

Experimental Philosophy

Joshua Knobe^{1,2}, Wesley Buckwalter³, Shaun Nichols⁴, Philip Robbins⁵, Hagop Sarkissian⁶, and Tamler Sommers⁷

¹ Program in Cognitive Science, Yale University; ² Department of Philosophy, Yale University; ³ Department of Philosophy, City University of New York, Graduate Center; ⁴ Department of Philosophy, University of Arizona; ⁵ Department of Philosophy, University of Missouri; ⁶ Department of Philosophy, City University of New York, Baruch College; ⁷ Department of Philosophy, University of Houston.

Abstract Experimental philosophy is a new interdisciplinary field that uses methods normally associated with psychology to investigate questions normally associated with philosophy. The present review focuses on research in experimental philosophy on four central questions. First, why is it that people's moral judgments appear to influence their intuitions about seemingly nonmoral questions? Second, do people think that moral questions have objective answers, or do they see morality as fundamentally relative? Third, do people believe in free will, and do they see free will as compatible with determinism? Fourth, how do people determine whether an entity is conscious?

Keywords: moral psychology, moral relativism, free will, consciousness, causation.

INTRODUCTION

Contemporary work in philosophy is shot through with appeals to intuition. When a philosopher wants to understand the nature of knowledge or causation or free will, the usual approach is to begin by constructing a series of imaginary cases designed to elicit prereflective judgments about the nature of these phenomena. These prereflective judgments are then treated as important sources of evidence. This basic approach has been applied with great sophistication across a wide variety of different domains.

Although this approach remains influential within the discipline of philosophy, it has inspired a growing ambivalence within the broader field of cognitive science. On the one hand, work using this approach has helped to shape a number of successful scientific research programs (Keil 1989, Rips et al. 2006, Xu 1997). On the other, there is a persistent worry that the key claims made about intuition are not being subjected to empirical testing

and that the approach as a whole is insufficiently attentive to psychological theories about how people's minds actually work ([Stich 2001](#)).

Experimental philosophy arose in part as a reaction to these worries. Experimental philosophers pursue the traditional questions of philosophy (free will, the mind-body problem, moral relativism), but they examine people's intuitions about these questions using the tools of contemporary psychology. Claims about intuition are tested in controlled experiments, and results are subjected to the usual statistical analyses. Most importantly, the patterns observed in people's intuitions are explained in terms of psychological processes, which are then explored using all of the usual methods: mediation analysis, developmental research, reaction time studies, patient studies, and so on.

At this point, it may seem natural to ask: "How exactly is the project of experimental philosophy, thus defined, distinct from that of social psychology?" The best answer is that this is precisely the sort of question that experimental philosophers want to reject. A guiding theme of the experimental philosophy movement is that it is not helpful to maintain a rigid separation between the disciplines of philosophy and psychology. Experimental philosophers explore issues that are central to traditional philosophical concerns, but in practice many papers in experimental philosophy are coauthored with psychologists, and many have been published in psychology journals. Much as in psycholinguistics or experimental economics, what we see emerging is an interdisciplinary research program in which philosophers and psychologists work closely together by combining the tools once thought native to each field in the pursuit of questions of renewed interest to both disciplines. (For contrasting perspectives on the more general nature of experimental philosophy, see [Alexander et al. 2009](#), [Knobe & Nichols 2008](#), [Nadelhoffer & Nahmias 2007](#), [Sosa 2007](#).)

Perhaps the best way to become acquainted with the field of experimental philosophy is to look in detail at the actual research findings. To illustrate the substantive contributions of experimental philosophy, this review focuses on research programs in four specific domains. Within each domain, recent work has involved a complex collaboration among philosophers and psychologists, and the resulting research draws on insights from both disciplines. Though research in each of the domains is concerned with a distinct substantive

question, our hope is that, together, they will serve to illustrate the general approach that has been characteristic of the experimental philosophy movement as a whole.

MORALITY AND CONCEPT APPLICATION

Moral deliberations about agents and actions often begin with a series of questions. Did the agent act intentionally or accidentally? Did the agent know what would happen when choosing a particular course of action? Was the agent causally responsible for the relevant outcome? People's answers to these questions frequently influence their moral judgments (Cushman 2008, Guglielmo et al. 2009).

However, one of the major findings in experimental philosophy is that this influence can also go in the opposite direction. Evaluative judgments, and in particular moral judgments, can themselves influence judgments about what was done intentionally, what agents know, and what agents cause.

Asymmetry in Folk Concepts

Perhaps the best-documented instance of the impact of moral judgment on concept application concerns the concept of intentional action. This body of research shows that people's moral evaluations about a particular action influence their judgments about whether that action was performed intentionally. To see this, consider the contrast between the following two vignettes (Knobe 2003).

(a) 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, and it will also help the environment." The chairman of the board answered, "I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program." They started the new program. Sure enough, the environment was helped. Did the chairman intentionally help the environment?

(b) 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, and 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.

Did the chairman intentionally harm the environment?

Participants presented with these cases make asymmetric intentionality judgments. Those who are given the help case typically say that the chairman helped unintentionally, whereas those who are given the harm case typically say that the chairman harmed intentionally. Yet it seems that the only major difference between the two cases lies in the moral status of the chairman's action. A broad array of researchers have therefore concluded that the moral status of the chairman's action is somehow affecting intuitions regarding whether or not that action was performed intentionally (Malle 2006, Nadelhoffer 2005, Nado 2008).

This effect has been replicated and extended in a number of subsequent studies (Cushman & Mele 2008, Feltz & Cokely 2007, Mallon 2008, Nadelhoffer 2005, Nichols & Ulatowski 2007, Phelan & Sarkissian 2008). Such studies show that the effect arises with different vignettes (Cushman & Mele 2008, Mallon 2008, Nadelhoffer 2005, Phelan & Sarkissian 2008), in different cultures (Knobe & Burra 2006), and in children as young as 3 years old (Leslie et al. 2006, Pellizzoni et al. 2009). It has also been shown that individual differences in moral judgment lead to corresponding differences in intuitions about whether an action was performed intentionally (Ditto et al. 2009).

It was originally thought that this asymmetry might be due entirely to certain aspects of the concept of intentional action. However, subsequent work indicates that the same basic effect arises for other concepts as well. Take the concept of knowledge. One can investigate the impact of moral judgments on ascriptions of this concept by simply giving participants the very same vignettes quoted above, but this time asking a different question:

Did the chairman know that the new program would help [harm] the environment?

Faced with this latter question, participants show the same asymmetry, indicating greater agreement with the knowledge ascription in the harm case than in the help case (Beebe & Buckwalter 2010). Just as for the concept of intentional action, this effect of moral judgment on knowledge attribution has also been found in a wide variety of other scenarios (Beebe & Jensen 2011, Buckwalter 2011a, 2011b).

Continuing research in this vein has shown impacts of moral judgment on people's use of numerous other folk-psychological concepts: desiring (Tannenbaum et al. 2009), valuing (Knobe & Roedder 2009), deciding (Pettit & Knobe 2009), weakness of will (May & Holton 2011), and happiness (Phillips et al. 2011). At this point, it is beginning to seem that

the effect found for intentional action is really just one symptom of a far more pervasive effect of moral judgment on the way that people understand each other's minds.

But the impact of moral judgment does not seem to be confined to the domain of folk psychology. Several studies indicate that moral considerations can also influence ordinary intuitions about causation (Alicke 2000, Buckwalter 2011b, Cushman et al. 2008, Hitchcock & Knobe 2011, Roxborough & Cumby 2009, Solan & Darley 2001). These studies show that an agent is more likely to be considered a cause of an event when that agent's action was first judged to be morally bad. For a simple example of this phenomenon (Knobe & Fraser 2008), consider the following vignette:

The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own. The administrative assistants typically take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly emailed them reminders that only administrative assistants are allowed to take the pens. On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk.

In this case of the missing pens, people typically agree with the statement that the professor caused the problem and disagree that the administrative assistant caused the problem. Yet the only thing that is different about the actions of both characters is their moral status, suggesting that moral judgments of the actions of the professor and the administrative assistant affect participants' judgments about the cause of the pen shortage.

In short, moral considerations appear to have a powerful and robust impact on the application of a wide variety of folk judgments that one might have expected to be quite independent of moral judgment.

Explanatory Theories

At this point, then, there is a considerable body of evidence indicating that moral judgments can in some way influence intuitions about what appear to be purely descriptive questions. The principal aim of continuing research on this topic is therefore to go beyond simply showing that the effect arises and provide a broader theory that can explain why it is arising.

Answers to this question fall into two broad groups. Distortion theories say that although people have an entirely nonmoral understanding of concepts such as intention, knowledge, and causation, there is some additional cognitive process that distorts people's intuitions and allows moral judgments to impact them (Adams & Steadman 2004, Alicke 2000, Ditto et al. 2009, Nadelhoffer 2006). By contrast, competence theories say that the impact of moral judgment revealed in these studies reflects people's fundamental way of making sense of the world (Cushman & Mele 2008, Halpern & Hitchcock 2011, Knobe 2010, Phelan & Sarkissian 2009). According to this latter group of views, there is no hidden nonmoral capacity that is distorted by moral factors. Instead, asymmetric application arises because morality informs a fundamental part of what it means to correctly apply these folk psychological and causal concepts. Much of the recent experimental work on these topics is devoted to testing hypotheses derived from specific views within either the distortion or competence theoretical framework. Although this has led to the development and profusion of theoretical proposals that invoke a wide variety of different cognitive processes, no single view has emerged unchallenged.

Beginning with the former group, some have suggested that the impact of morality arises because certain emotional or affective processes distort the normal application of these concepts (Malle 2006, Malle & Nelson 2003, Nadelhoffer 2006). The proposal is that, in the above chairman vignettes for instance, when the agent's action leads to harmful effects on the environment, this generates a negative emotional reaction that leads people to say that the agent intentionally harmed the environment. To put this hypothesis to the test, Young and colleagues (2006) conducted a study on patients with severe emotional deficits resulting from damage to the ventromedial prefrontal cortex (VMPC). Although such patients show highly unusual patterns of judgment when presented with moral decisions that are thought to rely on emotion (Koenigs et al. 2007), they showed no unusual behavior on the questions under discussion here. Just like normal participants, they tended to say that the chairman harmed the environment intentionally but helped unintentionally (Young et al. 2006). Such results cast doubt on the view that asymmetric judgments can indeed be explained by appeal to emotional response.

To address the challenge of these neuropsychological data, defenders of distortion theories have suggested that people's judgments are being distorted, not by emotion, but by

a desire to blame (Alicke 2008). This view holds that an immediate desire to blame the chairman for the bad outcome induces posthoc attributions of intentionality (for instance) in an attempt to justify prior assessments of blameworthiness. However, this kind of explanation also does not go unchallenged. A number of researchers have argued that although the effects observed in these cases seem to have some relation to morality, they do not have any special connection to blame in particular. First, reaction time studies demonstrate that people generally make judgments of blame after they make judgments regarding intentionality (Guglielmo & Malle 2010), suggesting that people actually make the intentionality judgments before they have even engaged in an assessment of blame. Second, studies involving the application of other folk psychological concepts such as knowledge have shown that the crucial asymmetry persists when the desire to blame is diminished (Buckwalter 2011a, Schaffer & Knobe 2011). Finally, the effects seem to emerge even when one looks at cases in which there is no opportunity for moral blame per se but only a tendency to conclude that the agent violated some other sort of norm or incurred some other sort of cost (Machery 2008, Uttich & Lombrozo 2010).

Accordingly, some researchers have concluded that it might be a mistake to understand these effects as arising from any kind of distortion. Instead, such theorists offer competence theories, according to which moral judgment actually figures in people's basic capacity for applying the relevant concepts (Cushman & Mele 2008, Halpern & Hitchcock 2011, Knobe 2010, Phelan & Sarkissian 2009). A variety of such theories have been proposed, but all of them seek to explain the relevant effects without appealing to a distorting influence of blame. Indeed, many of them assign no role to judgments of blame at all; they focus rather on some other sort of judgment (a judgment that the agent has violated a norm, or incurred a cost, or simply done something wrong; Knobe 2010, Machery 2008, Uttich & Lombrozo 2010). For example, one hypothesis is that, independent of anything about blame, people's judgments about norm violations can impact their counterfactual reasoning and that counterfactual reasoning plays an important role in the competence underlying causal intuitions (Halpern & Hitchcock 2011).

Yet this approach, too, has been met with criticism. Researchers have used structural equation modeling to show that the impact of condition on people's causal judgments can sometimes be mediated by blame attribution (Alicke et al. 2011). Although such an effect

could in principle be compatible with a competence theory, it is not predicted by any of the specific competence theories that have been developed thus far.

Summary

Work in experimental philosophy has provided strong evidence for the claim that moral considerations can impact the application of a number of important folk concepts. However, although a great deal of evidence has been amassed for and against theories attempting to understand this general phenomenon, no consensus has emerged.

Despite the theoretical progress in providing explanations for the observed moral asymmetry in concept application, the questions that originally framed the debate continue to occupy experimental philosophers: Could morality really be at the core of how people make sense of their world? Alternatively, could additional factors in association with moral considerations distort the normal application of concepts such as intentionality, knowledge, or causation, and if so, which factors? With the aid of more advanced techniques in the social sciences, research in experimental philosophy will undoubtedly lead to the further development of distortion and competence theories in an attempt to explain the observed impact of evaluative judgment on the application of these different folk psychological and causal concepts.

MORAL OBJECTIVISM AND MORAL RELATIVISM

Imagine two people having an argument. One claims that billiards is an exciting game to play, and the other claims that, quite the contrary, it is not. During the course of their argument each invokes many good reasons, each argues passionately and with conviction, yet at the end of it all they remain in stalemate. In such a case, we might conclude that there is no single fact of the matter---billiards is just exciting for some and dull for others. The whole question, one might say, is fundamentally relative.

Now suppose the two people move on to another topic---Venus's orbit. One claims that Venus orbits the sun faster than the Earth, while the other claims that it does not. Again, they argue to stalemate, with no resolution. This case seems different. Here, it seems correct to say that there is a single right answer, and so one of them must, in fact, be wrong. This second question might be said to be objective.

Now that we have at least a rough sense for these two categories, a question arises about the status of moral questions. Are moral questions entirely relative, like the question as to whether billiards is exciting? Or do moral questions have objective answers, like the question about the orbit of Venus? This issue has generated tremendous controversy in the philosophical literature, with some philosophers saying that moral questions are fundamentally relative (Dreier 1990; Harman 1975; Prinz 2007; Wong 1984, 2006) and others saying that moral questions are just as objective as the questions of science (Shafer-Landau 2003, Smith 1994).

Despite this continuing controversy about whether morality actually is relative or objective, researchers have shown a striking degree of consensus about how ordinary people see the issue. Both philosophers and psychologists have suggested that ordinary folks take moral claims to be objectively true (e.g., Brink 1989, Goodwin & Darley 2008, Mackie 1977, Nichols 2004a, Shafer-Landau 2003, Smith 1994). We refer to this as the thesis of folk moral objectivism.

The thesis of folk moral objectivism has played an important role in theoretical arguments both in philosophy and in cognitive science. But is the thesis correct? Experimental philosophers have conducted a range of studies to put it to the test. In one early experiment (Nichols 2004a), all participants were given a vignette about two people who held opposite views on a moral question:

John and Fred are members of different cultures, and they are in an argument. John says, “It’s okay to hit people just because you feel like it,” and Fred says, “No, it is not okay to hit people just because you feel like it.”

Participants were then asked to choose between three options: (a) It is okay to hit people just because you feel like it, so John is right and Fred is wrong. (b) It is not okay to hit people just because you feel like it, so Fred is right and John is wrong. (c) There is no fact of the matter about unqualified claims such as, “It’s okay to hit people just because you feel like it.” Different cultures believe different things, and it is not absolutely true or false that it’s okay to hit people just because you feel like it.

The first two options seem to accord with objectivism, whereas the third fits more with relativism. Just as one might predict, the majority of participants chose one of the first two options. So the results of this first study seemed to support the thesis.

However, more recent work suggests an unexpected and more complicated picture of folk morality. It simply does not seem to be the case that people in general show a strong and robust tendency to endorse objectivist claims about morality. Instead, the experimental results suggest that people's responses depend on a complex array of different variables, including the subjects' age and personality traits, the way they are asked about morality, and even the specific moral question at stake.

Age

Studies suggest that young children are objectivists about morality (Nichols & Folds-Bennett 2003, Wainryb et al. 2004). As early as the age of 5, children display greater intolerance of dissenting judgments or opinions when they concern moral matters as opposed to other matters, such as matters of taste and fact; indeed, children can be as objectivist about moral disagreements as they are about purely factual disagreements (e.g., disagreement about whether pencils fall down or shoot up when you drop them) (Wainryb et al. 2004). Other work shows a striking difference between children's judgments about matters of taste and their judgments about morality. When children are asked whether watermelon is "yummy for real" or just "yummy for some people," they respond that watermelon is only yummy for some people; but when children are asked a corresponding question about morality, they tend to reject the claim that certain actions are "simply good for some people"; they say that these actions are "good for real" (Nichols & Folds-Bennett 2003).

Although young children seem to consistently endorse objectivism, there appears to be a strong tendency for people's views to change over the course of development. As individuals enter adulthood, their commitment to moral objectivism sometimes falls away, and they come to respond more as relativists. Studies on college-aged adults show sizable minorities of relativists (Nichols 2004a). Individuals tend to embrace moral relativism in their late teens to early thirties, only to revert back to objectivism as they grow older; indeed, it seems the older one gets, the more objectivist one becomes (Beebe & Sackris 2010). It seems as though a person's commitment to moral objectivism is not fixed but instead ebbs and flows across the lifespan.

Personality

Recent work has also shown that one's meta-ethical commitments might be related to other facets of one's psychology---specifically, to one's personality traits. Some studies have suggested a correlation between being high on the personality trait of being open to new experience and embracing a form of moral relativism (Feltz & Cokely 2008). Relatedly, relativists score high on disjunctive thinking, which measures one's ability to unpack alternative possibilities when problem solving (Goodwin & Darley 2010). Relativists also tend to be tolerant of alternative points of view, as opposed to objectivists (Wright et al. 2008), and they were better able to explain these alternative points of view (Goodwin & Darley 2010). Taken together, these studies suggest that whether one is a moral objectivist will hinge upon a cluster of related personality traits, such as one's levels of tolerance and one's ability to imaginatively engage with differing perspectives; the higher one scores on these traits, the less likely one is to be an objectivist about morality.

Framing of the Issue

A common feature of many of these studies is that they use a disagreement task to probe people's meta-ethical commitments. The method presents subjects with individuals who have differing judgments on some moral matter and then ask whether these individuals can both be correct. Recent work has suggested that an important variable in these studies is whether the individuals who have differing judgments belong to the same culture. If so, then subjects seem to think that one of them must be wrong---that two individuals of the same culture can't disagree about a moral issue without one of them being mistaken. However, people's intuitions undergo a systematic shift as they begin considering individuals of radically different cultural backgrounds. As they come to think about individuals who are deeply dissimilar---individuals with radically different cultures, values, or ways of life---people shift away from objectivism and tend to think that the disagreeing individuals can both be correct (Sarkissian et al. 2011). Thus, how moral disagreement is framed can be an important variable in gauging folk views about morality.

Specific Moral Issue

Finally, even though previous studies have found overall high mean levels of objectivism about moral issues, a closer inspection of the pattern of results reveals a great deal of variation according to the moral issue being considered. For example, even while Goodwin

& Darley (2008) found high mean levels of objectivism about moral issues, a number of particular issues garnered extremely low scores of objectivism; some moral transgressions (such as cheating on an exam or opening gunfire in a crowd) seemed to be deemed objectively wrong, whereas other transgressions garnered far lower scores of objectivism. In fact, some of the most highly charged and divisive moral issues of recent times (such as abortion, assisted suicide, and stem cell research) yielded very relativistic responses. For these latter issues, individuals tended to allow that individuals with differing moral judgments might both be correct. Further work is needed to understand what makes certain moral issues seem more objective than others.

Discussion

At least at first glance, these experimental results seem to spell trouble for the thesis of folk moral objectivism. After all, if participants had been asked a simple question about, say, whether a certain English sentence was grammatically correct, we would have expected to find a strong and robust consensus, with almost no variance in responses and very little impact of subtle experimental manipulations. But that is not what one finds in the case of questions about moral objectivism. Instead, the experimental results show powerful effects of both individual differences and experimental manipulation, with certain people under certain circumstances giving seemingly objectivist answers and other people under other circumstances giving seemingly relativistic answers. How can we explain these results?

One hypothesis is that, despite what we see in people's explicit responses, the traditional view in philosophy and psychology was actually right all along (see, e.g., Nichols 2004b). Perhaps people have a core capacity for understanding morality---a capacity whose workings we see coming out clearly in developmental studies---and this core capacity yields an understanding of morality as objective. Later on, people can develop explicit theories according to which morality is relative. Nonetheless, it might be that these explicit theories override a more immediate understanding of morality that retains its objectivist core.

Alternatively, it might be thought that the experimental data correctly reveal the actual nature of people's moral understanding. If so, perhaps the best way to make sense of the results is to deny that there is a fact of the matter as to whether people are moral objectivists or moral relativists. On some occasions, and with regard to some issues, people may give objectivist responses, and on other occasions, and with regard to other issues, they may give

relativist responses. There might then be no straightforward answer to the question, “Are people moral objectivists?” The real question would be about which factors can draw people to one view or another.

For example, looking across the studies above, those responding as relativists seem to share certain features in common: They tend to be in their late teens to early thirties, are open to new experiences, are willing to engage with diverse ways of life, and are tolerant of people with opposite opinions. One possible hypothesis would be that all of these different findings are explained by a single underlying process. Specifically, it might be that there is a general effect whereby people become more inclined to endorse relativism to the extent that they are more inclined to open their minds to alternative perspectives ([Sarkissian et al. 2011](#)).

To decide between these opposing explanations, it may be necessary to adopt new methodologies that allow us to look not only at people’s final conclusions but also at the psychological processes that lead up to those conclusions. Such work could further illuminate the patterns of intuition observed in studies thus far.

FREE WILL

In 1924, Clarence Darrow defended Nathan Leopold and Richard Loeb for the kidnapping and murder of their 14-year-old schoolmate Bobby Franks. Because the defendants had pled guilty to the crime, Darrow’s task was to save them from the death penalty. The challenge was finding a basis for mitigation. Leopold and Loeb were rich, healthy, and highly educated; they seemed to have every advantage young men could have. So Darrow appealed to the only mitigating factor that was available: the deterministic nature of the universe itself. “Your Honor,” Darrow said during his famous 12-hour closing statement, “Why did they kill little Bobby Franks? Not for money, not for spite; not for hate... They killed him because they were *made* that way. Because somewhere in the infinite processes that go to the making up of the boy or the man something slipped, and those unfortunate lads sit here hated, despised, outcasts, with the community shouting for their blood” (Darrow 1988).

Darrow’s defense here touches on one of the oldest and most controversial questions of philosophy. If a person’s actions are completely determined, can that person still be morally responsible for what he or she is doing? This question, in various guises, has obsessed

philosophers since at least the time of the Ancient Greeks, and settling on an answer is just as difficult today as it was in fifth century B.C. The absence of a satisfactory resolution after all this time suggests that people's intuitions about free will are deeply conflicted. It seems that one set of intuitions leads us to attribute free will and moral responsibility to agents who meet appropriate conditions even if their actions are the result of deterministic processes, whereas another pulls us toward withdrawing these attributions once we recognize that the causes of behavior do not originate ultimately within the agent. Experimental philosophers have sought to get at the psychological roots of the free will debate by examining the underlying causes of this conflict. The aim is to arrive at a better understanding of the factors that can draw people's intuitions toward one side or the other.

One such factor is emotional salience. Even in cases where an agent's behavior is entirely determined, people appear to be inclined to ascribe moral responsibility as long as the behavior elicits a strong emotional response. Thus, in one study ([Nichols & Knobe 2007: 669](#)), all participants were asked to imagine a deterministic universe.

Imagine a universe (Universe A) in which everything that happens is completely caused by whatever happened before it. This is true from the very beginning of the universe, so what happened in the beginning of the universe caused what happened next, and so on right up until the present. For example one day John decided to have French Fries at lunch. Like everything else, this decision was completely caused by what happened before it. So, if everything in this universe was exactly the same up until John made his decision, then it had to happen that John would decide to have French Fries.

Participants were then assigned to either a concrete high-affect condition or to an abstract low-affect condition. In the low-affect condition, participants were simply asked if people in Universe A could be fully morally responsible for their actions in this deterministic universe. Here, a large majority (86%) of the subjects answered "no." In the high-affect condition, participants read about a specific man named Bill in Universe A who burns down his house, killing his wife and three children, so that he can be with his secretary.

Participants were then asked whether this specific man was fully morally responsible for his behavior. In this condition, 72% of the subjects answered "yes." These results suggest one possible explanation for the intractability of this age-old problem. When we consider the problem abstractly, one set of cognitive processes leads us to the conclusion that

determinism is incompatible with free and responsible action. But cases like the story about Bill trigger a different set of processes that dispose us to assign blame and responsibility for terrible crimes and worry less about how they were caused.

A related factor influencing free will judgments is psychological distance, the distance (either in space or time) between subjects and the event or object and events they are considering. [Weigel \(2011\)](#) asked participants to imagine hearing a lecture about a deterministic universe; participants were then asked if a murderer in this universe acted freely. Some participants were assigned to a condition in which the lecture on determinism was taking place in a few days; others were assigned to a condition in which the lecture was taking place in a few years. This seemingly small manipulation had a significant effect. The results showed that subjects were less inclined to say that this man freely decided to kill when they imagined hearing about it at a more distant time. Research on psychological distance suggests that greater distance triggers cognitive processes that deal with questions more abstractly, so these results lend support to the view that our conflicting intuitions on free will are the product of different cognitive processes (see also Roskies & Nichols 2011).

[Feltz & Cokely \(2009\)](#) adopt the same basic framework in their studies, but with the following twist: The authors investigate whether personality differences can affect intuitions on free will. Specifically, Feltz & Cokely (2009) predicted that subjects who were high in personal trait extroversion would be more likely to assign free will and moral responsibility to a murderer in the deterministic scenario. The results showed a significant correlation between extroversion and a willingness to attribute free will and responsibility for determined behavior. These results may also support the emotional salience model because extroverted people, due to their increased sensitivity to the social features of a scenario, may consider concrete cases less abstractly (and so have a greater affective response) than their introverted counterparts. Certainly, the results shed light on why philosophical reflection and debate alone have not led to more universal agreement about the free will problem.

In the above studies, participants are asked to imagine a world where human behavior is caused deterministically, but no detail is given about the nature of causes. Several studies have shown, however, that the type of causal explanation can influence free will and responsibility judgments. [Nahmias and colleagues \(2007\)](#) and Nahmias & Murray (2010) separated participants into two conditions: In one condition, the agents' decision-making is

described “in terms of neuroscientific, mechanistic processes”; in the other, decision-making is described “in terms of psychological, intentional processes.” They found that in both abstract and concrete cases, subjects found neuroscientific descriptions of decision-making to be more of a threat to freedom and responsibility than psychological ones. Nahmias and colleagues offer their results as evidence that participants are prone to confuse determinism with fatalism, the view that our conscious desires and deliberations does not causally influence our behavior and destiny. Because this is a mistake---determinism does not entail that our conscious deliberations are causally impotent---the authors conclude that folk intuitions might be more unified in favor of the view that we can be free and responsible as long as our actions are determined in the right way.

The common ingredient in all of the above studies is this: The more real and personal the case, the more prone we are to attribute free will and moral responsibility to agents even when their behavior is determined. At one end of the spectrum, we might imagine someone deliberately harming a family member or loved one. Few of us would withhold blame due to theoretical considerations about the deterministic nature of the universe. But when the case is more abstract, involving strangers in another time or universe, we find it less plausible to hold agents free and responsible when the causes of their actions trace back beyond their control. A few important questions remain, however. First, how should we regard the intuitions generated in high-affect cases? Should we view them as distortions of the folk concept of free will, or as reliable indicators of what we really believe? Second, what, if anything, can experiments like these tell us about the accuracy of our intuitions? In other words, can experimental philosophy shed light on the correct understanding of the relationship between free will and determinism? Following Nichols (2008), we may call these the descriptive and substantive questions, respectively.

The verdict on the descriptive question is mixed. Nichols and Knobe offer evidence suggesting that our judgments in the high-affect cases are the result of a performance error due to the distorting influence of our emotions. They tentatively conclude that the folk concept of responsibility is incompatible with the truth of determinism. Nahmias and colleagues take the opposite view: The performance error, they argue, occurs when people mistakenly assume that determinism rules out effective deliberation. Weigel argues for a middle position, holding that neither set of intuitions should be regarded as a distortion;

both reveal competencies with our concepts in different contexts. More philosophical analysis and experimental work are needed if we are to arrive at a confident resolution.

Thus far, we have been discussing empirical questions about the psychological roots of people's intuitions about free will. Some researchers, in addition, employ the results of these studies to address the substantive philosophical question of whether our beliefs in free will and moral responsibility are justified. The key suggestion here is that a proper understanding of the nature and origins of our intuitions will enable us to explain away or debunk the widespread belief that people can be free and responsible. It will show us that people's belief in free will arises from a psychological source that carries no warrant or justification and should therefore be dismissed as misleading. Sommers (2007), for example, argues that our beliefs in free will and moral responsibility are the product of adaptations formed in hunter-gatherer environments, and so there is no reason to think they reflect any kind of moral truth. Greene & Cohen (2004) argue that our responsibility judgments reflect a false but evolutionarily useful presupposition of a dualist agent-self. And Ross & Shestowski (2003) present evidence that our responsibility attributions are contaminated by a "dispositional bias" that overrates the influence of stable character traits and underestimates the power of situational factors to govern. All of the authors acknowledge the difficulty of overcoming these biases and beliefs given how deeply rooted they are in our psychologies. But like Darrow, they believe that when the stakes are high---in theories of criminal justice, for example---we are obligated to acknowledge the truth

The first thing to note about such debunking strategies is that they assume a particular answer to the descriptive question: namely, that most people regard free will and moral responsibility as incompatible with an accurate naturalistic understanding of human behavior. Otherwise, there is no basis for claiming that our current attributions of free will would (or should) change once we reject our false beliefs. Second, even granting these assumptions, we cannot yet infer that the rational response is to reject our assignments of free will and moral responsibility. Further argument is needed to show that rejecting the belief in free will and moral responsibility is preferable to revising our criteria for their application (Nichols 2007, Vargas 2007). For this, we need to take into account a wide range of factors, among them the practical implications of retaining or rejecting these concepts.

Several recent studies have been developed to explore these implications. Vohs & Schooler (2008) offer some evidence that denying free will may lead people to behave immorally, by providing “the ultimate excuse to behave as one likes.” Baumeister et al. (2009) expand on these results with a study that suggests that inducing disbelief in free will leads to an increase in aggression and a reduction in willingness to help. If the authors are correct, this may undermine the rationality of rejecting free will and moral responsibility, since the belief in the concepts would have important social functions (but see Nadelhoffer & Feltz 2007) Sommers 2010). However, because the studies are designed to test for short-term rather than long-term effects, it is not clear how worried free will skeptics should be about these results.

PHENOMENAL CONSCIOUSNESS

One of the oldest and thorniest questions in philosophy is the “problem of other minds.” How can we know that another entity has a mind? It seems clear that the person down the hall is capable of beliefs, intentions, and emotions, whereas the toaster in the kitchen is not capable of thinking or feeling anything at all, but what sorts of evidence can we use to tell the difference?

Just in the past few years, a number of researchers in both psychology and experimental philosophy have argued that this traditional question needs to be reformulated (Gray et al. 2007, Knobe & Prinz 2008, Robbins & Jack 2006). These researchers have suggested that it might be a mistake to suppose that there is one unified process involved in attributing a mind. Instead, they have suggested that there might be fundamentally different processes involved in attributing distinct psychological capacities.

In philosophy, it is common to distinguish between mental states that involve phenomenal consciousness and those that do not (Block 1995). Take the difference between the mental state “feeling upset” and the mental state “knowing that $2 + 2 = 4$.” The former state involves a certain kind of feeling or experience---there is something that it is like to feel upset at a particular time---whereas the latter does not directly involve any feeling or experience. Philosophers mark this distinction by saying that the former state involves phenomenal consciousness whereas the latter does not.

Recent work suggests that ordinary people also appreciate this distinction. Moreover, the work suggests that the judgment that an entity is capable of having states that involve phenomenal consciousness (e.g., feeling upset) is driven by different cues than the judgment that an entity is capable of having states that do not involve phenomenal consciousness (e.g., knowing that $2 + 2 = 4$). There might also be different mechanisms that are activated by these different sets of cues. In short, this research has proposed the hypothesis that phenomenal consciousness is special.

Two Dimensions of Mind

Initial evidence for the complexity of the folk conception of mind comes from an investigation of mental-state attribution by [Gray et al. \(2007\)](#). In a large-scale, online-survey-based study, participants provided ratings for a cast of characters (e.g., a normal adult, a child, a dog, a robot). Participants were presented with pairs of the characters and asked to rank them across a wide range of capacities: memory, planning, fear, pain, pleasure, and so forth.

Interestingly, the results did not show a single continuum whereby certain characters scored high on all capacities while others scored low on all capacities. Instead, a factor analysis revealed two distinct dimensions of mind---capacities for cognition, such as self-control and planning, and then, separately, capacities for phenomenal consciousness, such as pain and fear. Some characters (e.g., a baby) were rated relatively high on consciousness and low on cognition; others (e.g., God) showed the opposite pattern.

These results do not directly show that people think that there are two different kinds of minds---cognitive minds and conscious minds. But they do suggest that there are separate cognitive capacities for attributing states that have phenomenal consciousness and those that do not. This is precisely the suggestion of [Robbins & Jack \(2006\)](#), who also maintain that judgments of moral considerability depend preferentially upon judgments of the capacity for conscious experience (see also [Robbins 2008](#)), whereas judgments of moral responsibility have more to do with perception of a target's capacity for sophisticated cognition.

The studies by Gray and colleagues provide some evidence for this ([Gray et al. 2007](#); see also [Waytz et al. 2010](#)). In addition to ranking characters on a range of mental capacities, participants in that study also ranked characters on two dimensions of moral status: (a) moral agency, or the capacity to perform right or wrong actions and to be held accountable

as such, and (b) moral patency, or the capacity to receive such actions and to be given due consideration on that basis. Correlational analysis of the data revealed that these dimensions of morality were strongly positively correlated with different dimensions of mind---capacities for cognition and capacities for phenomenal consciousness, respectively. For example, entities judged to be high on cognition and low on consciousness (like God) tend to be categorized more or less exclusively as moral agents, whereas experientially rich but cognitively impoverished entities (like infants) tend to be categorized almost exclusively as moral patients (Gray & Wegner 2009, 2010). These studies of the relation between different dimensions of mentality and morality provide further support for the idea that there is a deep rift between attributions of states that have phenomenal consciousness and those that do not.

The Role of Embodiment

If attributions of phenomenally conscious mental states actually are deeply distinct from attributions of nonphenomenal states, it seems that these different sorts of attributions should rely on different cues. A considerable body of research suggests that one main cue people use to determine whether an entity has nonphenomenal mental states is its behavior (Johnson 2003). (Indeed, people will even attribute beliefs and goals to pictures of little triangles on a screen when these triangles are exhibiting the right sorts of behaviors; see Heider & Simmel 1944.) But what cues do people use to determine whether an entity has states with phenomenal consciousness? One possible answer is that the main cue here is not a matter of exhibiting certain behaviors---that it is instead a matter of having a biological body (Knobe 2011).

A natural way to test this hypothesis is to look at people's intuitions about entities that do exhibit complex rational patterns of behavior but that do not have the right sorts of bodies. For example, suppose that we look at attributions of mental states to corporations. A corporation might show all the right patterns of behavior (gathering information, reacting flexibly to achieve certain goals), but instead of having a biological body composed of flesh and blood, it is composed of far-flung committees and departments communicating with each other through emails and memoranda. The key question now is what mental states people will attribute to an entity like this one.

In one recent study ([Knobe & Prinz 2008](#)), participants were given a list of sentences ascribing mental states to corporations. Some of these sentences ascribed nonphenomenal mental states:

Acme Corporation believes that its profit margin will soon increase.

Acme Corporation wants to change its corporate image.

Other sentences ascribed states that involved phenomenal consciousness:

Acme Corporation is experiencing a sudden urge to pursue Internet advertising.

Acme Corporation is now experiencing great joy.

Participants tended to regard the ascription of conscious states as linguistically anomalous, whereas they had no qualms about ascriptions of nonphenomenal states. This asymmetry might be taken as evidence for the claim that the body plays a special role in attributions of phenomenal consciousness.

This first argument has been a controversial one. Subsequent work has shown that participants are willing to attribute certain kinds of apparently phenomenal states to corporations (e.g., “McDonald’s is feeling upset about the court’s recent ruling”) ([Arico 2010](#)). Cross-cultural studies show that the effect is significantly weaker in participants in Hong Kong than in participants in the United States ([Huebner et al. 2010](#)). Finally, on a more theoretical level, it has been suggested that the difference between corporations and individuals might have more to do with their differing patterns of behavior than with any difference in their embodiment ([Sytsma & Machery 2009a](#)).

However, additional evidence for the same hypothesis has been found in studies of mental-state attributions to artifacts, such as robots ([Huebner 2010, Sytsma & Machery 2009b](#)). Like corporations, robots lack a biological body, so by hypothesis they should elicit a similar pattern of attribution. And this does indeed appear to be the case. In one study ([Huebner 2010](#)), participants were given a brief vignette about a robot that behaves exactly like a human being in every way and asked whether they agreed with the statements:

He believes that triangles have three sides.

He feels happy when he gets what he wants.

As predicted, participants were willing to attribute the nonphenomenal state (belief) but unwilling to attribute the state involving phenomenal consciousness (feeling happy).

Finally, the embodiment hypothesis appears to find support in the results of the factor analysis described above (Gray et al. 2007). Participants saw newborn babies as having ample capacity for phenomenal consciousness but little capacity for nonphenomenal states, whereas they saw God as having the maximum possible capacity for nonphenomenal states but little capacity for phenomenal consciousness. This is exactly the result one would expect if one assumes that complex behavior is the main cue for attributions of nonphenomenal states whereas embodiment is the main cue for attributions of phenomenal consciousness.

Models and Mechanisms

Thus, a diverse range of data suggests that the attribution of phenomenal states is driven by different cues than the attribution of nonphenomenal states. What kinds of explanations are available for this striking pattern of results?

One proposal is that the findings reflect a deep difference in the mechanisms underlying the attribution of mental states. On this view, the capacity to attribute feelings to something rests on a functionally specialized mechanism, at least partially distinct from the mechanism responsible for the attribution of thoughts (Robbins & Jack 2006).

An alternative view is that there is really just one core mechanism that is responsible for the attribution of both nonphenomenal and phenomenal mental states. On this view, if an entity is identified as an agent with goals and thoughts, that will be sufficient to generate an inclination also to attribute feelings (Arico et al. 2011, Fiala et al. 2011). If that view is right, then when people deny conscious states to something they regard as an agent, it should be the case that at some deeper level, they really are inclined to attribute conscious states to the entity. A recent reaction time study provides some support for this---participants were significantly slower to deny conscious states to agents than to nonagents (Arico et al. 2011). If, however, there are entities---such as groups---for which people really have no inclination to attribute conscious states, then the single mechanism view must explain why people are willing to attribute ordinary nonphenomenal states such as goals and thoughts. One possibility is that attributions of goals and thoughts to groups are better interpreted as somehow figurative rather than literal ascriptions of goals and thoughts to groups (Phelan et al. 2011).

It remains quite unclear which of these approaches is correct, but the advent of experimental philosophy of consciousness has initiated a new way to investigate some

fundamental philosophical questions about how people ordinarily attribute psychological capacities.

CONCLUSION

This review has focused on four specific areas of research in experimental philosophy. Within each of these areas, one finds the emergence of an interdisciplinary conversation in which philosophers and psychologists work closely together to address a set of questions that lie at the intersection of the two fields.

Although we discuss work in four major areas, this is far from an exhaustive review. Experimental philosophers have also investigated cross-cultural differences in philosophical intuitions (e.g., [Machery et al. 2004](#), [Weinberg et al. 2001](#)), judgments about whether a person truly knows something as opposed to merely believing it (for a review, see [Pinillos 2011](#)), and people's ordinary conceptions of race ([Glasgow et al. 2009](#)). The possibilities for productive work in experimental philosophy are broad, and many of the topics explored by psychologists make contact with closely related philosophical issues. Future work in experimental philosophy could take the same interdisciplinary approach found in the four areas reviewed here and apply it to questions about causation, the self, religion, aesthetics, and elsewhere.

This interdisciplinary approach of the experimental philosophy movement has sometimes been characterized as a revolutionary new attack on the longstanding division between the disciplines of philosophy and psychology (e.g., [Appiah 2007](#), [Lackman 2006](#)). It seems to us that this characterization is not quite right. After all, philosophers have been concerned with psychological questions for thousands of years (think of the work of Plato and Aristotle), and this fluid boundary between philosophy and psychology persisted up through the twentieth century (think of William James). Perhaps then, the kind of interdisciplinary collaboration one sees in experimental philosophy is best understood not as a radical break with the past, but rather as a return to a more traditional conception of how philosophy and psychology should relate and develop.

DISCLOSURE STATEMENT

The authors are unaware of any affiliation, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

- Adams F, Steadman A. 2004. Intentional action in ordinary language: core concept or pragmatic understanding? *Analysis* 64:173--81
- Alexander J, Mallon R, Weinberg J. 2009. Accentuate the negative. *Rev. Philos. Psychol.* 1:297--314
- Alicke MD. 2000. Culpable control and the psychology of blame. *Psychol. Bull.* 126:556--74
- Alicke MD. 2008. Blaming badly. *J. Cogn. Cult.* 8:179--86
- Alicke MD, Rose D, Bloom D. 2011. Causation, norm violation and culpable control. *J. Philosophy*. In press
- Appiah KA. 2007. The new new philosophy. *N.Y. Times*. Dec. 9.
- Arico A. 2010. Folk psychology, consciousness, and context effects. *Rev. Philos. Psychol.* 1:371--93
- Arico A, Fiala B, Goldberg RF, Nichols S. 2011. The folk psychology of consciousness. *Mind Lang.* 18:327--352.
- Baumeister RF, Masicampo EJ, DeWall CN. 2009. Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personal. Soc. Psychol. Bull.* 35:260--62
- Beebe JR, Buckwalter W. 2010. The epistemic side-effect effect. *Mind Lang.* 25:474--98
- Beebe JR, Jensen R. 2011. Surprising connections between knowledge and action: the robustness of the epistemic side-effect effect. *Philos. Psychol.* In press.
- Beebe JR, Sackris D. 2010. Moral objectivism across the lifespan. *Exp. Philos. Soc. Meet. East. Div. Meet. Am. Philos. Assoc., Boston, MA*, Dec. 28
- Block N. 1995. On a confusion about the function of consciousness. *Behav. Brain Sci.* 18:227--47
- Brink DO. 1989. *Moral Realism and the Foundations of Ethics*. New York: Cambridge Univ. Press

- Buckwalter W. 2011a. Gettier Made ESEE. Unpublished Manuscript, City University of New York Graduate Center.
- Buckwalter W. 2011b. Experiments in Metaphysics, solving the puzzle of causation by absence. Unpublished Manuscript, City University of New York Graduate Center.
- Cushman F. 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108:353--80
- Cushman F, Knobe J, Sinnott-Armstrong W. 2008. Moral appraisals affect doing/allowing judgments. *Cognition* 108:353--80
- Cushman F, Mele A. 2008. Intentional action: two-and-a-half folk concepts? In *Experimental Philosophy*, ed. J Knobe, S Nichols, pp. 171--188. New York: Oxford Univ. Press
- Darrow C. 1988. *Clarence Darrow's Sentencing Speech in State of Illinois v. Leopold and Loeb*. Classics of the Courtroom Series. Minnetonka, MN: Prof. Educ.
- Ditto PH, Pizarro DA, Tannenbaum D. 2009. Motivated moral reasoning. In *Psychology of Learning and Motivation, Vol. 50: Moral Judgment and Decision Making*, ser. ed. BH Ross, ed. DM Bartels, CW Bauman, LJ Skitka, DL Medin, pp. —307--338. San Diego, CA: Academic
- Dreier J. 1990. Internalism and speaker relativism. *Ethics* 101:6--26
- Feltz A, Cokely ET. 2008. The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. McRae & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. 1771--1776 Austin, TX: Cognitive Science Society.
- Feltz A, Cokely ET. 2007. An anomaly in intentional action ascription: more evidence of folk diversity. *Proc. Cogn. Sci. Soc.* 1748.
- Feltz A, Cokely ET. 2009. Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Conscious. Cogn.* 18:342--50
- Fiala B, Arico A, Nichols S. 2011. On the psychological origins of dualism: dual-process cognition and the explanatory gap. *Mind Lang.* In press.
- Glasgow J, Shulman J, Covarrubias E. 2009. The ordinary conception of race in the United States and its relation to racial attitudes: a new approach. *J. Cogn. Cult.* 9:15--38

- Goodwin GP, Darley JM. 2008. The psychology of meta-ethics: exploring objectivism. *Cognition* 106:1339--66
- Goodwin GP, Darley JM. 2010. The perceived objectivity of ethical beliefs: psychological findings and implications for public policy. *Rev. Philos. Psychol.* 1:1--28
- Gray H, Gray K, Wegner D. 2007. Dimensions of mind perception. *Science* 315:619
- Gray K, Wegner D. 2009. Moral typecasting: divergent perceptions of moral agents and moral patients. *J. Personal. Soc. Psychol.* 96:505--20
- Gray K, Wegner D. 2010. Blaming God for our pain: human suffering and the divine mind. *Personal. Soc. Psychol. Rev.* 14:7--16
- Greene J, Cohen J. 2004. For the law, neuroscience changes nothing and everything. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359:1775--85
- Guglielmo S, Malle BF. 2010. The timing of blame and intentionality: testing the moral bias hypothesis. *Personal. Soc. Psychol. Bull.* 36:1635--47
- Guglielmo S, Monroe AE, Malle BF. 2009. At the heart of morality lies folk psychology. *Inquiry* 52:449--66
- Halpern J, Hitchcock C. 2011. Actual causation and the art of modeling. In *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, ed. R Dechter, H Geffner, J Halpern, pp. 383--406. London: College Publ.
- Harman G. 1975. Moral relativism defended. *Philos. Rev.* 84:3--22
- Heider F, Simmel G. 1944. An experimental study of apparent behavior. *Am. J. Psychol.* 57:243—59.
- Hitchcock C, Knobe J. 2011. Cause and norm. *J. Philos.* 106:587—612.
- Huebner B. 2010. Commonsense concepts of phenomenal consciousness: Does anyone *care* about functional zombies? *Phenomenol. Cogn. Sci.* 9:133--55
- Huebner B, Bruno M, Sarkissian H. 2010. What does the nation of China think about phenomenal states? *Rev. Philos. Psychol.* 1:225--43
- Johnson SC. 2003. Detecting agents. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358:549--59
- Keil FC. 1989. *Concepts, Kinds, and Conceptual Development*. Cambridge, MA: MIT Press
- Knobe J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63:190--94
- Knobe J. 2010. Person as scientist, person as moralist. *Behav. Brain Sci.* 33:315--29

- Knobe J. 2011. Finding the mind in the body. In *What's Next 2*, ed. M Brockman. New York: Vintage. In press.
- Knobe J, Burra A. 2006. Intention and intentional action: a cross-cultural study. *J. Cult. Cogn.* 6:113--32
- Knobe J, Fraser B. 2008. Causal judgment and moral judgment: two experiments. In *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity*, ed. W Sinnott-Armstrong, pp. —441—448. Cambridge, MA: MIT Press
- Knobe J, Nichols S. 2008. An experimental philosophy manifesto. In *Experimental Philosophy*, ed. J Knobe, S Nichols, Oxford Univ. Press.
- Knobe J, Prinz J. 2008. Intuitions about consciousness: experimental studies. *Phenomenol. Cogn. Sci.* 7:67--83
- Knobe J, Roedder E. 2009. The ordinary concept of valuing. *Philos. Issues* 19:131--47
- Koenigs M, Young L, Adolphs R, Tranel D, Cushman F, et al. 2007. Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature* 446:908—11.
- Lackman J. 2006. March 2. The X-Philes: Philosophy meets the real world. *Slate*
- Leslie A, Knobe J, Cohen A. 2006. Acting intentionally and the side-effect effect: “theory of mind” and moral judgment. *Psychol. Sci.* 17:421--27
- Machery E. 2008. Understanding the folk concept of intentional action: philosophical and experimental issues. *Mind Lang.* 23:165--89
- Machery E, Mallon R, Nichols S, Stich S. 2004. Semantics, cross-cultural style. *Cognition* 92:B1--12
- Mackie JL. 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin
- Malle BF. 2006. Intentionality, morality, and their relationship in human judgment. *J. Cogn. Cult.* 6:61--86
- Malle BF, Nelson SE. 2003. Judging mens rea: the tension between folk concepts and legal concepts of intentionality. *Behav. Sci. Law* 21:563--80
- Mallon R. 2008. Knobe versus Machery: testing the trade-off hypothesis. *Mind Lang.* 23:247--55
- May J, Holton R. 2011. What in the world is weakness of will? *Philos. Stud.* In press
- Nadelhoffer T. 2005. Skill, luck, control, and folk ascriptions of intentional action. *Philos. Psychol.* 18:343--54

- Nadelhoffer T. 2006. Bad acts, blameworthy agents, and intentional actions: some problems for jury impartiality. *Philos. Explorations* 9:203--20
- Nadelhoffer T, Feltz A. 2007. Folk intuitions, slippery slopes, and necessary fictions: an essay on Saul Smilansky's free will illusionism. *Midwest Stud. Philos.* 31:202--13
- Nadelhoffer T, Nahmias E. 2007. The past and future of experimental philosophy. *Philos. Explorations* 10:123--49
- Nado J. 2008. Effects of moral cognition on judgments of intentionality. *Br. J. Philos. Sci.* 59:709--31
- Nahmias E, Coates D, Kvaran T. 2007. Free will, moral responsibility, and mechanism: experiments on folk intuitions. *Midwest Stud. Philos.* 31:214--42
- Nahmias E, Morris S, Nadelhoffer T, Turner J. 2006. Is incompatibilism intuitive? *Philos. Phenomenol. Res.* 73:28--53
- Nahmias E, Murray D. 2010. Experimental philosophy on free will: an error theory for incompatibilist intuitions. In *New Waves in Philosophy of Action*, ed. J Aguilar, A Buckareff, K Frankish. Hampshire, UK: Palgrave-Macmillan
- Nichols S. 2004a. After objectivity: an empirical study of moral judgment. *Philos. Psychol.* 17:3-26
- Nichols S. 2004b. *Sentimental Rules*. New York: Oxford Univ. Press
- Nichols S. 2007. After incompatibilism. *Philos. Perspect.* 21:405--28
- Nichols S. 2008. How can psychology contribute to the free will debate? In *Psychology and Free Will*, ed. J Baer, J Kaufman, R Baumeister, London: Oxford Univ. Press
- Nichols S, Folds-Bennett T. 2003. Are children moral objectivists? Children's judgments about moral and response-dependent properties. *Cognition* 90:B23--32
- Nichols S, Knobe J. 2007. Moral responsibility and determinism: the cognitive science of folk intuitions. *Nous* 43:663--85
- Nichols S, Ulatowski J. 2007. Intuitions and individual differences: the Knobe effect revisited. *Mind Lang.* 22:346--65
- Pellizzoni S, Siegal M, Surian L. 2009. Foreknowledge, caring, and the side-effect effect in young children. *Dev. Psychol.* 45:289--95
- Pettit D, Knobe J. 2009. The pervasive impact of moral judgment. *Mind Lang.* 24(5):586--604

- Phelan M, Sarkissian H. 2008. The folk strike back; or, why you didn't do it intentionally, though it was bad and you knew it. *Philos. Stud.* 138:291--98
- Phelan M, Sarkissian H. 2009. Is the trade-off hypothesis worth trading for? *Mind Lang.* 24:164--80
- Phelan M, Arico A, Nichols S. 2011. Thinking things and feeling things: on an alleged discontinuity in folk metaphysics of mind. In press.
- Phillips, J, Nyholm, S, Liao, S. 2011. The ordinary concept of happiness. Unpublished manuscript. Yale University.
- Pinillos N. 2011. Some recent work in experimental epistemology. *Philos. Compass*
- Prinz J. 2007. *The Emotional Construction of Morals*. New York: Oxford Univ. Press
- Rips LJ, Blok S, Newman G. 2006. Tracing the identity of objects. *Psychol. Rev.* 113:1--30
- Robbins P. 2008. Consciousness and the social mind. *Cogn. Syst. Res.* 9:15--23
- Robbins P, Jack A. 2006. The phenomenal stance. *Philos. Stud.* 127:59--85
- Roskies AL, Nichols S. 2011. Bringing moral responsibility down to earth. *J. Philos.* In press
- Ross L, Shestowsky D. 2003. Contemporary psychology's challenges to legal theory and practice. *Nw. UL Rev.* 97:1081.
- Roxborough C, Cumby J. 2009. Folk psychological concepts: causation. *Philos. Psychol.* 22:205--13
- Sarkissian H, Park JJ, Tien D, Wright JC, Knobe J. 2011. Folk moral relativism. *Mind Lang.* In press.
- Schaffer J, Knobe J. 2011. Contrastivism surveyed. *Nous*. In press
- Shafer-Landau R. 2003. *Moral Realism: A Defence*. New York: Oxford Univ. Press
- Smith M. 1994. *The Moral Problem*. Oxford: Blackwell
- Solan L, Darley J. 2001. Causation, contribution, and legal liability: an empirical study. *Law Contemp. Probl.* 64:265--98
- Sommers T. 2007. The illusion of freedom evolves. In *Distributed Cognition and the Will*, ed. D Spurrett, H Kincaid, D Ross, L Stephens, Cambridge, MA: MIT Press
- Sommers T. 2010. Experimental philosophy and free will. *Philos. Compass* 5:199--212
- Sosa E. 2007. Experimental philosophy and philosophical intuition. *Philos. Stud.* 132:99--107
- Stich S. 2001. Plato's method meets cognitive science. *Free Inquiry* 21:36--38

- Sytsma J, Machery E. 2009a. How to study folk intuitions about phenomenal consciousness. *Philos. Psychol.* 22:21--35
- Sytsma J, Machery E. 2009b. Two conceptions of subjective experience. *Philos. Stud.* 151:299--327
- Uttich K, Lombrozo T. 2010. Norms inform mental state ascriptions: a rational explanation for the side-effect effect. *Cognition* 116:87--100
- Vargas M. 2007. Revisionism. In *Four Views on Free Will*, ed. J Fischer, R Kane, D Pereboom, M Vargas, Wiley Blackwell
- Vohs KD, Schooler JW. 2008. The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychol. Sci.* 19:49--54
- Wainryb C, Shaw LA, Langley M, Cottam K, Lewis R. 2004. Children's thinking about diversity of belief in the early school years: judgments of relativism, tolerance, and disagreeing persons. *Child Dev.* 75:687--703
- Waytz A, Gray K, Epley N, Wegner DM. 2010. Causes and consequences of mind perception. *Trends Cogn. Sci.* 14:383--88
- Weigel C. 2011. Distance, anger, freedom: an account of the role of abstraction in compatibilist and incompatibilist intuition. *Philos. Psychol.* In press.
- Weinberg J, Nichols S, Stich S. 2001. Normativity and epistemic intuitions. *Philos. Topics* 29:429--60
- Wong DB. 1984. *Moral Relativity*. Berkeley: Univ. Calif. Press
- Wong DB. 2006. *Natural Moralities: A Defence of Pluralistic Relativism*. New York: Oxford Univ. Press
- Wright JC, Cullum J, Schwab N. 2008. The cognitive and affective dimensions of moral conviction: implications for attitudinal and behavioral measures of interpersonal tolerance. *Personal. Soc. Psychol. Bull.* 34:1461--76
- Xu F. 1997. From Lot's wife to a pillar of salt: evidence that physical object is a sortal concept. *Mind Lang.* 12:365--92
- Young L, Cushman F, Adolphs R, Tranel D, Hauser M. 2006. Does emotion mediate the effect of an action's moral status on its intentional status? Neuropsychological evidence. *J. Cogn. Cult.* 6:291--304