader in empirical investigation of these problems, and though other scientists have subsequently joined in, Greene’s work is the focus in what follows.

Consider the following two cases:

Bystander

A runaway trolley is headed for five people who will be killed if it proceeds on its present course. The only way to save the five people is to hit a switch that will turn the trolley onto an alternate set of tracks where it will kill one person instead of five. Should you turn the trolley in order to save five people at the expense of one?

Footbridge

A runaway trolley is headed for five people who will be killed if it proceeds on its present course. You are standing next to a large stranger on a footbridge that spans the tracks. The only way to save the five people is to push this stranger off the bridge, onto the tracks below. He will die if you do this, but his body will stop the trolley from reaching the others. Should you save the others by pushing this stranger to his death?

Most people intuitively judge that it is permissible to turn the switch in Bystander, but it is not permissible to push the stranger to his death in Footbridge. But what are the underlying (tacit) information structures that give rise to these differing intuitive judgments? Greene's research program uses a variety of empirical methods to answer this question and in turn uses the answers generated to inform philosophically relevant conclusions.

Let me start by sketching the form of one of Greene's main arguments and then the individual premises will be scrutinized. Greene himself helpfully offers a three-step argument along the following lines (based on Greene 2014a; Greene in preparation):

1. Intuitions in response to Footbridge are strongly influenced by the presence of personal force (i.e., being the initiator of the causal force that harms another person).
2. The presence of personal force is morally irrelevant to the moral acceptability of what one should do in this case.
3. Conclusion. Since intuitions in response to Footbridge are strongly influenced by at least one morally irrelevant factor, personal force, they are therefore at least somewhat unreliable.

Before going forward, it is worth underscoring that the preceding is *one of* the main arguments that Greene has defended (he calls it the "Direct Route" argument, see Greene 2014a). There is in addition a substantially more ambitious argument that Greene has advanced that he calls the "Indirect Route" argument. This argument seeks to distinguish cases that elicit "characteristically deontological" intuitions from cases that elicit "characteristically consequentialist" intuitions. Greene understands these two categories of intuitive judgment as something like natural kinds, with distinctive psychological and neural etiologies and distinctive features of cases that they track. The goal of the Indirect Route argument is not simply to challenge the reliability of intuitions associated with individual cases (e.g., the Footbridge case) but rather to challenge deontological theories writ large (Greene's argument for this conclusion is complex and cannot be sketched here). Going forward, the more ambitious (and substantially more controversial) Indirect Route argument is put aside. The Direct Route argument provides a better illustration of the pattern of reasoning that occurs as part of the psychological program in experimental philosophy, so this argument will be our focus.

Returning now to the three-step argument presented above, Greene and his colleagues have conducted a number of studies in support of premise 1. In one set of studies (Greene et al. 2009), they presented participants with a series of systematically varied vignettes in order to identify the factors that might be driving the different judgments in Footbridge and Bystander. One factor that might influence people's intuitive judgments in Footbridge is physical contact, having to lay one's hands on another person and push him to his death. Another factor is what Greene calls "personal force," being the initiator of the causal force that harms another person. Still another potential factor is spatial proximity, being spatially near to the site where a horrible harm occurs.

Participants were presented with a series of scenarios in which these factors were selectively present or absent. In one scenario, the agent on the footbridge can only use a pole to knock the stranger off the bridge, and like in the original Footbridge case, the stranger's body stops the runaway trolley. In this scenario, the agent doesn't put his hands directly on the stranger (physical contact is absent), but he is the initiator of the causal force that harms the stranger (personal force is present). In another pair of scenarios, the agent has access to a switch that opens a trapdoor on the footbridge. If the switch is flipped, the stranger falls through and, like in the original Footbridge case, his body stops the runaway trolley. In one version of this "switch" scenario, the switch is right next to the stranger and in another version, the switch is far away from the footbridge; this pair of cases tests for the influence of spatial proximity. After tabulating results from across all the various scenarios, Greene and his colleagues found evidence that people's moral intuitions in these trolley problems are selectively sensitive to the personal force factor and not the physical contact and spatial proximity factors.

Why might we have moral intuitions that are highly sensitive to harms generated by personal force? Greene offers an evolutionary explanation in which there is a suite of relatively simple, "alarm-bell"-like emotions that serve to selectively detect such harms:

Given that personal violence is evolutionarily ancient, predating our recently evolved human capacities for complex abstract reasoning, it should come as no surprise if we have innate responses to personal violence that are powerful but rather primitive... In contrast, when a harm is impersonal, it should fail to trigger this alarm like emotional response.... As Josef Stalin once said, "A single death is a tragedy; a million deaths is a statistic." His remarks suggest that when harmful actions are sufficiently impersonal, they fail to push our emotional buttons, despite their seriousness.

(Greene 2008: 43)

Greene supports this evolutionary speculation with a review (see Greene 2008) of a large psychological literature that finds a highly similar "personal" versus "impersonal" distinction made across a number of domains: people's decisions to help victims in distress; people's decisions to invest in collective action games; people's decisions to punish violators of moral rules; and people's attitudes towards so-called "harmless" norm violations. In all these domains, psychologists have posited that simple emotional responses lead people to treat harms that happen to identifiable, concrete victims substantially differently than seemingly identical harms that happen to abstract, unnamed, unidentifiable victims.

Another line of evidence for Greene's hypothesis comes from a heavily discussed fMRI study that Greene and his colleagues conducted (Greene et al. 2001). They found brain regions linked to emotional responses were implicated in hypothetical cases that involve personal harm, while brain regions linked to "cool" deliberative reasoning were involved in hypothetical cases that—like Bystander—involve impersonal harms.

To sum up, Greene attempts to support premise 1 of his argument by combining psychological, evolutionary, and neuroscientific arguments. Other theorists have pushed back, offering alternative hypotheses. One of the most influential perspectives is advanced by the philosopher and legal theorist John Mikhail, who explains differing judgments in Footbridge versus Bystander in terms of the application of the doctrine of double effect. This principle draws a distinction between (1) bringing about a good outcome while foreseeing that an unintended harm will occur as a side effect (which is permissible); and (2) causing that same harm in order to bring about a good outcome (which is not permissible). In a number of studies, some of which were done in collaboration with psychologists and neuroscientists, Mikhail assembles a sophisticated and convergent theoretical and empirical case for his position (Mikhail 2007; Mikhail 2011). Greene, in turn, has offered trenchant replies (Greene 2014b), and the debate is ongoing.

For the purpose of continuing to explicate Greene's argument, let us assume that he is right about premise 1. Premise 2 of Greene's argument is a normative premise that Greene unabashedly grants is got from the armchair. According to Greene, it is obvious, simply by reflecting on the issue, that personal force ought not to be a morally relevant consideration. That is, we should not judge that an action that sacrifices one to save five is impermissible *only* because it involves personal force-type harm (while an otherwise similar action that does not involve personal force is judged to be permissible).

Finally, the conclusion of Greene's argument regarding the reliability of intuitions in Footbridge follows fairly directly from the premises. If the conclusion of Greene's argument is correct, this has implications for certain debates in normative ethics. Trolley cases along the lines of Footbridge are often used in arguments against utilitarianism. Utilitarianism would seem to predict that we should sacrifice the one to save five in Footbridge just as we should in Bystander. The fact that utilitarianism can't capture our intuitions about the case, it is argued, counts against the view. If the conclusion of Greene's argument is correct and intuitions in Footbridge are shown to be at least somewhat unreliable, then this helps to undercut one important source of evidence against utilitarian theories. Theorists relying on this source of evidence should, other things being equal, correspondingly increase their credence in utilitarian theories.

One sometimes hears criticisms of Greene that say he is attempting to extract a normative conclusion from empirical premises. This critique of Greene is misguided because premise 2 of his argument is straightforwardly a normative premise, not an empirical one. (It is in fact a normative premise that Greene says he gets from the armchair.) So there is no obvious way in which Greene is deducing an *ought* exclusively from an *is*.

Another objection is that Greene's argument is somehow circular or question begging. The objection begins by noting that Greene is employing a normative claim as premise 2 of his argument (and thus it is conceded he is not deducing an *ought* from an *is*). It is then argued that since Greene already presupposes a normative claim as a premise, any normative conclusions that follow from the argument beg the question.

The problem with this argument is that it fails to notice that premise 2 of Greene's argument relies on a normative claim that is quite a bit less controversial than the normative conclusions that follow from the argument. The normative claim in premise 2 is that personal force is not a morally relevant factor. This is a claim that has support given a wide range of plausible normative theories, including utilitarian and deontological theories. The conclusions of Greene's argument involve normative claims that are quite a bit more controversial. For example, earlier it was shown that if Greene's argument is correct, then at least some theorists (those who rely on intuitions in Footbridge as evidence against utilitarianism) should increase their credence in utilitarianism. In short, Greene's argument is not question begging because the normative premises are different from, and in an important sense "less weighty" than, the normative conclusions that the argument supports.

Perhaps the best-known critique of Greene's research was put forward by Selim Berker in his article "The Normative Insignificance of Neuroscience" (Berker 2009). It turns out, however, that Berker's argument has relatively little relevance to the aspects of Greene's research discussed here for two reasons. First, Berker's critique targets Greene's more ambitious (and more controversial) Indirect Route argument. We, however, have focused exclusively on Greene's more modest Direct Route argument, since the pattern of reasoning in this argument is more representative of the psychological project in experimental philosophy. Second, Berker is concerned specifically with Greene's *neuroscience* results. As we have seen, however, these results play at best a small (and largely inessential) role in the Direct Route argument, while Greene's psychological results play a much larger role. Readers interested in a detailed response to Berker's critique should consult Greene (in preparation).

From Greene's three-step argument, we can extract a more general three-step template for how the psychological program in experimental philosophy works. Start with the following background: Some philosophical debate is unfolding and theorists on one side appeal to an intuitive judgment *I* in order to advance their position. *I* exhibits a high degree of source opacity; there are mentally represented principles that are applied in producing *I* but the theorists can't articulate their contents. The first step in the psychological program is to use experimental methods from psychology, neuroscience, or other fields in the behavioral sciences, in order to identify these mentally represented principles that are the basis for *I*. The next step is to subject the contents so identified to reflective scrutiny. For example, are the contents reasonable in light of other principles, commitments, considered judgments about cases, etc.? The third step consists of adjustment of the evidential weight of the relevant intuitions upwards or downwards. For example, if the contents identified in the second step are found to be unjustifiable, then the intuitions that are rooted in these contents should commensurately lose some of their evidential status. This adjustment can in turn change theorists' credences in philosophical theories.

Thus far, we have been focusing on the information structures that are the inferential basis of intuitive judgments. There are, however, a variety of *other* factors that also influence one's intuitive judgments that are not naturally understood to be elements of the information structures associated with the judgment process. For example, one's judgments about whether to push the man in the Footbridge case might be influenced by one's mood (happy or depressed), state of inebriation, attitudes to the race of the person being pushed, and so on. An additional goal of the psychological program in experimental philosophy is to characterize the effects of factors such as these on intuitive judgment. Just like factors that are part of an intuition's information basis, uncovering the influence

of these “extraneous” factors on an intuition can, through the three steps outlined in the preceding paragraph, lead to adjustment of the intuition’s evidential worth. Information structures, nonetheless, continue to be emphasized going forward for ease of exposition, with the understanding that what is said applies equally well to influences exerted by these “extraneous” factors.

A critic of the psychological program in experimental philosophy might argue that empirical research is unnecessary; it is possible to uncover the contents of the underlying information structures that are the basis for intuitive judgments from the armchair alone. There are two versions of this argument. The first version says we can discover these contents from the armchair simply by “looking inside” our own minds. The information structures are represented in the head, so we can just introspect and see what their contents are. The problem with this argument is that, as the discussion of the Gettier case underscores, the kinds of mentally represented information structures at issue are *tacit*. They are not readily introspectable or available for verbal report.

The second version of this argument says theorists can pursue a quasi-experimental approach from the armchair. Consider how linguists make progress. They frequently test their own (armchair) intuitions systematically across a large number of cases, operating with the assumption that their own intuitive reactions are widely shared by the folk (or at least that they are good at predicting folk intuitive reactions; see Dunaway, Edmonds, and Manley 2013). They then use this armchair data to propose hypotheses about the mentally represented information structures that underlie grammaticality judgments. Why can’t philosophers do the same?

This second argument is on much sounder footing; philosophers often do make progress in uncovering tacit principles, commitments, and other informational elements from the armchair by using the quasi-empirical method. The problem with using this observation as part of an argument against the psychological project in experimental philosophy is that the argument assumes there is some kind of all-or-none competition: we must pursue the goal of uncovering the mentally represented information structures underlying intuitive judgments *either* from the armchair or from the lab. But there is no restriction that we pursue one or the other; we can pursue both.

Armchair methods have certain advantages. For example, the “subjects” in the experiments, i.e., philosophers themselves, tend to be reflective, have extensive background knowledge, and are motivated to engage deeply with cases.

Experimental methods have their own advantages. One of the most important is that they provide quantitative precision in measuring key variables, such as the strength of an intuitive response. Look at Greene’s study earlier using a series of cases to distinguish the role of physical contact, personal force, and spatial proximity in trolley problem judgments. Differences in intuitive responses to the cases are subtle, and it would be easy for a single philosopher in the armchair (especially one in the grip of a theory) to miss these subtleties. But by assembling hundreds of participants, Greene was able to identify clear quantitative differences in how people respond to these cases, which in turn points to a pivotal role for the personal force factor influencing their judgments. Experimental methods also have other advantages. For example, theorists can use a variety of methods (causal modeling, neuroimaging, reaction time measures) to understand how different psychological processes interact to produce the eventual judgment (Sinnott-Armstrong

et al. 2008; Sripada and Konrath 2011; Sripada 2012). This is something that cannot be easily achieved by from the armchair, for example by introspection.

Overall, since armchair methods and experimental methods each have advantages for pursuing the goals of the psychological project (i.e., uncovering tacit information structures that are the basis for intuitive judgments), and since we are under no pressure to pursue one or the other, then it seems reasonable that we should pursue them both.

Before concluding, it is worth raising an issue that is likely to be on the minds of many readers. Philosophers have been drawing on empirical observations for centuries, and in the last few decades, many moral philosophers have explicitly strived to propose theories that are richly informed by theories and findings from the natural sciences. For example, in “Wise Choices, Apt Feelings” (1992), Allan Gibbard proposes a theory of normative judgment that draws on evolutionary theory and psychological research on decision-making. In “Lack of Character” (2002), John Doris critiques virtue ethics based on a large body of findings from social psychology that seem to put the existence of morally relevant character traits in question.

On the conception of experimental philosophy put forward in this chapter, the preceding are examples of *empirically informed philosophy*, not experimental philosophy. The distinctive feature of experimental philosophy is that it seeks to better understand philosophically relevant intuitions specifically, especially their sociological distribution and psychological roots. This conception of experimental philosophy lines up with what at least some of its leading practitioners say (Alexander 2010; Nadelhoffer and Nahmias 2007), though others offer more expansive conceptions of the field (e.g., Knobe and Nichols 2008). Perhaps more importantly, it also seems to capture what most self-identified experimental philosophers actually *do*.

CONCLUSION

The core idea of experimental philosophy is that systematic investigation of intuitive judgment, using the methods of cognitive science, can be useful in making progress in philosophical debates. In this chapter, two major programs in experimental philosophy research, the sociological and psychological projects, were distinguished and their contributions to debates in moral theory were explored. The sociological project was illustrated with summaries of research on moral objectivity and motivational internalism. The psychological project was illustrated with research into the psychological basis of judgments in trolley problems conducted by Joshua Greene and his colleagues. Research by experimental philosophers into these topics, as well as other important issues in moral theory, is active and ongoing. Experimental philosophy is a relatively new field, but there is great promise that it can complement and enrich traditional armchair methods.

ACKNOWLEDGMENTS

Many thanks to Jamie Fritz, Tristram McPherson, and David Plunkett, whose extensive comments improved virtually every aspect of this chapter.

REFERENCES

- Alexander, J. (2010) Is Experimental Philosophy Philosophically Significant? *Philosophical Psychology*, 23(3): 377–389.
- Anderson, S.R. and Lightfoot, D.W. (2000) The Human Language Faculty as an Organ. *Annual Review of Physiology*, 62(1): 697–722.
- Berker, S. (2009) The Normative Insignificance of Neuroscience. *Philosophy & Public Affairs*, 37(4): 293–329.
- Björklund, F. et al. (2012) Recent Work on Motivational Internalism. *Analysis*, 72(1): 124–137.
- Björnsson, G. et al. (2015) Motivational Internalism and Folk Intuitions. *Philosophical Psychology*, 28(5): 715–734.
- Blackburn, S. (1984) *Spreading the Word*, Oxford: Clarendon Press.
- Brink, D.O. (1986) Externalist Moral Realism. *The Southern Journal of Philosophy*, 24(S1): 23–41.
- Cappelen, H. (2012) *Philosophy without Intuitions*, Reprint edition, Oxford: Oxford University Press.
- Doris, J.M. (2002) *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press.
- Dunaway, B., Edmonds, A. and Manley, D. (2013) The Folk Probably Do Think What You Think They Think. *Australasian Journal of Philosophy*, 91(3): 421–441.
- Foot, P. (1967) The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, 5.
- Gettier, E.L. (1963) Is Justified True Belief Knowledge? *Analysis*, 23: 121–123.
- Gibbard, A. (1992) *Wise Choices, Apt Feelings: A Theory of Normative Judgment*, Cambridge, MA: Harvard University Press.
- Goodwin, G.P. and Darley, J.M. (2008) The Psychology of Meta-Ethics: Exploring Objectivism. *Cognition*, 106(3): 1339–1366.
- Greene, J.D., in preparation. Notes on “The Normative Insignificance of Neuroscience” by Selim Berker. Available at: <https://joshgreene.squarespace.com/s/notes-on-berker.pdf>.
- Greene, J.D. (2014a) Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics. *Ethics*, 124(4): 695–726.
- Greene, J.D. (2014b) *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, Reprint edition, New York: Penguin Books.
- Greene, J.D. (2008) The Secret Joke of Kant’s Soul. In W. Sinnott-Armstrong, ed. *Moral Psychology, Vol 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press: 35–80.
- Greene, J.D. et al. (2009) Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment. *Cognition*, 111(3): 364–371.
- Greene, J.D. et al. (2001) An fMRI Investigation of Emotional Engagement in Moral Judgment. *Science*, 293: 2105–2108.
- Kelly, D. et al. (2007) Harm, Affect, and the Moral/Conventional Distinction. *Mind & Language*, 22(2): 117–131.
- Khoo, J. and Knobe, J. (2016) Moral Disagreement and Moral Semantics. *Noûs* (advance online publication).
- Knobe, J. and Nichols, S. (2008) *Experimental Philosophy*, Oxford: Oxford University Press.
- Machery, E. et al. (2015) Gettier Across Cultures. *Noûs* (advance online publication).
- Mikhail, J. (2011) *Elements of Moral Cognition: Rawls’ Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* 1st ed., Cambridge: Cambridge University Press.
- Mikhail, J. (2007) Universal Moral Grammar: Theory, Evidence and the Future. *Trends in Cognitive Sciences*, 11(4): 143–152.
- Nadelhoffer, T. and Nahmias, E. (2007) The Past and Future of Experimental Philosophy. *Philosophical Explorations*, 10(2): 123–149.
- Nagel, J. (2012) Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, 85(3): 495–527.
- Nichols, S. (2004) After Objectivity: An Empirical Study of Moral Judgment. *Philosophical Psychology*, 17(1): 3–26.
- Nichols, S. (2002) How Psychopaths Threaten Moral Rationalism: Is It Irrational to Be Amoral? *The Monist*, 85(2): 285–303.
- Nucci, L. (1981) Conceptions of Personal Issues: A Domain Distinct from Moral or Societal Concepts. *Child Development*, 52(1): 114–121.
- Sarkissian, H. et al. (2011) Folk Moral Relativism. *Mind & Language*, 26(4): 482–505.

- Shafer-Landau, R. (2003) *Moral Realism: A Defence*, Oxford: Oxford University Press.
- Sinnott-Armstrong, W. et al. (2008) Intention, Temporal Order, and Moral Judgments. *Mind & Language*, 23(1): 90–106.
- Smetana, J.G. (1981) Preschool Children's Conceptions of Moral and Social Rules. *Child Development*, 52(4): 1333–1336.
- Smith, M. (1994) *The Moral Problem*, Oxford: Blackwell Publishing.
- Sripada, C., in preparation. Intuitions, Tacit Theories, and Experimental Philosophy.
- Sripada, C. (2012) What Makes a Manipulated Agent Unfree? *Philosophy and Phenomenological Research*, 85(3): 563–593.
- Sripada, C.S. and Konrath, S. (2011) Telling More than We Know about Intentional Action. *Mind & Language*, 26: 353–380.
- Stich, S. and Weinberg, J.M. (2001) Jackson's Empirical Assumptions. *Philosophy and Phenomenological Research*, 62(3): 637–643.
- Strandberg, C. and Björklund, F. (2013) Is Moral Internalism Supported by Folk Intuitions? *Philosophical Psychology*, 26(3): 319–335.
- Svavarsdottir, S. (1999) Moral Cognitivism and Motivation. *The Philosophical Review*, 108(2): 161–219.
- Thomson, J.J. (1985) The Trolley Problem. *The Yale Law Journal*, 94(6): 1395–1415.
- Turiel, E. (1983) *The Development of Social Knowledge: Morality and Convention*, Cambridge: Cambridge University Press.
- Wainryb, C. et al. (2004) Children's Thinking About Diversity of Belief in the Early School Years: Judgments of Relativism, Tolerance, and Disagreeing Persons. *Child Development*, 75(3): 687–703.
- Weinberg, J.M., Nichols, S. and Stich, S. (2001) Normativity and Epistemic Intuitions. *Philosophical Topics*, 29: 429–460.
- Williamson, T. (2007) *The Philosophy of Philosophy*, Oxford: Blackwell Publishing.