Chris Weigel

Experimental Philosophy Is Here to Stay

Abstract: Experimental philosophy is comprised of two broad projects, the negative project and the positive project, each of which is a response to a kind of armchair use of intuitions. I examine two examples of the negative project—the analysis of knowledge and the theory of reference—and two examples of the positive project—free will and intentional action—and review criticisms of each example. I show how the criticisms can be met and argue that even if they could not have been met, experimental philosophy raises important questions about methodology, opening the door on new questions and new ways of looking at old questions. For that reason, experimental philosophy as a movement is robust and full of potential.

0. Introduction

The recent naturalistic movement known as experimental philosophy has become established enough now that to speak simply of 'experimental philosophy' is to use a convenient label that describes very diverse projects unified by family resemblance around empirical methods. To outline all of the projects and sub-projects and address not only the goals but also the criticisms would be far too ambitious for one paper. Nonetheless, it is possible to give the basic contours of two relatively exhaustive projects within experimental philosophy, and that is my undertaking in this paper.

But first, to understand experimental philosophy, it is helpful to step back a little and examine empirical philosophy more generally, to see the genus to which it belongs. There are many ways that philosophy can intersect with the empirical. By 'empirical philosophy' I have in mind something more specific. In empirical philosophy, a philosopher uses a scientific finding or body of scientific findings in order to advance a philosophical argument (for further discussion about the relationship between empirical/experimental philosophy and armchair conceptual analysis see Prinz 2008). An exemplar here is John Doris's work *Lack of Character* (2002). Doris marshals an impressive array of social psychological research on personality and the correspondence bias and finds patterns in it. Namely, he finds that there is massive amounts of empirical evidence for the fact that thinking we have a stable character is really a form of the correspondence bias. He argues for a version of situationism, the view that behavior is much

more a function of situation or circumstance than character, that character is to a large extent a philosophical fiction.

None of the experimental hypotheses he cites say, 'the Aristotelian tradition is wrong' or 'situationism is true'. How could they? That is not a claim that can be directly tested empirically using a controlled study. Doris's contribution is to critically examine the broad trend and show its philosophical import. He also weighs in on debates in the literature about the existence of personality, but again contributes philosophically by showing how these debates matter for the philosophical question of whether there is such a thing as character and how the answer matters for ethics. Doris's work is an exemplar: empirical philosophy also includes much of the literature on mental imagery, emotion, perception and mental representation, to name just a few examples. The characteristic that makes these works empirical philosophy is the use of empirical evidence. The use of empirical evidence in all of these topics influences the kinds of questions asked and changes what counts as a relevant way to look at a question. For example, contemporary virtue theorists have heard of Milgram's famous experiments (1974). Doris's contribution is to argue that the experiment and others like it help undercut virtue theory. Also, belonging to empirical philosophy is a matter of degree, and empirical philosophy does not use empirical science as its only tool: it participates in both elements of its moniker.

Empirical philosophy becomes experimental philosophy when the philosopher carries out the experiments needed. Philosophers have turned to carrying out their own experiments because empirical philosophical questions are not confined to the kinds of things that scientists typically work on. Primarily, there are empirical claims made in the philosophical literature that require empirical backing. These empirical claims typically involve claims about intuitions but they are traditionally made from the armchair. There are two overlapping purposes for armchair intuitions. First, they are used as evidence for a theory. An armchair philosopher may say, for example 'it is intuitive that x' or 'the folk intuitively think that y'. The goal in saying this is to bolster a particular claim. A causal theorist about names may claim that it is intuitive to think that in a particular situation a name refers to someone who does not fit any of the descriptions associated with the person. An epistemologist may claim that it is intuitive that a particular case of justified true belief is not knowledge. In each case—and examples abound in the literature—the appeal to intuition is supposed to give evidence for the theory in question. Second, claims about intuitions take place when a philosophical project is supposed to track folk concepts. For example, consider the free will literature. Philosophers repeatedly claim that although there may be many concepts of free will, the one they are interested in is the one that the folk use in ordinary practices, the concept that supports ordinary claims about moral responsibility. Claims about whether the folk are intuitively compatibilists or intuitively incompatibilists are made from the armchair. Folk psychology provides another example. In discussions about folk psychology, ordinary concepts like beliefs, desires, and intentions are precisely the target of investigation. Hence, the question of whether folk psychology requires an intentional action to involve trying on the part of the agent targets the folk concept.

Armchair philosophers use intuitions about cases to declare what the folk concept is.

The two broad projects alluded to above can be understood as a response to each of these types of armchair use of intuition. First, when armchair philosophers use a claim about what is intuitive to bolster a particular theory, various empirically related concerns arise. For whom is the claim intuitive? Are our intuitions cognitively respectable or are they subject to bias? Are they neutral or shaped by philosophically irrelevant considerations? In the negative project, or experimental restrictivism, experimental philosophers have discovered great diversity in intuition and cognitive bias on the part of armchair philosophers, and have therefore seriously undermined the method of using intuitions as evidence for the application of a concept.

Second, when armchair philosophers claim to be discussing folk concept as in the cases of moral responsibility and folk psychology, the experimental philosopher asks 'Are you sure?' and follows up with 'Let's find out'. The experiments not only show us something about the original folk concept in question, but also important things about how the mind works. Here this second, positive, project not only helps to clarify and deepen the nature of the original inquiry, but also opens new avenues for inquiry.

In *sections 1 and 2*, I explain the negative project through two examples. Both the analysis of knowledge and the meaning of proper names have been subject to experimental scrutiny. In both of these cases, philosophers have used experiments to uncover systematic diversity in intuitions and have gone on to argue that intuitions therefore are fundamentally flawed as a basis for a method for these questions. I show examples of two kinds of objections to the negative project and give examples of how experimental philosophers reply to each kind. Sometimes the objections to this project have an empirical basis, and I give examples of how experimental philosophers have met these objections empirically. Other objections take the basic project for granted but show that the experiment does not show what it claims to show. Here I show how this kind of objection leads to new experimental work that can bolster the original claims. I argue that because of the contributions the negative project makes, even a very critical stance toward the negative project shows that it is here to stay because of the important avenues and dimensions it opens.

In *sections 3–4*, I explain the positive project through two more examples. Discussions of free will and of folk psychology have relied on claims about folk concepts. Experimental philosophy has shown that folk concepts are more subtle and different than philosophers may have supposed. Unlike the negative project, the positive project does not call for an end to using intuitions. In fact, in these debates, intuitions are what we seek to understand, so it would be rather incoherent to eliminate them from the discussion. The positive program only says that claims about what the folk intuit are empirical claims. Criticisms of the positive project argue that the experiments do not show what they claim to show and criticize more broadly the method of using experiments to understand folk intuitions. I argue on the contrary that the experiments do get at the philosophically important concepts, but even if they didn't, the positive project would be here

to stay because it helps us better understand how our minds work and human nature more generally.

1. The Negative Project and the Analysis of Knowledge

Weinberg, Nichols and Stich (2001) assert that intuitions are used uncritically in a certain class of projects in epistemology and show that there is good reason to doubt the efficacy of this strategy. Hence, there is good evidence to doubt that these projects are philosophically justified because of they way they rely on the armchair. But how could this be? How could, for example, the whole tradition in analytic epistemology of trying to analyze the concept of knowledge be based on a questionable method?

WNS make their case in the following way. First of all, consider how Gettier-style cases are used. Suppose, for example, that you believe that knowledge is justified true belief. But then you are presented with the Gettier cases (1963). Intuitively, these do not seem to be cases of knowledge even though they are cases of justified true belief. So you add something to your analysis of the concept of knowledge, say a stronger justification condition, and now your intuition is satisfied. This story is familiar (Shope 1983), but WNS's insight is that there are empirical claims being made about intuitions. For example, the argument relies on the claim that the Gettier cases are intuitively not cases of knowledge. Introspection may reveal whether you yourself find this intuitive, but the argument relies on a more general claim about what is intuitive. After all, the discussion is not merely about what one subjectively believes that knowledge is; rather, it deals with the question of what knowledge really is. Hence, the claim about whether Gettier cases are intuitively cases of knowledge or not is a claim about what is naturally intuitive.

Previously, philosophers did not question whether the claim is naturally intuitive: after all, as the large literature responding to Gettier's paper attests, most philosophers do find his examples intuitively to challenge the claim that knowledge is to be analyzed as justified true belief. Other, more complicated cases may split philosophers on intuitive lines, but then the discussion is about what features you are supposed to be focused on, with the implicit assumption being that when looked at the right way, after long dialogue, reflective equilibrium will bring agreement in intuitions.

WNS did question the assumption that Gettier cases are intuitively not knowledge. They showed that what is intuitive is in part a matter of socio-economic status and of cultural background. Their experiment presents Western and East Asian subjects with scenarios that probe some of the issues in the analysis of knowledge debate. For example, subjects identified as Western and East Asian were presented with the following scenario (29):

"Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

REALLY KNOWS   ONLY BELIEVES"

WNS found that Western subjects intuit that Bob only believes Jill drives an American car (75% vs. 25% who said he really knows it), but the majority of East Asian subjects says that Bob really knows (57% vs. 43% who said he only believes). Several other probes produced similar results. A similar effect was found between high socio-economic and low socio-economic status (SES) groups: the low SES groups are much more likely to attribute knowledge than the high SES groups.

WNS's claims are bolstered by the fact that these phenomena are not isolated, but rather are explainable in terms of work in cultural psychology on culture and socio-economic status (Haidt/Koller et al. 1993; Norenzayan/Nisbett et al. 1999; Nisbett/Peng et al. 2001; Nisbett/Norenzayan 2002). WNS interpret their claims in light of this other work on culture that says that East Asians have a tendency toward 'holistic' thought while Westerners have a tendency toward 'analytic' thought (23) and in light of this other work on socio-economics that says that different social classes have different moral intuitions (24). WNS's results, then, far from demonstrating an anomalous curiosity, really demonstrate that epistemological intuitions are sometimes radically diverse.

Since there is no principled reason that Westerners should be better at intuiting whether a belief is also a case of knowledge, the use that epistemologists have made of intuitions seems to be undermined as an example of what may be called 'ethno-epistemology' (Nichols/Stich et al. 2003). At the very least, as Weinberg says, epistemologists who use this method owe us an account of why their intuitions are epistemically privileged.

Matters are worse though. If the culture of the person thinking about the claim 'Bill really knows that Jill drives an American car' is irrelevant to the truth of the claim, at least, it might be argued, culture provides a framework for thought. We can systematically study this framework and understand its scope and limitations. And at least we can see why it might serve a purpose for culture to play a role in shaping thought: culture may turn out to be cognitively respectable if limited. But what if intuitions are radically unstable? What if intuitions are subject to forces that are completely cognitively irrespectable? What if our intuitions are influenced by what we happen to be thinking about before the scenario in question? Swain, Weinberg and Alexander (2008) show that this is exactly the case: whether we intuit that a belief is a case of knowledge significantly depends on the case we considered immediately beforehand. They present subjects with the Truetemp case. This case is used by Keith Lehrer as an argument against reliabilism, the view that a true belief is knowledge if it is caused by a reliable cognitive process (1990). The thought experiment

posits that a man has a device inserted into his head without his knowledge that accurately reads the temperature and generates thoughts about what the correct temperature is. Lehrer thinks it is intuitive that this man does not really know the correct temperature. But SWA show that intuitions vary according to whether a clear case of knowledge is presented first (in which case, the mean shifts more toward saying this is not knowledge) or whether a clear case of non-knowledge is presented first (in which case, the mean shifts to saying this is knowledge).

In the traditional debate about the analysis of knowledge, then, intuitions are taken to have evidentiary force. If these experiments are right, intuitions are highly unstable, depending on philosophically and even culturally irrelevant factors. Again, if intuitions are to be used as evidence in light of these findings, an account is needed.

Not surprisingly, such accounts have been given, and they point to objections to the negative project generally. Some of them contain empirical claims. Experimental philosophers must take empirical claims seriously as their relevance is the cornerstone of experimental philosophy. Here I discuss the most prominent objection, which has come to be known as the Reflection Defense (see e.g., Sosa 2005; Kauppinen 2007; Sosa 2007).

The basic idea behind the Reflection Defense is that the intuitions of philosophers are the only ones that matter. After all, philosophers are being careful, thoughtful, and deliberate when they intuit. College students filling out a survey presumably are anything but careful, thoughtful, and deliberate. They are simply recording gut reactions, not reflective reactions.

The objection that unreflective intuitions are irrelevant rests on an empirical claim. Weinberg, Alexander and Gonnerman put the point succinctly:

> "The Reflection Defense turns on an empirical hypothesis: intuitions of interest to philosophers—that is, thoughtful judgments based on reflection on the case—won't suffer from the same kinds of instability that seem to afflict the intuitions studied by experimental philosophers." (forthcoming)

They test the empirical hypothesis in the following ingenious manner. They look at a personality trait known as Need for Cognition (NFC). This trait measures the degree to which a person "engages in and enjoys thinking" (Cacioppo/Petty 1982). The objection is hence operationalized by WAG to say that high NFC subjects will exhibit less instability than mid or low NFC subjects. The hypothesis is then tested.

First, WAG find that mid and low NFC subjects are susceptible to priming. These results replicate SWA's results. For high NFC subjects, the situation is more complex. They are also susceptible to priming, but *in the reverse direction.* That is, high NFC subjects seem to overcompensate for the default position. Nonetheless, priming effects are priming effects, and the empirical claim comes out against the reflection defense.

The NFC defense is very powerful, although questions can still be raised. Is the reflectiveness that philosophers display sufficiently captured by the need for

cognition instrument? Perhaps, after all, philosophical reflectiveness is shaped and fine tuned by years of considering the nuances of arguments, seeing how they fit into a larger debate, teasing out the implications, finding consistency with other views and so on. One hardly expects, that is, that Keith Lehrer would change his mind about the Truetemp case if he read about some poor example of knowledge presented as pseudo-scientific research in his morning newspaper. Also, one might ask whether NFC is a true personality trait or whether NFC is subject to situationist considerations. Finally one might argue that this overcompensation that philosophers do produces more stable intuitions. That is, much like Descartes used a worst-case scenario to show that knowledge is possible, high NFC subjects/philosophers do the same, so the instability actually produces more rigorous standards. Challenges to the NFC defense are, that is, possible, but this does not change the fact that the burden of proof is on the philosopher who uses intuitions to show why they can count as evidence: we cannot assume that intuitions are evidence uncritically any more. Moreover, there is strong evidence that such an account will not be forthcoming and that the entire armchair project will be hence undermined.

By making the question about intuitiveness empirical and by making the objection empirical as well, experimental philosophy gives rise to new avenues for research, as indicated by the questions in the previous paragraph. When it comes to perspectives, more is better. Even someone who disagrees with Doris's situationism has to agree that it raises philosophically important questions in a way that inherently relies on the empirical. So even though *prima facie* experimental restrictivism may seem to represent only a narrowing of philosophy, it also represents a broadening. That is, the fear that experimental philosophy will bring about the end of philosophy is unfounded. Because of the way it broadens the field, experimental philosophy is here to stay.

2. The Negative Project and Theories of Reference

The second example of the negative project comes from a pair of papers about the theory of reference. Mallon, Machery, Nichols and Stich (MlMNS) argue first of all that philosophers use what they call "the method of cases" when trying to find the correct theory of cases. The method of cases is that "the correct theory of reference for a class of terms T is the theory which is best supported by the intuitions competent users of T have about the reference of members of T across actual and possible cases" (2009, 338). In other words, the theory of reference is in a similar situation as the analysis of knowledge: each debate relies on intuitions to give evidence.

Like the analysis of knowledge, the theory of reference would also suffer from a serious methodological predicament if the intuitions about the reference of proper names (to take a very central case) turn out to be contingent upon non-philosophically relevant factors. Machery, Mallon, Nichols and Stich (McMNS) argue in *Semantics, Cross-Cultural Style*, that linguistic intuitions are in fact significantly different based again on whether the subject is Western or East Asian.

They use probes similar to Kripke's famous Gödel case, the case that elicits the intuition among philosophers that the name Gödel would refer to the historical Gödel even if none of the descriptions that common belief attaches to him are accurate. Western subjects have intuitions that fit Kripke's causal-historical account; East Asian subjects intuit significantly more like descriptivists (2004). Again, work in cultural psychology provides a possible way of understanding these differences: Western subjects tend to use a causal orientation when describing the world whereas East Asian subjects lean more toward perceiving similarity (Nisbett 2003).

MlMNS use these results to argue against the method of cases. The dialectic in the theory of reference for proper names rests on the implicit assumption that intuitions about cases are universal (339). But intuitions are not universal. So they should not be used as evidence for a theory of reference. To the extent that theories of reference rely on intuitions, they should be abandoned (352).

The problem, however, is much worse. For MlMNS also argue that there is a large array of philosophical debates that rely on arguments from reference. An argument from reference is one that starts by assuming a substantive theory of reference and uses that theory to derive another philosophical conclusion about whether the referent of a term, based on that theory, refers or fails to refer or has changed reference. Further substantive conclusions are then derived. For example, eliminative materialists assert that terms of folk psychology such as 'belief' and 'desire' associate certain descriptions with those putative mental states. But scientific findings indicate that there is no mental state that fits the descriptions that folk psychology associates with the term. Therefore, beliefs and desires do not exist.

This argument relies substantively on a theory of reference, which in turn relies substantively on intuitions about cases. Since McMNS have undermined the universality of those intuitions and hence the plausibility of relying on intuitions to establish philosophical conclusions, this defining argument for eliminative materialism is undermined. MlMNS make similar cases about scientific realism, the reality of race, and particular arguments about our knowledge of the nature of the good in ethics.

This second negative project allows for the opportunity to see a second type of criticism. The first class of arguments against experimental restrictivism concerns the relevance of ordinary intuitions. This Reflection Defense was addressed in the context of experimental philosophy and the analysis of knowledge. A second class of objections takes the methodology of experimental philosophy for granted and instead asks whether the experiments are designed sufficiently well to show what they say they show. To illustrate this objection, consider *Against Semantic Multi-Culturalism* (Martí 2009).

Martí argues that McMNS's experiment does not in fact probe intuitions about reference. Instead, it probes intuitions about *theories* of reference, which are a different beast. She does not dispute their description of the Gödel case used in the surveys. Rather, she finds fault with the questions used in the surveys. The questions ask people to agree with one of the following statements:

"When John uses the name 'Gödel', is he talking about:

(A) the person who really discovered the incompleteness of arithmetic? Or

(B) the person who got hold of the manuscript and claimed credit for the work."

McMNS claim that answering (A) amounts to a Kripkean intuition and that answering (B) amounts to having a descriptivist intuition. But Martí claims that if they wanted to probe Kripkean vs. descriptivist intuitions, then they have asked the wrong questions. In other words, she claims that their question "does not test the intuitions that could allow us to tell whether or not the participants in the experiment use names descriptively; rather the question tests their opinions as regards which theory of reference determination they think is correct" (44). The question only shows which theory subjects believe accords with how people think they will use names, but people can be wrong about that. And really, only intuitions about the use of names are relevant to the debate between Kripkeans and descriptivists. Since McMNS have not shown that there is cross-cultural variation in the use of names, they have not undermined the use of intuitions in the debate about proper names or presumably any of the other appeals to a theory of reference.

Martí does not point to, but easily could, literature on such phenomena as confabulation (e.g., Berrios 2000; Schneider 2008; Doris 2009) or unconscious determinants of thought (e.g., Bargh/Chartrand 1999; Bargh/Ferguson 2000; Bargh 2005; Bargh/Morsella 2008) to bolster her claim. These phenomena could give inductive evidence that we are not very good at theorizing about our own actions; the phenomena give rise to the question about whether we are good at theorizing about our semantic use of names. Her question, therefore, is empirically grounded and it carries empirical implications and thus is one that the experimental philosopher especially needs to take seriously.

Martí clearly holds that it is possible for empirical work to shed light on the topic: she ends her article by proposing a question designed to probe intuitions about the use of names. Machery, Olivola and de Blanc (2009) take up the challenge: empirical objections should be answered empirically. They first point out that Martí's criticism succeeds only if metalinguistic and linguistic intuitions are not in line with each other. Then they show how that claim can be tested experimentally. They looked at the percentages of people who hold Kripean intuitions about the theory of reference vs. descriptivist intuitions about the use of names by using the following questions (designed to be similar in all but superficial ways to the Gödel case):

> "When Ivy says 'Tsu Ch'ung Chih was a great astronomer,' do you
> think her claim is: (A) true or (B) false?" and "When Ivy uses the
> name 'Tsu Ch'ung Chih,' who do you think she is actually talking
> about: (A) the person who (unbeknownst to Ivy) really determined
> the solstice times? or (B) the person who is widely believed to have
> discovered the solstice times, but actually stole this discovery and
> claimed credit for it?" (691)

They report that among Indian, French, and Mongolian subjects, the difference
does not reach statistical significance, showing that the differences between the
two types of intuitions is not large in these groups (692). Hence, they conclude
that Martí's criticism does not pass the empirical challenge she poses: the expe-
riment shows that metalinguistic and linguistic intuitions match, so the original
MaMNS findings stand.

Here we have seen a type of challenge that addresses the experiment itself and
have shown how it can also be answered experimentally. Unlike the more general
Reflection Defense, this objection will vary from experiment to experiment. Thus,
experimental philosophy brings a new realm of thinking about methodology and
brings a clarification of the original dispute in ways that may not have happe-
ned or have been possible without experimental philosophy. Are people good
at theorizing about their linguistic practices in light of *prima facie* suggestions
to the contrary? How are linguistic intuitions shaped by culture, and what role
does culture play in circumscribing how we use proper names? Can arguments
from reference be saved by modifying the method of cases? Is there a way to
save the method of cases by saying that the reference of a name is relative to the
intuition group to which the speaker belongs? (MlMNS argue that the method of
cases cannot be saved by this move.) Can we know our own intuitions? MlMNS's
interpretation of the debate between Kripeans and descriptivists has also encou-
raged a re-thinking of the role of Kripke's cases. Do MlMS exaggerate the role
of Gödel-type cases at the expense of the more quotidian cases, as Devitt argues
(Devitt forthcoming)? Is there cross-cultural variation in metaphysical intuiti-
ons about modal properties (Devitt forthcoming)? Should philosophers refrain
from the implausible and outlandish when it comes to thought experiments for
new reasons (for an earlier argument, see Wilkes 1988)? Exploring how these
questions can all be addressed is beyond the scope of this paper. Here, my goal
is to show that experimental philosophy uses important tools for asking new
philosophical questions and for looking at old questions in new ways.

Hence, I argue that the negative project of experimental philosophy does
argue that certain doors should be closed on certain central philosophical pro-
jects. But the names 'negative project' and 'experimental restricitvism' do not
do full justice to this project. The negative project does call for an end to cer-
tain assumptions, but it also gives new tools and a new framework and therefore
is not merely negative.Those who do not accept the revisionist element in the
negative project need not, however, abandon all of experimental philosophy, for
the positive project does not contain the revisionist aspect. The positive project
attempts systematically to understand folk intuitions in order to gain new in-

sight into how the mind works and then to gain new perspectives on the original questions. Here the experiments are not used to undermine a particular kind of philosophical debate but rather to show how the debate arises or the structure of concepts.

3. The Positive Project and Free Will

The common assumption in the literature on free will is that the kind of free will that matters is the one that is commonly said to underwrite ordinary attributions of moral responsibility. The reason that *that* kind is in play while others are not is that the goal is to explain ordinary attributions of moral responsibility or show that those attributions are unjustified. So the sense of free will in question is the one that supports moral responsibility. Some compatibilists say our ordinary practices may even turn out to constitute the concept of free will (e.g., Watson 1987; Strawson 2008). Many incompatibilists are at pains to show that a lack of free will would still leave practices like criminal punishment with sufficient justification (e.g., Pereboom 2001). A technical concept that is divorced from the ordinary practices is undesirable, as that would then not be *the* problem of free will, the one that is *so* central and so vexing. An analysis of the doctrine of physicalism does not have to conform to ordinary notions or practices in the way that an analysis of free will does: the question of physicalism does not get its import from such a direct tie to our ethical lives.

From the outset, we can note that since these discussions target the ordinary concept of free will and moral responsibility, the experimental philosopher who examines folk intuitions is uniquely protected from the Reflection Defense. It is worth reiterating here that experimental philosophy is really a class of different projects: objections to one project do not undermine the entire enterprise (for a similar argument, see Nadelhoffer/Nahmias 2007). The idea here is not to give an analysis of free will, a set of necessary and sufficient conditions by coming up with something that conforms to every set of thought experiments. Rather the idea is to develop an understanding of free will grounded in practices of ascribing moral responsibility that shows whether or not it is compatible with determinism.

Nahmias, Morris, Nadelhoffer and Turner examine folk intuitions about compatibilsm and incompatibilism both as an object of interest in their own right and to support, albeit indirectly, compatibilism, to strengthen the compatibilists's rhetorical position (2006). Incompatibilism is more metaphysically demanding, they contend, since it requires a libertarian conception of free will if determinism is true. Moreover, they hold, metaphysical demand should be on a need only basis. Since incompatibilism is more metaphysically demanding, there support for it is weakened unless it is the intuitive notion. Hence, if incompatibilism is not intuitive, compatiblism, although not resolved, does carry a large advantage.

They examine whether incompatibilism is the intuitive view through a series of experiments. Subjects are presented with a determinist universe (although not in those terms exactly since determinism is often taken to be the opposite of free

will) and asked whether an agent in that universe is free or morally responsible. They found that the majority of people do think that the agent is free and responsible, regardless of whether the action is negative, positive, or neutral, and independently of various ways that determinism may be described.

To the extent that these results are definitive, philosophers who claim that incompatibilism is intuitive should be concerned. After all, if it is undesirable to assert irrelevant falsehoods it is even less desirable to assert falsehoods that carry argumentative weight. So NMNT show the importance of whether incompatibilism is intuitive along with giving preliminary evidence that it isn't.

That evidence is cast in a new light by Nichols's and Knobe's work (2007). NK are struck by the fact that many philosophers claim that ordinary people are natural incompatibilists even though NMNT's results and other experiments (e.g., Viney/Parker-Martin et al. 1988; Woolfolk/Doris et al. 2006) seem to indicate that compatibilism is the natural view. They think that something more subtle is going on: our natural incompatibilism can be undermined or changed by affective reactions.

To test this, they designed a study that postulates that there is a deterministic universe, one they call Universe A. They assigned subjects to one of two conditions in order to compare levels of compatibilist vs. incompatibilist responses. In the concrete condition, subjects answered the question following this scenario (111):

> "In Universe A, a man named Bill has become attracted to his se-
> cretary, and he decides that the only way to be with her is to kill
> his wife and three children. He knows that it is impossible to escape
> from his house in the event of a fire. Before he leaves on a business
> trip, he sets up a device in his basement that burns down the house
> and kills his family. Is Bill fully morally responsible for killing his
> wife and children?
>
>   YES   NO"

In the abstract condition, subjects answered this:

> "In Universe A, is it possible for a person to be fully morally respon-
> sible for their actions?
>
>   YES   NO"

Again, their concern is to see if these two questions bring out a difference in intuitions. They found very strong evidence that it does: in the concrete condition, 72% gave the compatibilist response; in the abstract condition, 86% gave the incompatibilist response (111). The folk are not simply compatibilists, as other experiments seem to indicate.

At this point, it may be tempting for the non-experimentally minded philosopher to object as follows. Certainly NK have shown that patterns of responses vary across conditions, but what does this have to do with the philosophical question of whether free will is compatible with determinism? This objection

may continue in one of two directions: the first says that the experiments are not in and of themselves philosophically relevant and the second says that the experiments are not philosophically relevant at all.

To the part of the objection that says that the experiments alone are not philosophically relevant, the reply is that NK are not interested in the experiments alone. As we will soon see, NK don't stop with a simple description of intuition patterns. The go on to ask about what role affect should have in our moral reasoning and whether affect presents a kind of performance error. Surely that is a philosophical question if any is.

To the part of the objection that says that experiments are not relevant at all, there are two replies. First, remember the point made at the beginning of this section, namely that the philosophical question gets its stronghold in ordinary practices. What ordinary practices are is an empirical question. There does not seem to be any principled reason why philosophers can't do their own empirical work. Second, experimental philosophy has shed light on new questions about how the mind works, specifically whether affect changes how we think about moral responsibility in a determinist universe. It addresses the issue of what factors make us reason the way we do. If this does not directly answer the question about whether free will is compatible with determinism, this is only because it has helped show how complex the question is and how it bears on so many other issues. We are in a much better position to answer whether free will is compatible with determinism if we know what our cognitive biases are and strive to eliminate them first.

Let us return to the topic of what NK derive from the experiments. They consider three broad models that might explain their results. The first, the performance error model, aligns the results with other results in social psychology that show that emotion impedes our ability to reason. On this model, the compatibilist results only demonstrate an error in reasoning, not a true theoretical belief in compatibilism (112–3). The second model says that affect provides a kind of competence. By looking at psychopaths, we see that affective reactions are necessary to understand moral reasoning. Hence, the responses given in the cases with high affect demonstrate our true competence. NK also consider hybrid models that assign to affect a role in generating competence while also generating error in certain situations (114–6). The third model agrees with the second that our competence is demonstrated in the concrete case, but denies that affect is the driving force. Perhaps we have an innate 'moral reasoning module' that does not communicate with the rest of the mind enough to interact with the theoretical understanding that a determinist universe is under consideration (114–5).

Teasing out which model is best is a complicated affair. To give evidence, NK look at another experiment. Specifically, they ask whether concreteness fully explains the compatibilist responses or whether affect is the driving force by holding the concreteness the same in both cases and only changing the level of affect. Subjects in this experiment consider again Universe A, but are asked one of the following: "As he has done many times in the past, Bill stalks and rapes a stranger. Is it possible that Bill is fully morally responsible for raping

the stranger?" "As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully morally responsible for cheating on his taxes?" The results are striking: 64% hold Bill responsible, while only 23% hold Mark responsible (117).

Since affect clearly plays a role, the question is whether it generates a competence or an error. They argue that the experiments support the performance error (or hybrid) model best. The performance error model makes better sense of the different responses when subjects consider Universe A vs. an indeterministic universe. 89% of subjects hold Mark responsible in the indeterminist universe versus the 23% in the determinist universe. 95% consider Bill morally responsible in the indeterminist universe, but in the determinist universe, the drop to 64% is nowhere near as dramatic. This provides evidence for saying that subjects are responsive to the difference between a determinst and indeterminst universe except when their affect skews their understanding (118).

So NK take this as preliminary evidence for the performance error model. They argue that this issue is not decided just by their experiment and is not the kind of question that can be resolved by one experiment alone (118). But they have shown that both compatibilism and incompatibilism "appeal to an element of our psychological makeup" (119). Plus, they have successfully argued that we can no longer declare *simpliciter* that people are natural compatibilists: those who think that compatibilist intuitions should be given more weight need to give arguments as to why. Similarly, those who say we are natural incompatibilists need to explain why the compatibilist intuitions are not relevant to that claim.

Finally, I would argue that this type of investigation opens up ways of looking at successor views to hard determinism (Double 1991; Strawson 1994; e.g. Smilansky 2000; Pereboom 2001; Honderich 2002; Mele 2002) more empirically. That is, it allows us to make empirically based arguments about conclusions that some new determinists derive non-empirically. For example, Smilansky (2000) argues that if we start from the 'Core Conception' of justice, we find that both determinism and compatibilism are insufficient. Hence we need a dualism that encompasses both. He makes his arguments on ethical and metaphysical grounds, but perhaps KN's assertion that there are elements of our psychological makeup that lead us in both directions can be seen as providing psychological grounding for Smilansky's claims, or as generating similar conclusions from different directions making it psychologically respectable.

Again, we have an instance of the positive project of experimental philosophy deepening the discussion of an old question and generating new sources of evidence and generating new questions. Unlike the negative project, experimental philosophers here aren't calling for an end to using intuitions: rather they are calling for a wiser use of them.

But what if the difference between the concrete and abstract cases does not reveal anything about the folk concept of moral responsibility, but rather uncovers a merely verbal dispute? This is exactly the challenge that Sosa poses toward NK's work (2007). Sosa argues that it is possible that the two conditions give rise to different patterns of response due to two different senses of moral responsibility. One sense, the attributability sense, is "inherently incompatibilist

in requiring only that the agent have caused his action, free of antecedent determinants, *free even of determination by his or her character*" (236, emphasis in original). The other accountability sense only requires that the person is properly held accountable for doing the action, that the person is a proper subject of praise or blame (236). Sosa's challenge then amounts to saying that all that may be involved in the NK studies is verbal disputes. Just as we have not discovered anything interesting about my newly purchased 2000 model car if we find that certain people call it new (in the condition where it is emphasized that I am driving it for the first time) and that others don't (in the condition where the fact that the model is 9 years out of date is emphasized), we also have not discovered anything about whether free will is compatible with determinism if we find that in certain conditions one sense of moral responsibility is invoked and in another condition another sense is invoked.

This is a more sophisticated version of the objection that surveys are not relevant for the philosophical debate. It can be met with the reply that *even if* NK have uncovered a verbal dispute only, they have revealed in an empirically respectable way that there is no one folk concept, contrary to the implicit assumption often made in the literature on free will when it is claimed that the folk are naturally incompatibilist (e.g., Strawson 1986; Kane 1999; Vargas 2006). They have, that is, highlighted the fact that the folk reason in ways that are guided by different concepts. And in so doing they have uncovered interesting facts about how the mind works. If there has been all along an unrecognized verbal dispute, that would be important to know.

The experimental philosopher doesn't, however, even need to grant that much. Characteristically, merely verbal disputes like the one about my new car (Sosa's example involves saying Mary went to the bank yesterday on different days) are easily eliminated because they are *merely* verbal. No deep philosophical debate about the concept of 'newness' arises from this debate. Discussions of free will are different, because at the end of the day we still need to decide whether someone was morally responsible. If I steal your pen and you blame me, it isn't going to go very far for me to point out that our dispute is verbal. This is not to say that it is impossible for there to be multiple senses of terms, or to say that people can't make mistakes, or to say that it is impossible to have a merely verbal dispute about whether someone is morally responsible. For such a confusion to have been the case in the NK survey, though, it would have to be likely that we could get someone together who said Bill is morally responsible with someone who said Mark is not and for them to say 'oh, we aren't really arguing about whether someone can be responsible in universe A'. It is an empirical question as to whether the folk would assent to that, and I do not have evidence, but I think it is highly implausible that it would happen. It seems highly implausible that people switch senses of moral responsibility when they switch from abstract to concrete descriptions of the same action. Hence I do think it is very unlikely for something so central and deep to be a matter of merely verbal disputes. NK have found a tension in a concept that is central and deep, so it seems implausible that the disputes they have uncovered could be merely verbal.

4. The Positive Project and Intentional Action

Folk psychology forms the basis for much reflection in the philosophy of mind.
Folk psychology is the way that people attribute beliefs, desires, intentions and
other psychological concepts to themselves and to others in order to make their
way about in the world, specifically to explain and predict behavior. One seminal
paper in experimental philosophy explores the folk concept of an intentional
action and leads to a surprising understanding that has lead philosophers to
rethink the nature of folk psychology.

The paper is Knobe's *The Concept of Intentional Action: A Case Study in
the Uses of Folk Psychology* (2006). He focuses on the concept of an intentio-
nal action to show that the aims of folk psychology are not entirely congruent
with the aims of scientific psychology. Knobe's goal is to show that folk psy-
chology, contrary to the vast consensus, is not simply to explain and predict
behavior. Instead, intentional action has a moral dimension. Knobe establishes
this through an experiment that generates what has come to be known as 'the
Knobe effect'. Subjects were given one of two vignettes (differences are indicated
in parentheses).

> "The vice-president of a company went to the chairman of the board
> and said, 'We are thinking of starting a new program. It will help us
> increase profits, but it will also harm (help) the environment.' The
> chairman of the board answered, 'I don't care at all about harming
> (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 harmed (helped). Now ask yourself:
> Did the chairman of the board *intentionally* harm the environment?"
> (130–1)

Knobe found vastly different intuitions depending on whether the moral com-
ponent is positive or negative. In the harm scenario, 82% of subjects said the
chairman harmed the environment intentionally, while in the help scenario, only
23% said he helped the environment intentionally (131). The key result is that
"people seem to be considerably more willing to say that the agent brought about
the side effect intentionally when they regard that side effect as bad than when
they regard the side effect as good" (133). The result that moral considerations
play a role in our attribution of intentionality is supported by much other work
in philosophy and psychology.

Like NK, Knobe here wants not only to describe patterns of intuitions, but
also to use those patterns to better understand how the mind works. He asks
what explains the effect he found: is it that our folk concept really is morally
laden, or should we rather say that it is not but that somehow moral considera-
tions intervene. He considers and rejects three models of how the latter option
would come into play, bolstering his claim that the moral considerations play a
key role in our folk psychological concept of intentional actions.

One model, proposed by Mele, says that our competence regarding the term
'intentional' might be side-railed by an explicit but erroneous belief that all

blameworthy actions are intentional (2001). Knobe tested this claim by having people assess that an unintentional behavior by a drunk person was nonetheless blameworthy. If Mele is right, then immediately after seeing that, the Knobe effect should be diminished, but Knobe found that it is in fact not diminished, thereby leading him to reject this model (135–6).

Another model says that conversational pragmatics gives rise to the effect. Adams and Steadman say that to say 'He didn't do that intentionally' is to imply that he is not blameworthy (2004). So intentional could still be a non-moral concept if they are right. Knobe puts this model to the test by seeing if the same effect arises if you remove the word 'intentionally', substituting it with 'in order to'. The same effect happens in this case as well (137).

One final model, proposed by Nadelhoffer (2004) and by Malle and Nelson (2003), says that the effect is a result of the effect of blame. Specifically, people erroneously assign blame before they decide whether an action was intentional. This would explain the data, but Knobe says, another more reasonable model also explains the data. This model says that "people's judgment that the behavior itself is bad can influence their intuitions as to whether the behavior was performed intentionally and that these intuitions can, in turn, play an important role in the process by which people determine whether or not to assign blame" (139). Knobe shows this by proposing another experiment, since both of these models make the same prediction with regard to the previous experiment. This experiment presents a case where the side effect is bad but the agent is not blameworthy. Here, the chairman of the board decides to implement a program that will greatly increase sales in Massachusetts but decrease sales in New Jersey. Even though there is a sense in which decreasing sales in New Jersey is bad, the chairman isn't thought to be blameworthy for doing it: after all, she deserves praise for increasing sales overall. Knobe found that in this case as well, people say that she intentionally decreased sales in New Jersey. Hence Knobe argues, his model better fits the patterns of responses in the data. Knobe urges that it is better to revise our understanding of folk psychology and acknowledge that the concept of intentional action is not simply a tool for predicting and explaining behavior.

In more recent work, Knobe uses not only these results but a vast array of other similar results about moral terms such as but also even concepts like causation to argue that moral considerations are key to our competencies regarding all of these terms, that we should think of people as moralists when they display their competencies, not simply as scientists (2009). Moral considerations, on his view, do not distort our folk psychology. They do not interfere with our amoral rational processes. Rather they are inherent in and inseperable from our very basic attempts to understand and explain the world.

But what if survey results are completely irrelevant to uncovering the nature of folk psychology: what if the survey model is so fundamentally flawed that it does not in fact reveal folk intuitions but rather just survey responses? This is what Kauppinen argues in his paper *The Rise and Fall of Experimental Philosophy* (2007). Kauppinen distinguishes between what he calls robust and surface intuitions and says that armchair philosophy aims and claims to use robust in-

tuitions whereas experimental philosophy only uncovers surface intuitions. His critique of surface intuitions has several parts, one important one of which has not yet been touched on so far in this exposition and applies at this juncture, so I will focus on that (see also Nadelhoffer/Nahmias 2007). Specifically, Kauppinen claims that robust, or genuinely philosophical intuitions are made under 'ideal conditions'.

Kauppinen emphasizes that ideal conditions are important because even when incompetent users are eliminated from the discussion, performance errors can arise (103). Ideal conditions are conditions that are conducive to avoiding performance errors and competent users are those who "apply the concept correctly to a sufficient number of cases" (102–3). Experimental philosophy does not get at what competent users say under ideal conditions. Instead, it only gets at what non-specialists who appear to understand the question would say under whatever circumstances or conditions they happen to be in, regardless of what considerations influence their response (105). Philosophers should instead, he says, use the Dialogue Model. This involves eliciting intuitions, but then subjecting them to dialogue. For example Knobe might ask someone whether the executive intentionally harmed the environment, but then if the person says yes, he would ask how that is consistent with the belief that the other executive did not intentionally help the environment. Only in this way would he uncover whether people robustly believe that there is an asymmetry. It may turn out that most people will decide that their answer that the CEO intentionally harmed the environment was the result of being distracted by the harm done, that once they realize that they were influenced incorrectly by thinking about hurt animals or the like, they will come to say that the CEO did not intentionally harm the environment.

Clearly, the Knobe effect would be highly suspect if Knobe's experiments did not elicit genuine folk intuitions. But is Kauppinen operating with the same understanding of a folk intuition as the experimental philosopher? For Kauppinen, folk concepts cannot be derived from what people say in response to survey questions because the conditions are not ideal. However, experimental philosophy relies on manipulation checks and experimental philosophers administer the surveys in a way that minimizes subject fatigue and other performance errors (except when, of course, the idea is to see whether something like affect produces a performance error). For Kauppinen, these conditions still fall short, and only philosophical dialogue can remedy the situation. Kauppinen acknowledges that there is the possibility of introducing bias and recommends avoiding it. But since philosophical dialogue inherently introduces philosophical sophistication, by subjecting survey answers to scrutiny through dialogue, we are no longer getting at folk conceptions. We are deepening them, changing them, shaping them. Ordinary intuitions and philosophical beliefs differ in fundamental ways. The later are more sophisticated, consistent and rational. Even if the dialogue Kauppinen recommends does not involve introducing philosophical theories, it still involves introducing a philosophical style of reflection. So in the course of the dialogue folk intuitions lose their folk character.

Moreover, the change that folk intuitions undergo during dialogue is likely to be temporary. That is, I believe that typically, people can and often do decide in the dialogue that Kauppinen suggests, that the CEO did not intentionally harm the environment, but then a few days or weeks or months later can and often do go on to attribute intention to action on the basis of moral considerations. Mere reflection does not overcome optical illusions; similarly, subjecting the folk to temporary dialogue is not likely to fundamentally, permanently change their folk intuitions.

Here are some examples that show the insufficiency of conscious reflection for shaping behavior. First, Bargh, Chen and Burrows show that subjects who perform a decoy task that involves many words associated with being elderly walk more slowly to the elevator after the experiment is supposedly over than subjects who do not read words associated with being elderly (1996) (for further discussion, see Doris 2009). If something so fundamentally voluntary as gait is influenced by automatic processes, there seems to be strong reason to think that reflective beliefs, which require time, attention and memory, are also going to be influenced by automatic processes outside the context of dialogue. Second, the wide body of work on psychological distance shows that whether you imagine a task as taking place here or now vs. there or then can have profound effects on how you categorize and construe the world (e.g., Liberman/Trope 1998; Bar-Anan/Liberman et al. 2006; Liberman/Trope et al. 2007; Liberman/Trope 2008; Liviatan/Trope et al. 2008; Jia/Hirt et al. 2009). Similarly, philosophical reflection is not likely to overcome all of these social and psychological forces. Whether philosophical reflection overcomes confounding forces outside the context of dialogue is an empirical issue, and I am open to the possibility that I may be wrong. But if I am right, then the dialogue model is not getting at folk concepts: folk concepts are the concepts that are used in everyday life, not just in the armchair or the study.

In short, Kauppinen's ideal conditions are inherently not ordinary conditions and as such they do not reveal folk concepts, the concepts that are a part of true folk psychology. This issue goes deep into the heart of questions about human nature and my take on it obviously reveals my experimental bent, so it is clearly not conclusive. Many empirical considerations come into play and there are many open questions. I hope I have shown, however, that the ideal conditions argument rests on a likely false assumption about human abilities because the ideal conditions argument requires that we shape folk concepts in a way that makes them not what the philosopher who grounds an argument in everyday practices requires.

I believe that the positive and negative projects of experimental philosophy are very different in terms of the conclusions they draw, to the extent that they should really be treated distinctly. What unites them is the fact that philosophers have taken to using controlled, systematic experiments. This at the same time explains why they are treated as the same because it is a radical difference from what has gone before. 'What has gone before' requires a caveat. Many experimental philosophers have argued that what they are doing is a return in spirit to what philosophers have always done. Appiah treats 20th century analytic

philosophy's emphasis on a priori, armchair reflection as an abberation (2008, see especially 5–32). KN, while granting that using the methods of 20th–21st century experimental psychology is new, argue that it is really not relevantly different from what philosophers have always done: namely, seek to understand human nature using whatever tools are available (2008, see especially 13–14).

I therefore argue that at best the arguments against the negative project of experimental philosophy limit its scope or say that it shows something other than it thinks, but nonetheless in any case it always teaches us something important. Arguments against the positive project can be met by explaining the role that folk intuitions play in these debates. Hence, even if all experimental philosophy makes the mistakes the critics accuse it of making, it has forced the armchair philosophers to clarify their projects and it has opened new avenues of investigation, and that is why it is here to stay.

Bibliography

Adams, F./A. Steadman (2004), Intentional Action in Ordinary Language: Core Concept or Pragmatic Understainding?, in: *Analysis 64*, 173–181

Appiah, K. A. (2008), *Experiments in Ethics*, Cambridge/MA

Bar-Anan, Y./N. Liberman/Y. Trope (2006), The Association between Psychological Distance and Construal Level: Evidence from an Implicit Association Test, in: *Journal of Experimental Psychology 135(4)*, 609–622

Bargh, J. A. (2005), Bypassing the Will: Toward Demystifying the Nonconscious Control of Social Behavior, in: R. R. Hassin/J. S. Uleman/J. A. Bargh (eds.), *The New Unconscious*, Oxford, 37–58

— /M. C. Burrows (1996), Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action, in: *Journal of Personality and Social Psychology 71(2)*, 230–244

— /T. L. Chartrand (1999), The Unbearable Automaticity of Being, *American Psychologist 54(7)*, 462–479

— /M. J. Ferguson (2000), Beyond Behaviorism: On the Automaticity of Higher Mental Processes, in: *Psychological Bulletin 126(6)*, 925–945

— /E. Morsella (2008), The Unconscious Mind, in: *Perspectives on Psychological Science 3(1)*, 73–79

Berrios, G. E. (2000), Confabulations, in: G. E. Berrios/J. R. Hodges (eds.), *Memory Disorders in Psychiatric Practice*, Cambridge, 348–368

Cacioppo, J. T./R. E. Petty (1982), The Need for Cognition, in: *Journal of Personality and Social Psychology 42(1)*, 116–131

Devitt, M. (forthcoming), Experimental Semantics, in: *Philosophy and Phenomenological Research*

Doris, J. (2002), *Lack of Character*, Cambridge

— (2009), Skepticism about Persons, in: *Philosophical Issues 19(1)*, 57–91

Double, R. (1991), *The Non-Reality of Free Will*, Oxford

Gettier, E. (1963), Is Justified True Belief Knowledge?, in: *Analysis 23*, 121–123

Haidt, J./S. H. Koller/M. G. Dias (1993), Affect, Culture, and Morality, or Is It Wrong to Eat Your Dog?, in: *Journal of Personality and Social Psychology 65(4)*, 613–628

Honderich, T. (2002), Determinism as True, Compatibilism and Incompatibilism as False, and the Real Problem, in: R. Kane (ed.), *The Oxford Handbook of Free Will*, Oxford, 461–476

Jia, L./E. R. Hirt/S. C. Karpen (2009), Lessons from a Faraway Land: The Effect of Spatial Distance on Creative Cognition, in: *Journal of Experimental Social Psychology 45(5)*, 1127–1131

Kane, R. (1999), Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism, in: *Journal of Philosophy 96*, 217–240

Kauppinen, A. (2007), The Rise and Fall of Experimental Philosophy, in: *Philosophical Explorations 10(2)*, 95–118

Knobe, J. (2006), The Concept of Intentional Action: A Case Study in the Uses of Folk Psycology, in: Knobe/Nichols (2007), 129–147

— (2009), Person as Scientist, Person as Moralist, unpublished manuscript

— /S. Nichols (2008), An Experimental Philosophy Manifesto, in: Knobe/Nichols (2007), 3–16

— / — (2007) (eds.), *Experimental Philosophy*, Oxford

Lehrer, K. (1990), *Theory of Knowledge*, Boulder

Liberman, N./Y. Trope (1998), The Role of Feasibility and Desirability Conditions in Near and Distant Future Decisions: A Test of Temporal Construal Theory, in: *Journal of Personality and Social Psychology 75(1)*, 5–18

— / — (2008), The Psychology of Transcending the Here and Now, in: *Science 322*, 1201–1205

— / — /E. Stephan (2007), Psychological Distance, in: E. T. Higgins/A. W. Kruglanski (eds.), *Social Psychology: A Handbook of Basic Principles*, Guilford, 353–381

Liviatan, I./Y. Trope/N. Liberman (2008), Interpersonal Similarity as a Social Distance Dimension: Implications for Perception of Others' Actions, in: *Journal of Experimental Social Psychology 44*, 1256–1269

Machery, E./R. Mallon/S. Nichols/S. Stich (2004), Semantics Cross-Cultural Style, in: Knobe/Nichols (2007), 47–58

— /C. Y. Olivola/M. de Blanc (2009), Linguistic and Metalinguistic Intuitions in the Philosophy of Language, in: *Analysis 69(4)*, 689–694

Malle, B. F./S. E. Nelson (2003), Judging Mens Rea: The Tension between Folk Concepts and Legal Concepts of Intentionality, in: *Behavioral Sciences and the Law 21*, 563–580

Mallon, R./E. Machery/S. Nichols/S. Stich (2009), Against Arguments from Reference, in: *Philosophy and Phenomenological Research 79(2)*, 332–356

Martí, G. (2009), Against Semantic Multi-Culturalism, in: *Analysis 69(1)*, 42–48

Mele, A. (2001), Acting Intentionally: Probing Folk Notions, in: B. F. Malle/L. J. Moses/D. Baldwin (eds.), *Intentions and Intentionality: Foundations of Social Cognition*, Boston, 27–44

— (2002), Autonomy, Self-Control and Weakness of Will, in: R. Kane (ed.), *The Oxford Handbook of Free Will*, Oxford, 529–548

Milgram, S. (1974), *Obedience to Authority*, New York

Nadelhoffer, T. (2004), Praise, Side Effects, and Intentional Action, in: *Journal of Theoretical and Philosophical Psychology 24*, 196–213

— /E. Nahmias (2007), The Past and Future of Experimental Philosophy, in: *Philosophical Explorations 10(2)*, 123–149

Nahmias, E./S. G. Morris/T. Nadelhoffer/J. Turner (2006), Is Incompatibilism Intuitive?, in: Knobe/Nichols (2007), 81–104

Nichols, S./J. Knobe (2007), Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions, in: Knobe/Nichols (2007), 105–126

— /S. Stich/J. Weinberg (2003), Metaskepticism: Meditations in Ethno-epistemology, in: S. Luper (ed.), *The Skeptics*, Farnham, 227–247

Nisbett, R. E. (2003), *The Geography of Thought: How Asians and Westerners Think Differently ... and Why*, New York

— /A. Norenzayan (2002), Culture and Cognition, in: D. Medin/H. Pashler (eds.), *Stevens' Handbook of Experimental Psychology: Memory and Cognitive Processes 2*, 561–597

— /K. Peng/I. Choi/A. Norenzayan (2001), Culture and Systems of Thought: Holistic versus Analytic Cognition, in: *Psychological Review 108(2)*, 291–310

Norenzayan, A./R. E. Nisbett/E. E. Smith/B. J. Kim (1999), *Rules vs. Similarity as a Basis for Reasoning and Judgment in East and West*, University of Michigan

Pereboom, D. (2001), *Living without Free Will*, Cambridge

Prinz, J. (2008), Empirical Philosophy and Experimental Philosophy, in: Knobe/Nichols (2007), 189–208

Schneider, A. (2008), *The Confabulating Mind: How the Brain Creates Reality*, Oxford

Shope, R. (1983), *The Analysis of Knowing*, Princeton

Smilansky, S. (2000), *Free Will and Illusion*, Oxford

Sosa, E. (2005), A Defense of the Use of Intuitions in Philosophy, in: M. Bishop/D. Murphy (eds.), *Stich and His Critics*, Oxford, 101–112

— (2007), Experimental Philosophy and Philosophical Intuition, in: Knobe/Nichols (2007), 231–240

Strawson, G. (1986), *Freedom and Belief*, Oxford

— (1994), The Impossibility of Moral Responsibility, in: *Philosophical Studies 75(1/2)*, 5–24

Strawson, P. F. (2008), Freedom and Resentment, in: *Freedom and Resentment and Other Essays*, New York, 1–28

Swain, S./J. Alexander/J. Weinberg (2008), The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp, in: *Philosophy and Phenomenological Research 76(1)*, 138–155

Vargas, M. (2006), On the Importance of History for Responsible Agency, in: *Philosophical Studies 127*, 351–382

Viney, W./P. Parker-Martin/S. D. H. Dotten (1988), Beliefs in Free Will and Determinism and Lack of Relation to Punishment Rationale and Magnitude, in: *The Journal of General Psychology 115(1)*, 15–23

Watson, G. (1987), Responsibility and the Limits of Evil, in: F. D. Schoeman (ed.), *Responsibility, Character, and the Emotions*, Cambridge, 256–286

Weinberg, J./S. Nichols/S. Stich (2001), Normativity and Epistemic Intuitions, in: Knobe/Nichols (2007), 17–45

— /J. Alexander/C. Gonnerman (forthcoming), Unstable Intuitions and Need for Cognition: How Being Thoughtful Sometimes Just Means Being Wrong in a Different Way, in: *Philosophical Psychology*

Wilkes, K. V. (1988), *Real People: Personal Identity Without Thought Experiments*, Oxford

Woolfolk, R. L./J. M. Doris/J. M. Darley (2006), Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility, in: Knobe/Nichols (2007), 61–80