EXPERIMENTAL PHILOSOPHY

Wesley Buckwalter
City University of New York, Graduate Center

Joshua Knobe
Yale University

Shaun Nichols
University of Arizona

N. Ángel Pinillos
Arizona State University

Philip Robbins
University of Missouri

Hagop Sarkissian
City University of New York, Baruch College

Chris Weigel
Utah Valley University

Jonathan M. Weinberg
University of Arizona

Table of Contents

Introduction
General Overviews
Impact of Morality and Non-Moral Judgment
   Moral Asymmetry in Intuitions
   Distortion Explanations- Affect and Blame
   Distortion Explanations- Pragmatics
   Competence Explanations
   Individual Difference Explanations
   Many-Explanations Explanations
Metaethics
   Moral Objectivism and Relativism
Free Will
   Foundations of Compatibilism vs. Incompatibilism:
   Recent Developments in Compatibilism vs. Incompatibilism
   Folk Reductionism
INTRODUCTION
Experimental philosophy is a new movement that uses systematic experimental studies to shed light on philosophical issues. In other words, experimental philosophers apply the methods commonly associated with psychology (experimentation, statistical analysis, developmental research, reaction time studies, patient studies, and so on), but they use those methods to address the kinds of questions that have been traditionally associated with philosophy. The experimental philosophy movement is united more by a shared methodology than by a shared research agenda or metaphilosophical viewpoint. Thus, while work in experimental philosophy makes use of systematic empirical study, this methodology has been applied to a wide array of different philosophical questions, and researchers have offered quite different views about the way in which such experimental work can prove philosophically valuable.

GENERAL OVERVIEWS
For those looking for an introduction to the field as a whole, there are a number of resources that either collect existing papers, as in PhilPapers and Knobe and Nichols 2008, or offer brief summaries of some of the major findings, as in Knobe et al. 2012 and Nadelhoffer and Nahmias 2007.

*PhilPapers: Experimental Philosophy[http://philpapers.org/browse/experimental-philosophy]*
A continually updated database of links to experimental philosophy papers, broken down by category.

An up-to-date review of research in experimental philosophy. The paper was written for an audience of psychologists and therefore tends to emphasize empirical findings rather than philosophical implications.


A collection of experimental philosophy papers. Includes some of the early papers that established the field, as well as new commentaries and responses from a broader range of philosophers.


An early overview of the field, emphasizing the very different ways in which different experimental philosophy projects proceed.

**IMPACT OF MORALITY AND NON-MORAL JUDGMENT**

Perhaps the most well-known finding in experimental philosophy to date involves showing how people’s prior evaluative judgments, and in particular their moral judgments, can impact the subsequent application of a number of different folk psychological and causal concepts. The section is broken down into six parts, including an broad overview of which domains have displayed this kind of moral asymmetry in intuitions, as well as the rival types of answers theorists have given for why they arise, including distortion-affect and blame, distortion-pragmatics, competence, individual differences, and many-explanation explanations.

**Moral Asymmetry in Intuitions**

The *side-effect effect*, first presented in Knobe 2003, demonstrates that people’s moral judgments about an action influence their intuitions about whether or not that action was performed intentionally. This basic effect of moral judgment on intentionality judgments has been replicated developmentally, in children as young as four years old by Leslie et al. 2006. While it was originally thought that these effects were restricted to judgments about intentional action, subsequent research has suggested that this effect extends to several concepts including causation (Roxborough and Cumby 2009), knowledge (Beebe and Buckwalter 2010), valuing (Knobe and Roedder 2009), deciding and advocating (Pettit and Knobe 2009), weakness of will (May and Holton 2011), and freedom (Phillips and Knobe 2009).


Demonstrates an extension of the side-effect effect of intentionality judgments in Knobe 2003 for attributions of knowledge. Authors consider how this moral component might impact extant epistemic theories, and speculate that variation in intuitions may be tracking an important epistemological truth.

The paper in which the side-effect effect for intentionality is first reported, whereby moral considerations are shown to affect people’s judgments about whether or not an action was brought about intentionally.


Argues that ordinary attributions of valuing involve normative evaluations of the object to be valued.


A series of experimental studies showing that people’s intuitions about weakness of will are impacted by both judgment violations and by moral considerations.


Developmental research demonstrating that judgments made by four and five years old about whether or not a foreseen side-effect was done “on purpose“ were influenced by moral considerations.


Demonstrates that the side-effect effect for intentionality also arises for the concepts desire, deciding, intending, in favor of, opposed to, and advocating.


Argues that people’s intuitions about the concept of freedom are influenced by moral considerations.


Presents experiments suggesting that variation in people’s causal judgments is due to both moral norm violation and judgments concerning event atypicality.

**Distortion Explanations- Affect and Blame**

*Distortion theories* explain these effects of evaluative considerations on folk psychological and causal judgments by positing some additional cognitive process that distorts or interrupts the ordinary application of these concepts. Nadelhoffer 2006 suggests that these additional biasing processes arise as a result of emotional or affective reaction. Young et al. 2006 challenge this explanation with neuropsychological evidence from patients with severe emotional deficits. Alternatively, Alicke 2008 and Malle and Nelson 2003 suggests that distortion based judgments
are buttressed by a desire to blame. This approach has also been challenged by the reaction time studies of Guglielmo and Malle 2010, as well as experiments by Machery 2008 and Uttich and Lombrozo 2010 involving cases that seem to lack explicitly blameworthy norm violations.

Alicke, Mark D. “Blaming badly.” *Journal of Cognition and Culture* 8 2008: 179-186. Argues that the influence of moral considerations on people’s causal judgments arises because the desire to blame leads people to attribute causation as a punishment for negative outcomes.


Uttich, Kevin, and Tania Lombrozo. “Norms inform mental state ascriptions: A rational explanation for the side-effect effect.” *Cognition* 116 2010: 87–100. Argues that the influence moral judgments have on people’s intentionality judgments arises due to the violation of general norms.


**Distortion Explanations– Pragmatics**
Other kinds of distortion theories have explained these effects in terms of conversational pragmatics. Adams and Steadman 2004a and Adams and Steadman 2004b argue that when people are presented with questions regarding intentionality, the presence of moral considerations influence the application of certain conversational rules. For instance, they speculate that experimental participants infer that the “real question” experimenters are asking involves moral responsibility, and use intentionality measures to answer accordingly. Zalla and Leboyer 2011 challenge such explanations with evidence from patients with significant deficits in their capacity to understand conversational pragmatics. Driver 2008 argues that conversational pragmatics may explain the influence of moral considerations causal judgments specifically. Knobe and Fraser 2008 reply with further experiments to test Driver’s hypothesis.

Argues that the folk have no clearly articulated concept of intentional action, and that results of Knobe 2003 (see *Moral Asymmetry in Intuitions*) are best explained by the pragmatic features of intentional language.

An expansion of the argument in Adams and Steadman 2004a whereby asymmetry in judgments about intentional action arise due to the pragmatics of intentional talk, even when the word ‘intentional’ is not explicitly mentioned.

Argues that the influence of moral judgment on causal judgments may arise due to judgments about action atypicality or by appealing to the pragmatic pressures associated with causal talk.

Knobe, Joshua and Ben Fraser. “Causal judgment and moral judgment: Two experiments.”
Presents experiments suggesting that variation in people’s causal judgments is due specifically to moral norm violation, and not due to atypicality or conversational pragmatics.

Demonstrates that individuals with Asperger’s syndrome and high functioning autism continue to display the same pattern of intentional judgments in Knobe 2003 (see *Moral Asymmetry in Intuitions*) as normal individuals.
Competence Explanations
The competence theories of Hitchcock and Knobe 2011; Knobe 2010 hold that the application of concepts on the basis of evaluative considerations like moral judgment is not due to any kind of cognitive distortion, but alternatively, somehow reflects people’s fundamental understanding of the world. Specifically, Halpern and Hitchcock 2010 have suggested that judgments about norm violations play an important role in counterfactual reasoning, which in turn underlies the application of the concepts in question. However Alicke and Rose 2010 challenge these competence views by proposing a rival explanatory model for results in the causal domain based on blame judgments. Through the use of advanced techniques in structural path modeling, Sripada and Konrath 2011 argue that the effect is not driven by normative assessments but rather by judgments people make about an agent’s character, Also see Alexander, Malet al. 2010 for a discussion of the general difficulties associated with identifying conceptual competences.

Researchers argue for a “culpable control model” of blame whereby negative evaluative reactions result in judgments of blame, and that these attributions of blame mediate the application of causal concepts independently of the moral valence of that case’s outcome.

A discussion of structural equations to investigate the outcomes and appropriate variables selected that make a difference to the conclusions we draw about causality.

Presents evidence that professional philosophers display the effect whereby moral considerations influence their judgments about causation, and argues that these data support an explanation based on the way people evaluate counterfactuals.

Provides a review of the explanation of the effects that moral judgments have on the application of folk psychological and causal concepts and presents and explanation of those effects based on people’s fundamental conceptual competences.

Argues for the “deep self concordance model”, which holds that people’s intentionality judgments in the “chairman cases” are best explained by people’s assessments of the
chairman’s underling character, values, and dispositions. Includes a discussion of the method of structural path modeling and its relevance to experimental philosophy.

**Individual Differences Explanations**
A number of researchers have suggested that the influence of moral considerations on intentionality judgments is best explained by stable individual differences. Cushman and Mele 2008 speculate that there exists at least two different folk concepts of intentional action. Nichols and Ulatowski 2007 hypothesize that the effect above arises because the word ‘intentional’ has multiple interpretations. Cokely and Feltz 2008 find that intentionality judgments are predicted by certain kinds of personality traits. Lasty Pinillos et al. 2011 demonstrate that the effect may arise due to differences in intelligence, as measured by scores on cognitive reflection tasks.

Authors present experiments attempting to demonstrate the existence of multiple ordinary concepts of intentional action, one based on an agent’s belief, the other based on an agent’s desire for a side-effect’s occurrence.

Argues that individuals possessing the personality trait extraversion are more likely to display the effect found in Knobe 2003 (see *Moral Asymmetry in Intuitions*). Evidence also suggests these judgments are susceptible to priming manipulations.

Argues that the word ‘intentional’ has multiple, yet consistent, interpretations among different groups of people. Evidence from participant self-reports suggests that one interpretation is based on foreknowledge, and another is based on judgments concerning motives.

Finds that experimental participants with higher scores on certain kinds of cognitive reflection tasks are less likely to display asymmetric intentionality judgments.

**Many-Explanations Explanations**
When accounting for the influence people’s moral judgments have on concept application, many of the theories reviewed above have attempted to show how one factor (for instance, blame or emotional response) could explain all of the effects discovered regarding non-moral concepts like intentional action. However Phelan and Sarkissian 2009 and Phelan 2011 argue
that these researchers should abandon the desire for theoretical parsimony. Citing the number of effects discovered, and the lack of consensus regarding their cause, they suggest we are unlikely to find one unified explanation for these phenomena.

  Presents side-effect effect experiments as a critical response to Machery 2008 (see *Distortion Explanations- Affect and Blame*), and argues against the theoretical parsimony of such explanations.

  A general discussion of the results experimental philosophers have uncovered regarding intentional action, and how the desire to provide one explanation of these judgments might lead researchers astray.

METAETHICS

Metaethics is a research area in philosophy that explores topics such as the nature and status of moral judgments, as well as the presuppositions and commitments that may be at play in ordinary moral discourse. One of the main questions experimental philosophers have investigated in metaethics concerns moral relativism. Philosophers have long debated whether morality is fundamentally subjective (akin to matters of personal taste) or objective (akin to matters of scientific fact), and whether moral standards can apply across cultures and eras or are instead local and applicable only to one’s contemporary society. Though philosophers disagree in these matters, many have assumed in their arguments that non-philosophers are objectivists about morality, that ordinary folk view moral issues as having a single correct answer as opposed to several correct answers that are all true relative to a given perspective or culture. This empirical claim has been akin to an article of faith, a datum that must be captured or explained by any metaethical theory. But is the claim correct? Are ordinary people objectivists about morality?

Moral Objectivism and Relativism

Experimental philosophers have begun to address the question of folk morality by studying the extent to which—and in what specific ways—ordinary people really are committed to objective moral truth. Some of the earliest studies by Nichols 2004 appear to support the claim philosophers had previously made regarding folk objectivism, and work by Goodwin and Darley (Goodwin and Darley 2008, Goodwin and Darley 2010) on canonical moral judgments seem to support this conclusion. Developmental studies on young children by Wainryb 2004 as well as experiments looking at moral judgments across the lifespan by Beebe and Sackris 2010, follow the same trend. However, Feltz and Cokely 2008 suggest that these intuitions supporting folk objectivism may be predicted by stable personality traits. Similarly, Wright et al. 2012 argue that certain individuals are more objectivist than others, particularly when evaluating certain sorts of moral acts. Lastly Sarkissian et al 2011 provide evidence suggesting that ordinary
people are actually more likely to adopt relativistic answers when prompted with more information about cultural practices.


A study of people’s metaethical views across a range of age groups, from young teenagers to mature adults. They found that most people were objectivists about morality throughout their lives, except for during their 20s.


Using the same stimulus materials as Nichols 2004, Feltz and Cokely found that a majority of subjects were moral objectivists. However, objectivism was uniquely predicted by a single personality trait—namely, openness-to-experience: the lower subjects scored on being open to new experiences, the higher they scored as being moral objectivists.


The authors found that participants treat factual and moral statements similarly, and both differ from statements of convention and taste. The authors conclude that this suggests that people are objectivist about many canonical moral transgressions.


Investigated how participants explain their judgments concerning moral disagreement and found the following correlations: those who responded as objectivists did not seem to think of alternative reasons or points of view when reasoning through a moral issue, whereas those who responded as relativists did consider such alternative reasons.


A study using a disagreement task to assess the extent of people’s commitment to moral objectivism. Participants tended to respond as objectivists. However, those who responded as relativists still distinguished between morality and convention.


Authors argue that apparently objectivist responses can be explained in terms of cultural proximity. When asked about the views of peers, people responded in
predictably objectivist ways. But when one of the disagreeing individuals was depicted as being from a different cultural group, people began to give more relativist responses.

Authors found that children were more intolerant towards disagreement when it concerned moral beliefs than other types of beliefs, including matters of taste/preference, and beliefs about the world. This suggests that individuals tend to be objectivists about morality at a very early age.

Participants classified statements as moral or not moral, and the authors measured participants’ meta-ethical commitments for the issues that they themselves classified as moral. Some moral issues (such as racial discrimination and robbery) to be far more objective than others (such as abortion and assisted suicide).

FREE WILL
Because philosophers often appeal to intuitions when arguing about free will, experimental philosophers have taken up the task of empirically examining these intuitions. Answering the question “What do ordinary people intuit about free will?” turns out to be an incredibly nuanced and complex undertaking. Besides seeking to understand what ordinary people think, experimental philosophers also seek to understand how the mind works and to examine the role of intuitions. These projects have enhanced the more traditional debates about the nature of free will. Among the overviews of experimental philosophy, a few focus specifically on the question of free will. Nichols 2011 presents the state of the field not only for philosophers but also for the broader scientific community. The article focuses on the role of abstraction in intuitions about free will and moral responsibility. Sommers 2011 presents the literature in depth while offering an analysis of its contributions. The article suggests further avenues for research and has a useful bibliography.

Nichols, Shaun. “Experimental Philosophy and the Problem of Free Will.” Science 331 (2011): 1401-1403. This review article examines the impetus for experimental philosophy on free will and explores recent trends in the debate.

Sommers, Tamler. “Experimental Philosophy and Free Will.” Philosophy Compass 5 (2011): 199-212. This comprehensive essay covers the relevance of experiments to the problem of free will, reviews the experimental literature, and explores the insights and limitations within the literature.

Foundations of Compatibilism vs. Incompatibilism
Much of the literature focuses on the question of whether compatibilism or incompatibilism is intuitive. Nahmias et al. 2004, Nahmias et al. 2005, Nahmias et al. 2006, and Woolfolk et al. 2008 provide early and important work that launched the contemporary experimental work on the question of compatibilism. Vargas 2006 gives an overview, exploring what the data mean for discussions about compatibilism vs. incompatibilism.


Nahmias, Eddy, Morris, Stephen, Nadelhoffer, Thomas, and Turner, Jason. “Surveying Freedom: Folk Intuitions About Free Will and Moral Responsibility.” Philosophical Psychology 18 (2005): 561-584. Reports on experiments that present subjects with a determinist universe and ask whether free will and moral responsibility exist in both high and low affect scenarios. Argues that the folk are intuitively compatibilists.


Recent Developments in Compatibilism vs. Incopatibilism
The question of compatibilism continues to receive attention. Monroe and Malle 2010 and Stillman et al. 2011 think that the question needs broader exploration, so these papers examine the contours of folk intuitions beyond just the question of compatibilism. Faraci and Shoemaker 2010 explore Susan Wolf’s “sane deep self view,” showing that a case that supports her argument does not have as much intuitive support as she claims. Sripada 2011 examines the basis for the intuition that manipulated agents are unfree, finding that volitional capacities are central to the intuition.

Monroe, Andrew and Malle, Bertram. “From Uncaused Will to Conscious Choice: The Need to Study, Not Speculate About People’s Folk Concept of Free Will.” *Review of Philosophy and Psychology* 1 (2010): 211-224. Broadens the question of whether the folk find free will possible in a deterministic universe to the question of what the folk concept of free will is generally. Finds that people think of free will as an unconstrained choice that follows from desire.


**Folk Reductionism**

Whether or not the folk misunderstand determinism as entailing reductionism is an important precursor to understanding whether people are intuitively incompatibilists. Nahmias 2006 advocates carefully distinguishing determinism and reductionism, Nahmias et al. 2007 make the case for distinguishing determinism and mechanism, Feltz et al. 2009 show the importance of distinguishing determinism and fatalism, and Nahmias 2011 gives an overview of much of this literature, showing the problems with implying that determinism means our psychological states play no causal role.

Feltz, Adam, Cokely, Edward, and Nadelhoffer, Thomas. “Natural Compatibilism Versus Natural Incompatibilism: Back to the Drawing Board.” *Mind and Language* 24 (2009): 1-23. Critiques previous experiments for not being careful enough about determinism versus fatalism, for not looking at both free will and moral responsibility, and for not clarifying what it means to say that an agent could have done otherwise. Argues that such issues need to be clarified before adjudicating whether the folk are natural compatibilists or natural incompatibilists.

incompatibilist intuitions are really the result of conflating determinism with reductionism.


**Affect and Abstraction**
The roles of affect and abstraction have received a great deal of attention. Much of the discussion centers on whether they create a performance error or a competence. Nichols and Knobe 2007 come down on the side of saying that affect both generates compatibilist intuitions and creates a performance error. Roskies and Nichols 2008 find that the abstraction generated by thinking about a hypothetical world generates incompatibilist intuitions. Nahmias and Murray 2011 argue to the contrary that abstraction both generates incompatibilist intuitions and creates a performance error. Sinnott-Armstrong 2008 argues that abstraction and concreteness both serve important functions, so neither should be considered error-producing. Weigel 2011 argues that abstraction is tied to both compatibilism and incompatibilism, so neither abstract nor concrete mental representations create an error. Nichols 2006 argues that conflicting intuitions are generated by different psychological mechanisms. Knobe and Nichols 2011 argue that the way we think about free will is tied to the way we think about the self. Depending on how broad the context under consideration is, emotions may or may not be considered part of the self, so there is no single, stable view. Knobe and Doris 2010 also explore the ways variance is generated in judgments of moral responsibility. Nelkin 2007 on the other hand, suggests that the amount of variance may be overstated.


Argues that we do not have a stable view about the bounds of the self, and this affects how we understand moral responsibility. Finds that when we consider the broad context, emotions are taken to be part of the self, but when we consider actions in isolation, emotions are not taken to be part of the self.


**Personality**

Experimental philosophers sometimes find that philosophically irrelevant factors affect intuitions. In the context of free will, the role of personality has gained attention. Feltz and Cokely 2009 find a connection between extraversion and compatibilism, whereas Nadelhoffer et al. 2009 argue that further work needs to be done to understand the role of affect in the original results.


**Cross-Cultural and Developmental Studies**

Experimental philosophers often use cross-cultural studies to explore whether intuitions are unified across culture and developmental studies to explore how we acquire our intuitions. Nichols 2004 argues that children believe in agent-causation, while Turner and Nahmias 2006 argue that children might be compatibilists instead. Sarkissian et al. 2010 bolster the case that the folk are intuitive incompatibilists by showing that in four diverse countries, most people give incompatibilist responses.


Turner, Jason and Nahmias, Eddy. “Are the Folk Agent-Causationists?” *Mind & Language* 21 (2006): 597-609. Presents evidence that children’s implicit beliefs might accord more with compatibilism while agreeing that the notion of moral obligation is likely an important part of the explanation of how children acquire their beliefs.

Science
Experimental philosophy on free will pays close attention to scientific developments. In particular, it examines whether advances in neuroscience will affect folk intuitions and asks whether undermining the belief in free will can lead to problematic moral situations. This section is limited to articles that (a) deal with the question of the threat of nihilism and (b) either present new experiments or explicitly discuss or propose experimental philosophy. Most believe that neuroscience will not lead to nihilism. Roskies 2006, Nichols 2007, Nadelhoffer and Feltz 2007 and De Brigard et al. 2009 all make the case that neuroscience is not a threat in various ways. Nahmias 2010 clarifies the issues surrounding the question of whether science shows that free will is an illusion. Mele 2010 gives a careful analysis of the kinds of claims neuroscientists make about free will, showing that the bridge between the experiments and the claims about free will is not as strong as generally believed.

Argues that neuroscientific advances will not lead to a general excusing of behavior, presenting experiments in support.

Considers neuroscientific research that purports to undermine the causal powers of conscious intentions and presents three tests that would help establish whether that research shows what it is typically taken to show.

Empirically examines Smilansky’s claims that most people have illusory beliefs about the existence of libertarian free will and require those beliefs to maintain moral ties to others. Casts doubt on the empirical foundations of Smilansky’s theory.

Explores putative scientific threats to free will and considers whether these threats bear fruit. Discusses work in experimental philosophy that helps clarify the issue of whether incompatibilism is intuitive, in order to show that more nuance is required when scientists say that determinism is a threat.

Nichols, Shaun. “After Incompatibilism.” Philosophical Perspectives 21 (2007): 405-428. Explores empirical and experimental literature to argue that a belief in determinism neither would nor should lead to changes in our practices of ascribing moral responsibility.

question of free will, but even if people mistakenly think they do, this will not lead to a general abandonment of moral responsibility.

CONSCIOUSNESS
There is a long-standing philosophical puzzle concerning our knowledge of other minds. It seems obvious that certain sorts of things—like people—have thoughts and feelings, whereas certain other sorts of things—like pencils—do not. What’s less obvious is how we decide whether or not something has a mind at all, and if different processes are implicated in the attribution of different kinds of psychological capacities. These questions have led experimental philosophers to explore the factors relevant in ordinary mind perception. What they have found is that there seems to be important differences between attributions of consciousness versus attributions of non-conscious mentality, especially when it comes to minds and morality. Additionally, other researchers have focused on the relationship between minds and bodies, or the role that an entity’s physical constitution has one attributions of mindedness. Lastly, experimental philosophers have also investigated the machinery of attribution, attempting to isolate the specific mechanisms that underlie people’s theory of mind judgments.

Minds and morals
Initial evidence for the folk-psychological ‘specialness’ of attributions of consciousness as opposed to non-conscious mentality comes from an investigation of mental-state attribution by Gray et al. 2007. Their results point to a distinction between two ‘dimensions of mind,’ those capacities for conscious experience (such as pain and joy) and capacities for cognition (such as self-control and planning). Additionally, they also found important correlations between moral agency and patiency with these different dimensions. This pattern of correlation was later shown to be linked to the phenomenon of ‘moral typecasting’, according to which perceptions of moral agency and moral patiency are negatively correlated, as in Gray and Wegner 2009. Collectively these results suggest that the folk conception of the mind has a multi-dimensional structure, whereby one dimension of which is distinctively associated with phenomenal consciousness (see Robbins and Jack 2006).

A study of mental-state attribution arguing for a dimensional approach to mind perception.

Building on Gray, Gray, and Wegner 2007, this paper reports the results of follow-up studies suggesting that perceptions of moral agency and moral patiency are inversely correlated.
Drawing on evidence from neuropsychology and brain imaging, the authors distinguish between the ‘phenomenal stance’ (the capacity to attribute feelings) and the ‘intentional stance’ (the capacity to attribute thoughts), emphasizing the connection between seeing something as a locus of conscious experience and seeing it as a target of moral consideration. This distinction plays a central role in their account of the psychological origins of the mind-body problem.

Minds and bodies
Experimental philosophers have also explored the degree to which folk attributions of mindedness depend on the whether or not the entity under consideration has a specific kind of physical body. Initial evidence from studies asking people to rate different kinds of mental-state attributions to entities that lack physical bodies, such as corporations (Knobe and Prinz 2008) or robots (Huebner 2010) suggest that bodies play an important role in mind perception. However, subsequent work in this area has raised the possibility that these results may be an experimental artifact (Sytsma and Machery 2009a; Sytsma and Machery 2009b; Arico 2010), as well as specific to Western culture (Huebner et al. 2010). Similarly, people think of the mind of God, who presumably lacks a physical instantiation altogether, as cognitively rich but experientially poor (see Gray et al. 2007 and Gray and Wegner 2009 cited under *Minds and Morals*).

Targeting Knobe and Prinz 2008, Arico argues that the apparent anomalousness of phenomenal-state ascriptions to corporations is an experimental artifact.

Presents the results of experimental studies suggesting that folk attributions of phenomenal states like pain are sensitive in complex ways to facts about embodiment in a way that attributions of non-phenomenal states like belief are not.

This paper reports the results of a cross-cultural study suggesting that non-Westerners regard the attribution of phenomenal consciousness to corporations as less anomalous than their Western counterparts.

The experimental philosophy of consciousness literature begins with this paper. In it, the authors present the results of a variety of studies, arguing (inter alia) that attributions of mindedness are not all of a piece, and that attributions of phenomenal
consciousness in particular typically depend upon perception of the target’s physical constitution.


In this critique of Knobe and Prinz 2008, the authors argue on methodological and interpretive grounds that robust evidence for the anomalousness of phenomenal-state ascriptions to corporations is lacking.


Argues on empirical grounds that the philosophical conception of phenomenal consciousness at play in discussions of the mind-body problem has no analog in common sense, and that the folk conception of consciousness — which differs from the philosophical conception in a fundamental way — has more to do with valence, or hedonic tone, than phenomenal character.

The Machinery of Attribution

Experimental philosophers have also wondered whether the attribution of conscious experience to a target is implemented by a psychological mechanism specialized for that purpose. According to the ‘agency model’ attributions of conscious experience are automatically triggered by default whenever a target shows signs of animacy, such as biological motion and contingent interactivity (Arico et al. 2011; Fiala et al. 2011). This model, supported by the reaction times studies of Arico et al. 2011, suggests that a single mechanism is responsible for the attribution of both mindedness in general and conscious mindedness in particular. Importantly however, the agency model does not exhaust the possibilities for explaining attributions of mindedness. It could be that the capacity to attribute consciousness to something— or to adopt the ‘phenomenal stance’ towards that thing—rests on a functionally specialized mechanism, at least partially distinct from the mechanism responsible for the attribution of thoughts (see Robbins and Jack 2006 cited under *Minds and Morals*).


Presents and defends the ‘agency model’ of mental-state attribution, according to which phenomenal consciousness is automatically attributed by default to any entity exhibiting animacy cues — a default that is only overridden as a consequence of controlled processing reflecting the attributor’s beliefs about which sorts of entity are conscious and which are not.

Deploys the ‘agency model’ (described above) in a dual-process account of the psychological origins of the mind-body problem.

**EPISTEMOLOGY**

Within the western philosophical tradition, epistemology is the subfield that mainly studies the philosophical aspects of knowledge and justified belief. One important methodological commitment of epistemology, at least as has been practiced recently, involves the use of intuitions about hypothetical cases to gather evidence in favor of theses concerning knowledge and other epistemic notions. Since experimental philosophy involves the experimental study of philosophical intuitions, its methods are naturally applied to epistemology. Experimental epistemologists have mainly focused on specific first-order debates regarding contextualism and pragmatic encroachment, but have also begun to explore a mix of moral general issues such as intuitions concerning know-how, knowledge entailing belief, and moral encroachment. Lastly, a subsequent portion of the work in experimental epistemology regards issues in philosophical methodology, where the reliability of epistemic intuitions are called into question and debated in light of certain experimental results. These latter contributions are discussed in *Metaphilosophy*. Pinillos 2011 reviews and challenges some recent work in the field, specifically focusing on how these data might bear on current debates between contextualists and supporters of pragmatic encroachment.


This paper summarizes some experimental work that suggests that people’s intuitions do not accord with predictions made by some interpretations of Contextualism and Pragmatic Encroachment. The paper is somewhat critical of this experimental work.

**Contextualism and Pragmatic Encroachment**

Experimental philosophers have carried out research that aims to illuminate various first order theories in epistemology. These include contextualism (the thesis that knowledge attributions can vary in content in epistemically interesting ways depending on the context of utterance), and pragmatic encroachment (the thesis that whether someone who believes P also counts as knowing P may depend on their practical interests). For experimental evidence which goes against Pragmatic Encroachment and certain types of contextualism, see Buckwalter 2010, Feltz and Zarpentine 2010, and May et al. 2010. For experimental evidence in favor of Contrastivism (a type of Contextualism), see Schaffer and Knobe 2010. A good critique of this work can be found in Keith DeRose 2011. For an argument that a lot of the experimental work cannot, in principle, count against pragmatic encroachment, see Jessica Brown 2011. And for experimental work supporting pragmatic encroachment, see Pinillos 2012.

Authors report on experiments challenging certain predictions made by theories of pragmatic encroachment and contextualism. Results suggest that agents do not use knowledge in a way that is sensitive to what is at stake for the purported knower, or in a way that is sensitive to error salience.


The author reports on experiments suggesting that people’s use of knowledge may not be sensitive to what is at stake for the purported knower, raising questions about the evidentiary support for some versions of contextualism and pragmatic encroachment based on ordinary intuitions.


The authors present evidence from many different types of experiments that knowledge attributions are not sensitive to what is at stake for the purported knower. This is thought to be problematic for certain versions of Pragmatic Encroachment.


The authors defend Contrastivism about knowledge (a type of contextualism) using experimental methods. They also report on evidence that folk knowledge ascriptions are sensitive to salience of error.


The author defends Contextualism against some experimental philosophy results. The defense does not report on any new experiments. The discussion is particularly useful since it discusses various possibilities for Contextualism.


The author argues that the sort of experimental work which has been taken by some to impugn Pragmatic Encroachment cannot do so. This is because Pragmatic Encroachment, according to the author, does not make predictions about folk usage of the sort the experiments were aimed to probe. The author argues the same cannot be said about contextualism.

This principle has been taken to strongly support pragmatic encroachment. Some of the data presented also goes against certain versions of contextualism.

**Know-How, Knowledge Entailing Belief, and Moral Encroachment**

Experimental work has also been taken to support other theories in Epistemology. For example, a Rylean argument against the reduction of know-how to propositional knowledge has been challenged by Bengson at al. 2009. Also, the idea that there is moral encroachment in knowledge attributions has been defended in Beebe and Buckwalter 2010 (see *Moral Asymmetry in Intuitions*)

John Bengson, Marc A. Moffett, and Jennifer C. Wright. “The Folk on Knowing How.” Philosophical Studies 142 (3) 2009 387–401

According to Gilbert Ryle, knowing-how does not analyze to propositional knowledge. One reason he gave for this is that such a reduction over-intellectualizes our mental life. The authors provide some experimental evidence that this charge is without merit.

**METAPHILOSOPHY**

While the majority of the work conducted under the rubric of experimental philosophy have positive projects aimed to secure various positive philosophical theses, one project has had a more specifically negative metaphilosophical goal: raising empirically-based concerns about the widespread practice in philosophy today of appealing to intuitions about cases as evidence. In its most basic form, this negative project (also known as “restrictionism”, because of its stated goal of curtailing that target philosophical practice) argues from a starting point of the empirical evidence unexpected sources of error in intuition. This section reviews several central areas of research in experimental philosophy relating to these challenges. First, experimental philosophers have argued that the demographic diversity of intuitions, and instability of intuitions in terms of framing, context, order, and materials effects have important metaphilosophical implications. Second, experimental philosophers have used these data to phrase the challenge this evidence poses to traditional philosophical methods in a number of ways. Third, armchair defenses have been given to meet those challenges, with considerable emphasis on what might be called “wrong intuitions” defenses. Lastly, some philosophers have argued that these findings point to an irresolvable instability of intuitions, further questioning their reliability.

**Demographic Diversity**

Some experimental work suggests ways in which intuitions vary with such factors as the intuiter’s ethnicity (Weinberg, Nichols, and Stich 2001; Machery et al. 2004, extended in Mallon et al. 2009; see also Sarkissian et al. 2010 under *Cross-Cultural and Developmental Studies*) or gender (Zamzow and Nichols 2009), or other demographic dimensions, such as philosophical background (Nichols, Stich, and Weinberg 2003) or personality (see Feltz and Cokely 2009 cited under *Personality*).

The authors report evidence of cultural and SES variability in intuitions about classical epistemology thought-experiments. There is also evidence of philosophical universality: participants across all demographics were uniformly unwilling to attribute knowledge to someone who was just lucky at guessing whether a coin would land heads or tails.


The authors report evidence of cross-cultural variation in referential intuitions on variants of Kripke’s “Goedel/Schmidt” case.


Explores the substantial range of debates in philosophy that would be radically undermined by results suggesting widespread demographic variation in intuitions about reference.


Participants with less philosophical training seemed more willing to accept a Moorean response to external-world skepticism than did those with more coursework in philosophy. This is perhaps the only negative project paper that makes an argument based on differences between philosophically-trained and philosophically-naïve subjects.


The authors report evidence of a number of differences in how male and female subjects judge a set of trolley cases. They also explore whether variation in ethical intuitions could actually be overall an epistemically good thing for the community to have.

**Framing, context, order, and materials effects**

In addition to variation across persons, there is also evidence of instability of intuitions, in which the philosophical intuitions are susceptible to a number of contextual, framing, or other environmental effects (Petrinovich and O’Neill 1996; Liao *et al.* forthcoming; Swain, Alexander, and Weinberg 2008, which is responded to, with new data, in Wright 2010; see also Beebe and Buckwalter 2010 under *Know-How, Knowledge Entailing Belief, and Moral Encroachment*). Also, a great deal of the work on free will discussed above is also relevant: Nichols and Knobe 2007 and Roskies and Nichols 2008(under *Affect and Abstraction*), and Weigel 2011.

The authors report evidence that framing trolley cases in terms of who will be saved vs. in terms of who will die, can influence judgments about the cases. They also found some influence on judgments from what cases preceded the case being judged.

Liao, S. Matthew, Alex Wiegmann, Joshua Alexander, and Gerard Vong. Forthcoming. “Putting the Trolley in Order: Experimental Philosophy and the Loop Case.” *Philosophical Psychology*. The authors report evidence that intuitions concerning Thomson’s “Loop Case” vary according to which cases are considered preceding it. They explicitly draw the inference that intuitions about this case should not appealed to in arguments about the standing of moral principles like the Doctrine of Double Effect.

Swain, Stacey, Joshua Alexander, and Jonathan Weinberg. 2008. “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp.” *Philosophy and Phenomenological Research* 76:138-155. The authors report evidence of an order effect on a “Truetemp” case Roughly speaking, when the case follows a clear case of knowledge, subjects are less willing to attribute knowledge in the case than they are when the case is the first one they see.

Wright, Jennifer. 2010. “On intuitional stability: The clear, the strong, and the paradigmatic.” *Cognition*, 115: 491-503. Evidence that degree of confidence in an intuitive judgments may track its instability. Philosophers may thus, it is conjectured, exploit this correlation to help determine where they may be more prone to methodologically unfortunate influences like order effects.

**The Challenge**
The raw materials of the experimental findings themselves are only one part of the overall argument concerning the use of intuitions in philosophy. There are a number of theoretical questions that arise, in terms of diagnosing the specific nature of the (alleged) epistemological or methodological deficiencies in the traditional practices; in understanding just how severe or wide-reaching the threat of the results are meant to be; and seeing what possible responses can be made on behalf of the traditional practices. So what sorts of epistemological and methodological commitments and assumptions must be made, in order for the sorts of results canvassed above to begin to make trouble for traditional philosophical practices of appealing to intuitions? This is explored in Alexander and Weinberg 2007; Goldman 2010; Mallon 2007; Appiah 2008; and Machery 2011. Just which aspects of these practices are under threat, and to what extent (Feltz 2008; Mallon et al. 2009 [cited under *Demographic Diversity*] as well as Horvath 2010 and Ichikawa, et al. 2011)?

Alexander, Joshua and Jonathan Weinberg. 2007. “Analytic Epistemology and Experimental Philosophy.” *Philosophy Compass*, 2: 56-80. Presents a basic framework for distinguishing the negative project (“restrictionist”) uses of experimental philosophy from positive approaches; different models of the
philosophical practice of appealing to intuitions; and a survey of some early responses on behalf of that practice.


A wide-ranging exploration of a number of the theoretical issues involved in bringing scientific work to bear on ethical questions, arguing for a position that can respect both the contribution of the scientific work, and more traditional approaches.


Targets some *a priori* arguments that would purport to show that no experimental demonstration of the unreliability of intuitions in philosophy.


Presents a naturalistic and reliabilist framework for thinking about the use of intuitions in philosophy, with an extended exploration of what it means to understand intuitions as distinctively *classificatory* judgments.


Offers a construal of the debate that suggests only a “conservative restrictionism” is well motivated: not that the practices on the whole are challenged, but only the specific cases that have been experimentally examined.


Contends that intuitions about reference of the sort that have been studied in experimental philosophy do not, in fact, play much of a role in philosophical debates about reference, or that appeal to causal-historical theories of reference as a premise.


Construes the negative project as offering a limited and conditional version of skepticism about thought experiments. The author argues that there several different ways that ordinary cognition might be inadequate for some thought experiments.


Frames the concern about philosophical intuitions not in terms of whether philosophers’ judgments about which premises to accept in an important class of arguments is itself dependent on their desired conclusions.
Reactions to and Refinements of the Challenge
Further statements of this challenge have also explored its specificity and scope. Are the practices already sufficiently empirical to avoid the challenge (Stich and Weinberg 2001; Weinberg 2007; Weinberg 2008 [cited under *Framing, context, order, and materials effects*], Grundman 2010; Talbot 2010; Brown 2011; Ichikawa 2011)? And does the challenge apply not just to armchair philosophy, but also to some forms of experimental philosophy as well (Alexander, Mallon, and Weinberg 2010)?

Points out that not just traditional ‘armchair’ methods are under pressure from these experimental results, but also a number of positive projects in experimental philosophy itself as well.

Points out several cases in which it looks like we might want to endorse a methodology even if it were not hopeful, and the pragmatic nature of the case for hopefulness entails that its importance can in principle be overridden by other pragmatic factors.

Argues that a high degree of reliability entails the existence of a capacity to detect and correct for errors, and so Weinberg’s attempt to distinguish hopefulness as a further condition on trustworthiness, above and beyond reliability, fails.

Suggests a number of ways, with close attention to specific papers in the literature, in which current philosophical practice already contains a fair capacity to check for its own errors.

Points out that philosophers may be relying illegitimately on unsystematic and uncontrolled observations of people’s intuitions, in a way that will lead them to overestimate the degree of agreement about those intuitions.

Emphasizes the importance of investigating not just the pattern of intuitions themselves, but moreover the underlying psychological mechanisms that produce them, if we are to understand where they are and are not trustworthy

Presents a framework for debating the trustworthiness of intuitions in terms of “hopefulness”: not the reliability of a source of evidence, but our capacity to detect, quarantine, and correct errors in that the source. The author contends that the practice of appealing to intuitions in philosophy is substantially without “hope”.

**Armchair Defenses**

Unsurprisingly, the challenge of the negative project has received many responses looking to defend the traditional intuition-deploying practices from the experimental critique. Ernest Sosa has articulated in a series of papers an attempt to assimilate the negative project arguments to more traditional forms of skeptical argument, which thus should be seen as failing for the same reason that those skeptical arguments do (Sosa 2007; Sosa 2009; Sosa 2011). Alternatively, some philosophers have addressed the design of the existing experiments themselves, suggesting ways in which they might not yield appropriate results, perhaps because of the pragmatics of survey design (Cullen 2010); or because the experiments as run have not successfully targeted the philosophical claims they meant to address (see DeRose 2011 in *Contextualism and Pragmatic Encroachment*). Still others have argued that philosophers do not, in fact, make any serious reliance on intuitions, and so all of the experimental philosophy work is simply irrelevant to evaluating its methodology (Deutsch 2010; Cappelan 2012), though this line turns out to be somewhat hard to sustain, as the experiments may still be relevant under different construals as one also reconstrues philosophical practice (Alexander 2010).


Defends armchair philosophy by arguing that much of the diversity discovered by experimental philosophers can be explained as merely verbal disagreement.


Argues that experiments that the apparent diversity in epistemic intuitions might be caused by artifacts of the experimental materials. Furthermore, where there are genuine differences in intuitions, these may reflect judgments about epistemic standards rather than about the nature of knowledge.


Explicates justified intuition as a form of epistemic competence and maintains that experimental philosophy might play an important role in exposing distorting influences on intuitions.

The pragmatics of survey design can, in principle, induce context effects even if there are no underlying context effects on the actual intuitions.

Philosophers don’t generally appeal to intuitions, and so experimental work that studies intuitions is methodologically irrelevant.

Based on a close, linguistically-sophisticated examination of the many uses of “intuition”, “intuitive” and the like in philosophy, as well as a set of case studies of particular putative uses of intuition in the philosophical literature, the author concludes that there is no such methodology as the appeal to intuitions.

Responds critically to a version of this argument found in Timothy Williamson’s *Philosophy of Philosophy*.

The “Wrong Intuitions” Defenses
A number of philosophers have argued against one specific part of the negative project’s commitments, i.e., that the patterns found in the experimental work can be expected to be found in professional philosophical intuitions as well. These can be taxonomized in terms of what sorts of differences they posit between what is measured in the experiments, and what gets used by the philosophers: intuiting only upon sufficient reflection (Kaupinnen 2007; Livengood et al. 2010; Weinberg et al. 2010) appealing to a phenomenology of seeming (Bengson forthcoming;; or the deployment of training that makes one a philosophical expert (Ludwig 2007; Weinberg et al. 2010; Williamson 2011; Schulz, Cokely, and Feltz 2011).

The author argues that surveys of the sort used by experimental philosophers can only measure “surface” intuitions, but “robust” intuitions of the sort had by thinking harder and reflecting on the scenarios may not have the kind of instability and variation reported in the experimental studies.

A large-scale survey indicating that, even controlling for other factors like general education level, philosophers are more reflective than non-philosophers, as measured by performance on the Cognitive Reflection Task (CRT).

Analogizes philosophers to expert mathematicians making judgments about infinity, who we would clearly trust over any untrained folk making their own judgments about such matters.

Surveys portions of the extant psychological literature on expertise; then renders a pessimistic verdict on the likelihood that philosophers have expertise of the very specific sort needed to preempt the challenge of the negative project.

Contends that, since performing actual investigations of the philosophical expert population would be so difficult that it would be unreasonable to expect anyone to pursue such investigations, we should instead grant substantial deference to the philosophical experts.

The authors report an actual recent investigation of the philosophical expert population, and report evidence of persistent order effects in their judgments about trolley cases, and even a heightened order effect in their judgments about moral principles.

The authors report an actual recent investigation of the philosophical expert population, and report evidence indicating persistent influence of the personality trait of extraversion on judgments concerning free will even in that population (see also Feltz and Cokely 2009, under *Personality*).

Irresolvable Instability?
Several philosophers have suggested that some of the effects suggested by experimental philosophy work could be understood as signs that we as individuals do not have any one stable inner epistemology, morality, metaphysics, etc. Thus, it may be futile to try to separate that which are the “competent” intuitions from the ones that may be mere performance errors: these instabilities may themselves be part and parcel of our competences This is explored in Nichols and Ulatowski 2007, cited under *Individual Difference Explanations*, Gendler 2007, and Spicer 2010. See also Sinnott-Armstrong 2008, cited under *Affect and Abstraction*.

If different processes are activated by different ways of framing a thought experiment, we may simply be unable ever to attain cognitive equilibrium regarding it.

Suggests that the cultural diversity in epistemic intuitions need not reflect a deep cultural diversity in folk epistemology, but can be explained in terms of other cognitive factors.