

THE FOLK PROBABLY DO THINK WHAT YOU THINK THEY THINK¹

Billy Dunaway, Anna Edmonds, and David Manley

Much of contemporary experimental philosophy involves taking surveys of ‘folk’ subjects to test their intuitions involving philosophically relevant concepts. The results of these surveys are often claimed to be surprising, and treated as evidence that the relevant folk intuitions cannot be predicted from the ‘armchair’. We conducted an experiment to test these claims, and found that a solid majority of philosophers could predict even results that were claimed to be surprising in the literature. We discuss some methodological implications as well as some possible explanations for the common surprisingness claims.

How good are philosophers at predicting ordinary intuitions involving philosophically relevant concepts? This question has recently been the subject of considerable debate, without being directly tested. Some armchair philosophers have reported moderate optimism about their skills in this area, encouraged by the thought that they were once non-philosophers themselves (that is, members of ‘the folk’) and typically still interact with the folk on a regular basis.² But this kind of optimism has come under heavy fire from within the ranks of experimental philosophy. Armchair philosophers, it is claimed, simply ‘assume without evidence that they know what the folk think’, ‘merely speculate’ about folk intuitions, or simply place ‘their *own intuitions* into the mouths of the folk in a way that supports their own position’.³ As a result, it is claimed, debates in metaphysics (for example) are only ‘*nominally* constrained by so-called ordinary notions of identity, free will, and the like, while *really* [they are] checked only by the often peculiar intuitions of metaphysicians themselves’ [Livengood and Machery 2007:108]. This attitude is summed up nicely by Adam Feltz [2009: 203]: ‘Experimental philosophers normally hold that philosophers are not very good at knowing from the armchair... which intuitions are widely shared.’

This pessimism about armchair access to folk intuitions appears to be based on a number of widely heralded results from the branch of experimental philosophy devoted to conducting surveys among non-philosophers. These results are said to contrast with what armchair philosophers would have

¹We would like to thank Steve Campbell, David Chalmers, Kenny Easwaran, Sarah Moss, Ángel Pinillos, Chandra Sripada, David Weins, and many others for helpful suggestions and discussion. The project was supported by a Rackham Research Grant.

²See, for example, [Jackson 1998: 36ff].

³The first remark is from [Livengood and Machery 2007: 107]; the second is from [Alexander and Weinberg 2007: 73]; and the third is from [Nahmias et al. 2005: 562].

predicted, and this is taken in turn to be a symptom of the ignorance of philosophers about folk intuitions:

Again and again, these investigations have challenged familiar assumptions, showing that people do not actually think about these issues in anything like the way philosophers had assumed. [Knobe and Nichols 2007: 3]

Jonathan Livengood and Edouard Machery describe the ‘surprising’, ‘striking’ and ‘unexpected’ results of their folk surveys,⁴ and draw the following lesson (the first clause of which is also the title of their paper):

The folk probably don’t think what you think they think; so rather than guess from the comfort of your armchair, you ought to go out and check. [126]

In short, this line of reasoning has led to a claim that is ‘widely endorsed by experimental philosophers’: *only* survey-based empirical research ‘can deliver the intuitions that can serve as evidential basis for or against philosophical claims’.⁵

Armchair philosophers have resisted this conclusion, but usually without challenging the assumption that the survey results at issue were *surprising* in the relevant sense— i.e. *not predictable by philosophers*.⁶ Instead, the typical response has been to point out other ways in which the survey results might not be as significant as they are made out to be. For example, the relevant survey responses may be driven by pragmatic rather than semantic considerations.⁷ Or maybe the folk were employing different concepts from philosophers when they answered the survey questions.⁸ Or again, maybe the responses can be explained by conceptual mistakes, performance errors, unconscious biases, or a simple failure to adequately reflect on the survey questions.⁹ (For example, see Pinillos et al. [2011], which reports that among the non-philosophers in their study, those

⁴[108, 121]. See pp. 8-11 below on the applicability of these adjectives.

⁵[Alexander and Weinberg 2007: 61]. See also [Nadelhoffer and Nahmias 2007: 125]; [Stich and Weinberg 2001: 640].

⁶Though at times the concession may be facetious: see for example Sosa’s description (in his [2007: 104]) of what he calls the ‘shocking results’ and ‘amazing divergence’ described by Knobe and Nichols [2007]. That there would be a divergence in the relevant surveys was predicted by 85.5% of philosophers in our study (see §2 below).

⁷For example, see [Adams and Steadman 2004a, 2004b]; [McCann 2005]; and especially [Kauppinen 2007]; though see also [Knobe 2003b] and [Nadelhoffer 2004]. Capellen [2012] argues that what appear to be appeals to intuition are often not.

⁸For example, see [Goldman 2001: 477]; [Jackson 2001: 661]; [Sosa 2007: 102]. Sosa argues that some of the relevant results may only reflect differences in how people use certain words like ‘knowledge’ (for example) as opposed to disagreements over the nature of knowledge itself.

⁹See also [Mele 2001, 2003]; [Jackson 2001: 662]; [Alicke 2008], and the discussion in [Nichols and Knobe 2007].

who were less reflective and less aware of their own tendency to make mistakes were more likely to display the ‘Knobe effect’.)

We are sympathetic to some of these defensive moves, and will have more to say about them in §3 below. But we think it is also worth challenging the assumption that the relevant survey results were not predictable by philosophers. After all, this assumption must be in place if such results are to illustrate the ignorance of philosophers about how the folk employ philosophical concepts. And since the truth of this assumption is clearly an empirical matter, we designed a study to shed some light on it.

1. Four case studies

Our hypothesis was that philosophers would, for the most part, correctly guess the responses of non-philosophers to surveys in the experimental philosophy literature, even where those responses had been claimed to be surprising. This hypothesis was confirmed by our study.

We chose four published surveys of folk subjects that (i) had been claimed in the literature to have surprising results, and (ii) collectively covered a variety of topics that have received significant attention from experimental philosophers. We used the following method to identify our studies. We began by searching for claims of surprisingness in the literature; when we identified such a claim, we included the study referenced as surprising unless it concerned a topic already covered by an included study, or was too complex to incorporate into our short survey. We also omitted cross-cultural studies (see §3). We stopped our search when we had identified the desired number of studies.

Some of the surveys we included are well-known, but respondents were firmly instructed in a very salient way to opt out of a given question if they found it ‘familiar’ or if their ‘answer might be influenced by prior exposure to results involving similar cases’.¹⁰ (We also took steps, as far as possible, to avoid communicating to our respondents that the folk surveys we were asking about had already been conducted: see §2.)

Our study set a high bar for philosophers’ ability to predict folk intuitions. To begin with, we included only surveys whose results had been touted as surprising. But even the set of published surveys does not cover a random sample of folk intuitions, because of publication bias. It is plausible that researchers are more likely to write up a survey and submit it, and more likely to have the

¹⁰In addition to a bold-faced warning on the instructions page, each multiple-choice question had the option to choose ‘unable to provide an unbiased answer’. The numbers of people who selected this option, in order of the questions presented below, are: 32, 83, 3, 4, 31, 31.

resulting paper accepted, if the result is perceived as surprising or striking.¹¹ Thus even if philosophers had failed to perform well on our survey, that would not have established any sweeping claims about the inability of philosophers to predict folk intuitions in general.

Our survey was completed by 200 faculty and graduate students in philosophy departments in the English-speaking world.¹² We asked these philosophers to suppose that ordinary, non-philosophical folk are presented with the relevant cases, and to say how they thought the folk would respond. 77% to 95.8% of philosophers correctly predicted the surprising results. We provide details below. To help avoid confusion, we will reserve the expression ‘subjects’ for folk subjects of the original studies, and ‘respondents’ for philosophers who answered our meta-survey.

KNOBE AND FRASER ON CAUSATION AND MORALITY

The first question on our survey involved a study reported in the 2008 paper ‘Causal Judgment and Moral Judgment: Two Experiments’ by Joshua Knobe and Ben Fraser. The paper begins with the following discussion of the ‘surprising results obtained in recent studies’ about causation:

It has long been known that people’s causal judgments can have an impact on their moral judgments...But recent experimental work points to the existence of a second, and more surprising, aspect of the relationship between causal judgment and moral judgment. It appears that the relationship can sometimes go *in the opposite direction*. That is, it appears that our moral judgments can sometimes impact our causal judgments [441]

One of the authors’ aims was to address the concern that previous tests on causal judgments did not distinguish between assessments of *typicality* and assessments of *morality*. Thus, they altered one of the original cases to rule out typicality as a driving force for the causal intuitions. The goal was to find out whether a ‘small difference in perceived moral status can—all by itself, with no help from typicality judgments—have any impact on people’s causal judgments’.¹³

¹¹See [Sprouse and Almeida 2012: 2] for a similar point about criticisms of data collection in linguistics. See also §3 below.

¹²We emailed invitations to faculty and graduate students in departments on the Philosophical Gourmet Report’s list of departments in the English-speaking world with web-pages containing readily accessible email addresses. Slightly more than half of the invitations were sent to philosophy faculty, and the rest to philosophy graduate students. We proceeded down the ranked list of departments until 200 people reached the last question on the survey, at which point the survey closed. (Because some respondents did not reach the last question, some of our questions had more than 200 respondents.) A total of 1271 philosophers were sent invitations to the survey.

¹³Note that, if Knobe and Fraser are correct, the results of this study should be even more difficult for philosophers to predict than the results of the earlier studies, since it

We presented the philosophers in our study with the more discriminating case used by Knobe and Fraser, described as follows:

Suppose subjects are presented with the following case:

The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take the pens, but faculty members are supposed to buy their own.

The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist has repeatedly emailed them reminders that only administrative assistants are allowed to take the pens.

On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message... but she has a problem. There are no pens left on her desk.

All subjects are then asked how much they agree with each of the following statements:

1. Professor Smith caused the problem
2. The administrative assistant caused the problem

Subjects respond to each question by selecting a number on a scale ranging from -3 (no agreement) to 3 (full agreement).

Given the above description, philosophers were asked to guess whether there would be a significant¹⁴ difference between subjects' level of agreement with the two statements. Here is the wording of our question and the distribution of responses, among the 190 who did not opt out of the question due to possible influence by previous exposure to similar studies:

On average, agreement with 1 would be:

	total responses	% responses
<i>significantly greater than their agreement with 2.</i>	182	95.8%
not significantly different from their agreement with 2.	8	4.2%
significantly lower than their agreement with 2.	0	0%

discriminates the relevant intuition more finely. Indeed, the earlier case they discuss in detail is the same in every relevant respect except that it potentially runs together the possible effects of assessments of morality and typicality [Knobe 2006a: 68].

¹⁴Here, and for all subsequent questions in the survey, subjects were told to interpret 'significantly' in terms of statistical significance, where for significance, $p < .05$, two-tailed.

Here we have indicated the correct answer in italics. Nearly all philosophers in our study correctly predicted the result that the authors take to suggest ‘that moral judgments actually do play a direct role in the process by which causal judgments are generated’. (In the folk study, subjects’ average level of agreement with 1 was 2.2; and the average level of agreement with 2 was -1.2.)¹⁵

KNOBE ON INTENTIONALITY

The second question in our survey was taken from Joshua Knobe’s 2003 paper ‘Intentional Action and Side Effects in Ordinary Language’. The significance of the experiment is described in Knobe’s more recent [2006b]:

An outcome can be considered a ‘side-effect’ when (1) the agent was not specifically trying to bring it about but (2) the agent chose to do something that she foresaw would involve bringing it about. The question is: Will people think that the agent brought about such an outcome intentionally?

But when we study these cases systematically, we end up with a surprising result: people’s intuitions appear to be influenced by the *moral* qualities of the side-effect itself. Specifically, 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.

This is the key result of the experiment... where a vignette about environmental harm elicited very different intuitions from a quite similar vignette about environmental help. And the same effect arises for other cases that have the same basic structure. [p. 209]

Claims to the effect that the results of this study are striking or surprising abound in the experimental philosophy literature, including Knobe’s work.¹⁶ At the same time, the original paper introduces the results of the chairman case with the phrase ‘As predicted’ [Knobe 2003a]. For a proposal that makes sense of this apparent tension, see §3 below.

Most philosophers in our study likewise predicted the results. They were presented with Knobe’s original survey, as follows:

Suppose some subjects are presented the following case:

¹⁵ Knobe and Fraser [2008, fn. 4] report that 18 subjects participated in the study; the result was statistically significant as $p < .001$.

¹⁶For example the results are called ‘striking’ in [Nichols and Ulatowski 2007:346], [Adams and Steadman 2004b], [Rose et al. 2012], and [Guglielmo and Malle 2010], and called ‘surprising’ or prefaced with ‘surprisingly’ in Sverdlik [2004], Turner [2005], [Nadelhoffer 2006:205], [Feltz and Cokely, n.d.], and [Knobe 2006a, 2006b]. Indeed, they are often used as paradigms of the unexpected results obtained through surveys, as in the Feltz quote from our Introduction. In [Knobe 2004], they are apparently included as examples of ‘extremely surprising’ results in experimental philosophy.

Harm: The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.’

The chairman of the board answered, ‘I don’t care at all about harming the environment. I just want to make as much profit as I can. Let’s start the new program.’

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

These subjects are then asked to determine how much blame the chairman deserved for what he did (on a scale from 0 to 6) and to answer:

1. Did the chairman of the board *intentionally* harm the environment?

A different set of subjects are instead presented with the following case:

Help: The vice-president of a company went to the chairman of the board and said, ‘We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.’

The chairman of the board answered, ‘I don’t care at all about helping the environment. I just want to make as much profit as I can. Let’s start the new program.’

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

These subjects are then asked to determine how much praise the chairman deserved (on a scale from 0 to 6) and answer:

2. Did the chairman of the board *intentionally* help the environment?

Given the above description, philosophers were then given the opportunity to guess whether there would be a significant difference between the way the two groups answered the questions. Here is the wording of our question and the distribution of responses, among the 83 respondents who did not opt out of the question due to possible influence by previous exposure to similar studies:¹⁷

Subjects asked question 1 would respond ‘yes’

	total responses	% responses
<i>significantly more often than subjects asked question 2 would respond ‘yes’.</i>	69	83.1%

¹⁷The total number of respondents exposed to question 2 is low partly because it is well-known and many respondents opted out, but also because the first 42 surveys that were taken contained a mistake in the wording of that question: it did not correctly describe the wording presented to the folk subjects. After discovering the mistake, we rewrote the question and discarded all responses to the flawed version.

not significantly more or less often than subjects asked question 2 would respond 'yes'.	10	12%
significantly less often than subjects asked question 2 would respond 'yes'.	4	4.8%

The vast majority of philosophers would not have been surprised by the result that Knobe interprets as suggesting that people's intuitions about intentionality are 'influenced by the *moral* qualities of the side-effect itself'.¹⁸

LIVENGOOD AND MACHERY ON THE METAPHYSICS OF CAUSATION

The third section of our survey asked about two studies found in a 2007 paper by Jonathan Livengood and Edouard Machery, entitled 'The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence'.

Livengood and Machery set out to test folk intuitions about causation by absence. It is uncontroversial that the folk are sometimes disposed to cite absences as causes, and other times not. This fact has lead metaphysicians to offer explanations for what they take to be folk intuitions: on the one hand, theorists who posit widespread causation by absence must explain why the folk are often reluctant to cite absences as causes, and on the other hand, theorists who deny that absences can be causes must explain why the folk sometimes do cite absences as causes. Livengood and Machery focus on two explanations of this sort.

Helen Beebee argues that the relationist view of causation can be saved if philosophers interpret the folk tendency to speak of absences as causes as a confusion between causation and *causal explanation*. Livengood and Machery investigated this argument:

We thought Beebee's approach promising enough to be put to the test. We experimentally tested two claims. First, we tested the claim that the folk deny the causal status of some absences that count as genuine causes according to non-relationist theories of causation. Second, we tested the claim that the folk really fail to distinguish causation from causal explanation. The first prediction was upheld, but, surprisingly, the second was not, suggesting that pace Beebee, the folk do not conflate causation and causal explanation. We argue that our results stand as a challenge to both Lewis and Beebee.

¹⁸Knobe reports that in the original study, more people answered 'yes' when answering the first question (83%), while more people answered 'no' when answering the second question (77%) [2003: 192].

As the title of their paper and the following quote indicate, they take the results of their studies to be different from what ‘the often peculiar intuitions of metaphysicians’ would have indicated beforehand [108]:

We should also highlight a striking, unexpected feature of our results. Even when the absence was made clearly explanatorily relevant, as it is in the unsafe rope case, people were not disposed to assent to a sentence causally explaining an event by reference to this absence. [121]

We do not agree that this is a problem for Beebee,¹⁹ but for present purposes the relevant question is how striking and unexpected Livengood and Machery’s result would be to philosophers in general. To test this question, we first presented those in our study with the following description:

Suppose subjects are presented with the following case:

The broken rope case. Susan had to climb a rope in gym class. Susan was a very good climber. She started climbing, but the rope broke before she reached the rafters. She fell on the ground.

Subjects are asked one of two questions. Some subjects are asked:

On a scale of 1 to 7, 1 indicating that you totally disagree and 7 indicating that you totally agree, how much do you agree with the following claim?

1. ‘The rope breaking caused Susan to fall.’

Other subjects are asked:

On a scale of 1 to 7, 1 indicating that you totally disagree and 7 indicating that you totally agree, how much do you agree with the following claim?

2. ‘Susan fell because the rope broke.’

We then asked philosophers to respond to this question:

On average, agreement with statement 1 would be:

significantly greater than agreement with statement 2.

not significantly different from agreement with 2.

¹⁹Beebee’s goal is to explain some cases where the folk *do* cite absences as causes; she is not committed to the claim that any time the folk find an absence explanatorily relevant, they will be willing to cite it as a cause. Causal explanatory relevance is an intricate and context-dependent phenomenon, and arguably the rope’s breaking is more explanatorily relevant in the context of a discussion of the Broken Rope Case than the rope’s not breaking in the context of a discussion of the Unsafe Rope Case. In addition, it may well be that the folk *are* willing to cite the rope’s not breaking as *a* cause; Livengood and Machery simply do not test whether subjects would tend to agree with ‘The rope’s not breaking was *a* cause of Susan’s reaching the rafters’. Plausibly, on a view like Beebee’s, the degree of causal explanatory relevance required for assent to ‘A caused B to V’ (or ‘A is *the* cause of B’s V-ing’...) is higher than that required for assent to questions about whether A was *a* cause of B’s V-ing.

significantly lower than agreement with **2**.

Next, we asked about a second case, described as follows:

Suppose subjects (a different group from those asked about the broken rope case) are presented with the following case:

The unsafe rope case. Susan has to climb an old, worn-out rope in gym class. She wondered if it would support her weight. Susan was a very good climber. Though nervous, she climbed all the way to the rafters.

Subjects are asked one of two questions. Some subjects are asked:

On a scale of 1 to 7, 1 indicating that you totally disagree and 7 indicating that you totally agree, how much do you agree with the following claim?

1. ‘The rope not breaking caused Susan to reach the rafters.’

Other subjects are asked:

On a scale of 1 to 7, 1 indicating that you totally disagree and 7 indicating that you totally agree, how much do you agree with the following claim?

2. ‘Susan reached the rafters because the rope did not break.’

Philosophers were again asked to respond to this question:

On average, agreement with statement **1** would be:

significantly greater than agreement with statement **2**.

not significantly different from agreement with **2**.

significantly lower than agreement with **2**.

Livengood and Machery report that in their study there was no significant difference in folk agreement with the two sentences in the Broken Rope Case.²⁰ However, there was a significant difference in folk agreement with the two sentences in the Unsafe Rope Case; folk-subjects’ agreement with sentence **1** (‘The rope not breaking caused Susan to reach the rafters’) was significantly lower than agreement with sentence **2** (‘Susan reached the rafters because the rope did not break’).²¹

Both results were predicted by philosophers in our survey. In the Broken Rope Case, 202 respondents in our survey reported that they could respond

²⁰The average response to **1** in the Broken Rope Case was 5.77, and the average response to **2** was 5.47.

²¹The average response to **1** was 3.06, and the average response to **2** was 4.00.

without bias, and their responses were distributed as follows (correct answer in italics):

	total responses	% responses
significantly greater than agreement with statement 2.	9	4.5%
<i>not significantly different than agreement with statement 2.</i>	158	78.2%
significantly lower than agreement with statement 2.	35	17.3%

In the Unsafe Rope Case, 198 respondents in our survey reported that they could respond without bias,²² and their responses were distributed as follows (correct answer in italics):

	total responses	% responses
significantly greater than agreement with statement 2.	5	2.5%
not significantly different than agreement with statement 2.	22	11.1%
<i>significantly lower than agreement with statement 2.</i>	171	86.4%

Livengood and Machery conclude by exhorting us to leave the comfort of our armchairs because ‘the folk probably don’t think what you think they think’ [126]. Our results suggest that they overstepped in drawing this conclusion from their folk surveys alone.

NICHOLS AND KNOBE ON MORAL RESPONSIBILITY AND DETERMINISM

The last section of our survey involves a study reported in the 2007 paper ‘Moral Responsibility and Determinism: the Cognitive Science of Folk Intuitions’ by Shaun Nichols and Joshua Knobe. Nichols and Knobe gloss the results as follows:

These initial experiments replicated the finding (originally due to Nahmias et al.) that people have compatibilist intuitions when presented with vignettes that trigger affective responses. But they also yielded a new and surprising result. When subjects were presented with an abstract vignette, they had predominantly *incompatibilist* intuitions. [671]

To test whether this result was surprising, we presented philosophers with these cases, as follows.

²²³ respondents who answered the first rope question skipped the second one entirely. 1 respondent chose the ‘opt out’ option for the second but not the first.

Suppose subjects are presented with descriptions of two different ‘universes’, A and B:

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

Now imagine a universe (Universe B) in which *almost* everything that happens is completely caused by whatever happened before it. The one exception is human decision making. For example, one day Mary decided to have French Fries at lunch. Since a person’s decision in this universe is not completely caused by what happened before it, even if everything in the universe was exactly the same up until Mary made her decision, it *did not have to happen* that Mary would decide to have French Fries. She could have decided to have something different.

The key difference, then, is that in Universe A every decision is completely caused by what happened before the decision – given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does not have to happen* the way that it does.

Some subjects are presented with the question:

1. In Universe A, is it possible for a person to be fully morally responsible for their actions?

Philosophers in our study were then asked to respond to the question:

Subjects would answer ‘yes’

significantly more often than they would answer ‘no’

not significantly more or less often than they would answer ‘no’

significantly less often than they would answer ‘no’

Our meta-survey then continued:

Other subjects, who are presented with the same description of the two universes, are not asked question 1, but are instead presented with the following case:

In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 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.

These subjects are then asked the following question:

2. Is Bill fully morally responsible for killing his wife and children?

Philosophers were once again asked whether subjects would answer ‘yes’ (that Bill is morally responsible), and given the same three options as before.

In the original study, 86% of respondents answered ‘no’ to the first question (i.e., said that one cannot be morally responsible in Universe A), while 72% answered ‘yes’ to the second question (i.e., said that Bill is morally responsible for killing his wife and children in Universe A). Once again, our respondents predicted this contrast. For the first question, 163 respondents in our survey reported that they could respond without bias, and their responses were distributed as follows (correct answer in italics). Subjects would answer ‘yes’:

	total responses	% responses
significantly more often than they would answer ‘no’.	15	9.2%
not significantly more or less often than they would answer ‘no’.	22	13.5%
<i>significantly less often than they would answer ‘no’.</i>	126	77.3%

For the second question, 160 respondents in our survey reported that they could respond without bias, and their responses were distributed as follows (correct answer in italics). Subjects would answer ‘yes’:

	total responses	% responses
<i>significantly more often than they would answer ‘no’.</i>	133	83.1%
not significantly more or less often than they would answer ‘no’.	19	11.9%
significantly less often than they would answer ‘no’.	8	5%

More importantly, out of those philosophers who responded to both questions, 85.5% correctly expected that the first group of subjects would be more inclined to give the incompatibilist answer than the second. (This includes those who gave the middle answer to one question but the correct answer to the other.) In short, it does not appear that the contrast reported by Knobe and Nichols would in fact be surprising to most philosophers.

2. Objections

Let us turn to two possible objections about the design of our study.

OBJECTION 1: EXPECTING THE UNEXPECTED?

One worry is that our respondents may have treated the fact that they were being presented with questions about these surveys as evidence that the results were interesting or unexpected. In particular, they might have taken it as evidence that the surveys were published, and reasoned that surveys are more likely to be published if they have surprising results. This kind of subject-expectancy effect may have lead respondents to predict the very answers that they would have otherwise found surprising. (We will set aside the question whether similar worries apply to survey results in much first-order experimental philosophy.)²³

We cannot rule out that such an effect was present, though we doubt it played a very significant role. First, we did our best to avoid communicating that the folk surveys had already been conducted. Each question was put in hypothetical form: respondents were asked to *suppose* that ordinary non-philosophers are presented with a certain case, and then asked to say how they thought such people *would* respond. Moreover, the wording of the instructions carried an implicature to the effect that some of the surveys had not been carried out.²⁴ And we deliberately omitted any further information about the ‘hypothetical’ subjects at issue— such as the number of subjects in each study— precisely to avoid giving our respondents reason to think the studies had already been conducted.

Second, the hypothesized effect would not explain the highly accurate responses to our meta-survey. Every question had three possible answers: in each case there was one answer that the original experimenters appear to have thought would be the *predictable* answer, leaving two possible *surprising* answers— but in our study there is little or no pull towards the incorrect surprising answer. (Note that in Example 4 we asked about each condition individually, so various combinations of response were possible, more than one of which would presumably have counted as surprising.) In addition, recall the Broken Rope case, which was included in the folk survey simply as a control to make sure that

²³It is well-known that survey respondents do not simply focus on the content of the question but try to work out what kind of information the experimenter is trying to elicit. (See for example [Groves et al. 2009:228-9.]) Asked a series of questions about (say) their intuitions concerning intentionality, they may expect there to be something ‘tricky’ about the question and be more inclined to provide a tricky or interesting response rather than an apparently boring or obvious one. This kind of effect could theoretically produce artificially surprising results.

²⁴The instructions contained the following language: ‘Some of the cases you are asked about may be taken from the existing experimental philosophy literature. Questions are worded in a way that does not give away whether the relevant study has been conducted or not. (Any necessary citations, as well as the actual results of the relevant studies, will appear after you complete the survey.)’

the ‘striking’ differences in causal judgments in the Unsafe Rope case did not stem from a grammatical feature of the question.²⁵ The Broken Rope case indicated that they did not: there was no analogous ‘striking’ result in that case. For this reason, the Broken Rope case serves as a kind of control for our meta-study. If our respondents were influenced by the survey setting to select a result they would find surprising, they would not have been highly accurate even in their response to that question.²⁶

OBJECTION 2: NOT ENOUGH INFORMATION?

The previous worry was that our respondents had too much information communicated to them simply by being presented with our survey. But others might worry that our respondents weren’t given *enough* information. We omitted information about the number of participants in each study, as well as any demographic facts that had been reported. (As noted above, we did this to help avoid the worry raised by Objection 1). But the lack of specific information regarding the original study forces respondents to make guesses about these aspects of the original studies.²⁷ In particular, the folk surveys were quite small but still showed statistical significance: to take two extreme examples, the Knobe and Fraser case was put to 18 subjects, and the first question in the Nichols and Knobe study was put to 21 subjects. Now, suppose a philosopher presented with our study were to wrongly assume that the number of subjects in these studies was very high. Then, even if she wrongly thought there was only a very slight difference in the average levels of agreement between the two conditions, she might still arrive at the correct conclusion that the study would show a *statistically significant* difference. In this way, incorrect assumptions about the number of participants might exaggerate one’s predictions about the statistical significance of folk responses.

To control for the possibility of this kind of mistake, we conducted an additional survey, which was completed by 100 philosophers. The questions differed only in that our description of each folk survey was accompanied by the number of subjects rounded to the nearest 10, as well as whatever demographic information had been provided about the folk subjects. (For the most part they were American undergraduates.) Moreover, respondents were told at the outset

²⁵The concern was that subjects might find causal claims involving a gerundive construction along with the verb ‘cause’ to be more strange than two clauses connected with ‘because’, and as a result tend to agree with the second more often. See [Livengood and Machery: 117-8].

²⁶Note also that in our study the middle alternative (the ‘no statistical significance’ answer) was nearly always an incorrect answer, and there is some reason to think that when subjects have low credences about a survey question, they gravitate towards the middle alternative. See [Schuman and Presser 1981: ch. 6].

²⁷Thanks to David Chalmers and Kenny Easwaran for raising this concern.

that if a given study had not been conducted, we would use an arbitrarily chosen number of subjects and fake demographic information. We conjectured that if the respondents to our first survey had been making the incorrect assumptions described in Objection 2, they would have been more likely to predict statistically significant results than the respondents to the second survey. (This is because slight differences in levels of agreement would be less likely to yield statistical significance in studies with small survey populations.) But this was not the case: there was no significant decrease in the choice of the ‘no statistical difference’ response.²⁸

3. Discussion

QUALIFICATIONS

We wish to begin by stating what we take our study *not* to show. It clearly does not show the irrelevance of empirical research to philosophical theorizing. In fact, we enthusiastically endorse much of what Knobe and Nichols write on behalf of experimental philosophy in their ‘Manifesto’; for example, that philosophy is well-served by empirical investigations into the psychological processes governing human intuitions about central philosophical issues [Knobe and Nichols 2007:5]. Our results do not in any way subvert this guiding idea, for several reasons.

Most obviously, our study concerns only *folk surveys*, and in particular surveys in which the subjects were non-philosophers from a socioeconomic background similar to that of most members of the target philosophical audience. But there are many other kinds of empirical data that may be relevant to philosophy, from cross-cultural surveys to brain scans and skin-conductance tests.²⁹ We have no reason to think that philosophers could predict the results of any of these other methods of investigation.

²⁸The percentages of respondents selecting ‘no statistical difference’ in the first study, in order of questions asked, was 4.2, 12, 78.2, 11.1, 13.5, 11.9 (mean 21.82, median 11.95); in the second study they were 7.2, 15.4, 72.2, 15.5, 9.2, 10.6 (mean 21.68, median 13). There was, however, a small uptick in the number of respondents predicting statistical significance in the *wrong* direction (for the first survey the mean was 7.2, while for the second it was 10.7). Perhaps including details about the study populations induced the kind of subject-expectancy effect discussed in Objection 1. (This would vindicate our original decision to omit such details.) The only other difference between the studies was that, in soliciting respondents for the second study, we continued down the PGR where we had left off with the first study (see fn. 13), but sent emails only to faculty. (We also included top M.A. programs.)

²⁹See e.g. [Greene et al. 2001; Sripada and Konrath 2011].

Even surveys on American undergraduates may still have their uses.³⁰ First, philosophers who predicted a statistically significant effect may not have been very accurate in predicting the *extent* of that effect. Moreover, as prevailing theories about the kinds of processes that underlie philosophical intuitions become more fine-grained, adjudicating between theories may require very subtle studies whose results could not be predicted from the armchair.

We therefore do not draw any sweeping conclusions from our study about the relevance of experiments to philosophy. Our lesson is quite narrow: folk surveys can provide evidence against philosophers' ability to access ordinary intuitions only under the assumption that their results are not predictable by philosophers. And that assumption is far from innocent.

ON SURPRISINGNESS

Why, then, is it so common to claim that survey results are surprising? It cannot simply be because the folk report denying some consequence of a popular philosophical view. Philosophers employing a method of reflective equilibrium are well aware that their conclusions are often at odds with things the folk tend to say under certain conditions. And sometimes minimal revisionism should not count against a theory at all, as when, taken collectively, folk intuitions on the matter are incoherent. (Few would try to vindicate both of the apparently contradictory folk intuitions about whether someone can be morally responsible in Universe A.)³¹ A philosophical puzzle can often be stated as an inconsistent set of sentences, each of which the folk will find independently plausible under some conditions. In such a case every consistent theory will be subject to Alexander and Weinberg's criticism that it 'stray[s] from what intuitions people actually have' [2007: 73].

Here is a more plausible source for the surprisingness claim. Sometimes philosophers expound on what they take to be the 'intuitive' view, using claims of the form 'intuitively, P', 'we would be inclined to say that P', 'it would be counter-intuitive to conclude that not-P.'³² Can't we then assume that such philosophers would be surprised by a set of folk responses that apparently conflict with what they call the 'intuitive' view? Not if these philosophers would reject the relevant surveys as adequate tests for the intuitions at issue. For example, as Antti Kauppinen has argued, such philosophers may mean that P is

³⁰Though it may not be in their best interest for those who pursue such studies to endorse Knobe and Nichols' claim that 'the real measure of a research program depends on whether the program generates exciting new discoveries' [2007:14].

³¹We can imagine views about ethical language that would vindicate both judgments, but they would surely also generate consequences that the folk would be inclined to deny in various conditions.

³²See [Nadelhoffer and Nahmias 2007: 125; Livengood & Machery 2007: 107-8; and Nahmias et al. 2006: 29].

intuitive upon careful reflection, having avoided various conceptual pitfalls and pragmatic distractions [2007].³³

This point applies even in cases where experimental results seem to conflict with explicit claims about what is intuitive by ‘ordinary’ or ‘pretheoretic’ lights. After all, a member of the folk can reflect carefully and avoid conceptual pitfalls without thereby becoming a philosopher and therefore non-ordinary. As noted earlier, for example, non-philosophers who were more reflective or more aware of their own ability to make mistakes were significantly less likely than others to display the ‘Knobe effect’ (see [Pinillos et al. 2011]). For these reasons, a philosopher may think that P is the pretheoretically intuitive view, while being in a position to predict that, when asked a particular question intended to elicit an opinion as to whether P (in the setting of a quick survey), most ordinary folk would respond that not-P.

Note that being in a position to predict the results of a study does not entirely eliminate the element of surprise. A philosopher might, upon reflecting on the kind of case described in a folk survey, find *her own* reaction surprising. She may find unexpected conceptual connections, or discover that a concept influences certain judgments in a surprising way. And if she is not unusual in her deployment of that concept— or if she is good at simulating folk reactions— she may then correctly predict the relevant folk results. For example, in the chairman case, a philosopher may be surprised to find herself more drawn to an ascription of intentionality in the **harm** vignette than in the **help** vignette. She will probably be more inclined than ordinary folk to *resist* the Knobe effect—because of the ‘Pinillos-et-al effect’—but this does not preclude *noticing* by introspection that she herself is affected by it. And indeed, Knobe initially reported the results of his ‘chairman study’ with the phrase ‘as predicted’ [Knobe 2003a: 192]. Likewise, a variation on Knobe and Frazer’s pen case can be found in [Knobe 2006a], where he simply reports on his own reaction to the case, apparently anticipating the result later discovered empirically.³⁴ This suggests that even for the authors, these studies were only surprising in the second way— that is, from the armchair.

In short, perhaps the source of surprise in these studies has simply been mislocated. Perhaps the *initial* stage of each experiment— itself a piece of *a priori* analysis—is sufficient to reveal an unexpected feature of our conceptual

³³Another possibility: what the philosopher means by ‘S’ in ‘Intuitively, S’ may be hard to get across in a quick survey by using that English sentence. As Sosa puts it: ‘The experimental results really concern in the first instance only people’s responses to certain words. But verbal disagreement need not reveal any substantive, real disagreement, if ambiguity and context might account for the verbal divergence’ [2007:102].

³⁴The case involves a computer that crashes when used by more than one person. Knobe simply writes: ‘Here we would attribute the crash more to [Jane]’s behavior than to [Lauren]’s behavior. But why?’ [68].

apparatus. But *this* kind of surprise hardly provides evidence that armchair reflection is radically inadequate to access ordinary intuitions.

THE METHODOLOGICAL UPSHOT

We have chosen to sidestep the recent debate about the proper place of ordinary intuitions in philosophy. Some armchair philosophers make claims about ordinary intuitions in their arguments, others think ordinary intuitions should be superseded by tutored intuitions, and still others think it is a methodological confusion to make claims about intuitions about P rather than making claims about P.³⁵ Only philosophers of the first stripe are likely to hold that appeals to ordinary intuition are a crucial component of philosophical methodology. But we will grant that assumption for the purposes of this paper.

As discussed under ‘Qualifications’ above, we also allow that surveys can be useful for philosophy. Even so, it is far from obvious how *pervasive* a role should be played by surveys of primarily middle-class English speakers,³⁶ which make up the bulk of experimental philosophy to date.³⁷ A philosophical book might contain dozens of appeals to ordinary intuition, and the difficulty that would be involved in conducting careful experiments for each of these is enormous. Presumably we must make a trade-off between a higher risk of error on the one hand, and a drain of resources on the other. Frank Jackson takes an inclusive approach to this trade-off:

I am sometimes asked... why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious public opinion polls on people's responses to various cases? My answer is that I do — when it is necessary. [1998:36]

Meanwhile, Stich and Weinberg reject Jackson's ‘singularly implausible assumption’ that ‘when consulting intuitions about possible cases of philosophical interest, [he] and other philosophers can often simply rely on their own intuitions’ [2001: 640].³⁸ We think this kind of dismissal is premature

³⁵For some critiques of appeals to intuition in philosophy see [Cappelen 2012; Williamson 2004].

³⁶Again, we have no reason to think philosophers could predict the results of cross-cultural surveys, not to mention brain scans, skin conductance tests, etc.

³⁷There is a related question whether reporting the results of such a survey should constitute the primary content of a published philosophy paper, given the predictability of the results we studied.

³⁸They also reject Jackson's method of running cases by ‘students in the classroom’, since students are not a representative sample of the ‘folk’. This point, of course, applies equally to experimental philosophers who run surveys entirely on undergraduates in American universities, and though it is usually not reported how those undergraduates were selected, the sample size is often about that of a classroom.

given the lack of empirical reasons to doubt philosophers' reliability in predicting ordinary intuitions.

(Testing whether *philosophers'* intuitions differ from those of the folk is *not* an adequate substitute for a study testing whether philosophers can *predict* folk intuitions. A theorist may have different intuitions from the folk without being *ignorant* of folk intuitions. For example, with some effort a linguist may be able to parse a sentence like 'The rat the cat the dog chased ate died' and judge it to be grammatical, while predicting that the folk will judge it to be ungrammatical.³⁹ This makes sense if her best theory treats the folk judgment as a kind of performance error.)

To assess the proper frequency of surveys in philosophy, it would be useful to know how accurate philosophers are when it comes to the *average* philosophically-relevant folk intuition. It is hard to see how this would even be tested, but given that the surveys in our study were selected for their alleged surprisingness, it is reasonable to think philosophers would perform even better on average, *in vivo*. One experiment worth undertaking would test how accurate philosophers are when they *publish* claims about ordinary intuitions. (As discussed above, unadorned claims about what is 'intuitive' do not necessarily fit this criterion.) Such a study would comb through a philosophical journal looking for such claims and run a formal survey for each one.

Notably, the analog of this study for *linguistics* has already been performed [Sprouse et al., n.d.]. Like philosophy, linguistics has recently seen increased criticism about traditional methods. After all, syntacticians and semanticists often simply draw on their own intuitions about the acceptability of sentences, and perhaps those of some friends and colleagues, rather than running formal experiments. But insofar as the goal is to obtain intuitions matching those of ordinary speakers, there are concerns about the reliability of this method.⁴⁰

Sprouse, Schütze, and Almeida collected a random sample of nearly 300 sentence types from articles in ten years of *Linguistic Inquiry*. (The sentences were only sourced from articles in which acceptability judgments were based on informal methods.) The authors tested all of these sentences to see if the acceptability judgments assigned to them informally comported with judgments obtained experimentally from the folk. They found that 95% of the linguistic phenomena identified informally were replicated under experimental conditions: 'This means that even under the (likely unwarranted) assumption that all of the discrepant results are false positives that have found their way into the syntactic literature due to the shortcomings of traditional methods, the maximum proportion of such false positives in LI 2001-2010 is 5%' [Sprouse et al., n.d.]. A

³⁹The example is from [Miller and Chomsky 1963:286]. See also [Hudson 1996].

⁴⁰See [Sprouse and Almeida 2012: 2] and the citations therein.

similar study performed on a standard syntax textbook had a replication rate of 98% [Sprouse and Almeida, 2012]. In short, the authors find ‘no evidence of a reliability problem for acceptability judgment data in syntax’ [Sprouse et al., n.d.]. At the same time, they don’t deny that some syntactic questions should be investigated with the help of formal experiments:

We suggest that syntacticians abandon the idea that there is a single method for every research question (or research environment).... Syntacticians need to evaluate each methodology based on its costs and benefits to decide which method is most appropriate for their specific research question. [*ibid.*]

This multi-pronged approach to linguistics fits well with Jackson’s view that formal experiments are sometimes—but not always—necessary. Of course, in the absence of studies like that of Sprouse et al., it remains an open question whether philosophers are as reliable in predicting ordinary intuitions as linguists appear to be. But in the meantime it would be unfounded to insist that ‘only the results of [empirical] research can deliver the intuitions that can serve as evidential basis for or against philosophical claims’—even assuming that philosophical claims must be based on ordinary intuitions.⁴¹

4. Conclusion

We examined four survey results presented as surprising in the experimental philosophy literature, and found that all were predicted by a large majority of philosophers. Results deemed surprising by experimental philosophers should not be treated uncritically as evidence against the reliability of informal access to ordinary intuitions.

University of Michigan

REFERENCES

- Alexander, Joshua and Jonathan M. Weinberg 2007. Analytic Epistemology and Experimental Philosophy, *Philosophy Compass* 2/1: 56–80.
- Alicke, M.D. 2008. Blaming Badly, *Journal of Cognition and Culture* 8/1-2: 179-186.

⁴¹The claim in quotation marks is ‘widely endorsed by experimental philosophers’, according to Alexander and Weinberg [2007: 61].

- Adams, F., and Steadman, A. 2004a. Intentional Action in Ordinary Language: Core Concept or Pragmatic Understanding? *Analysis* 64:173– 81.
- _____. 2004b. Intentional Actions and Moral Considerations: Still Pragmatic, *Analysis* 64: 264– 6.
- Beebee, Helen 2004. Causing and Nothingness, in *Causation and Counterfactuals*, ed. L. A. Paul, E. J. Hall, and J. Collins: 291–308. Cambridge, MA: MIT Press.
- Cappelen, Herman 2012. *Philosophy without Intuitions*, Oxford: Oxford University Press.
- Feltz, A. 2009. Experimental Philosophy, *Analyse & Kritik* 31: 201-219.
- Feltz, A. and E. Cokely, unpublished. The Terror of ‘Terrorists’: An Investigation in Experimental Applied Ethics.
- Goldman, Alvin 2001. Replies to the Contributors, *Philosophical Topics* 29: 461–511.
- Greene, Joshua, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley and Jonathan D. Cohen 2001. An fMRI Investigation of Emotional Engagement in Moral Judgment, *Science* 293/5537: 2105-2108.
- Groves, Robert, F. Fowler, M. Couper, J. Lepkowski, E. Singer, R. Tourangeau 2009. *Survey Methodology*, 2nd Edition. Hoboken: Wiley.
- Guglielmo, S., & B. Malle 2010. Can Unintended Side Effects be Intentional? Resolving a Controversy Over Intentionality and Morality, *Personality and Social Psychology Bulletin* 36: 1635–1647.
- Harman, Gilbert 1976. Practical Reasoning, *Review of Metaphysics* 29, 431-63.
- Hudson, R. 1996. The Difficulty of (So-called) Self-embedded Structures, *UCL Working Papers in Linguistics* 8.
- Jackson, Frank 1998. *From Metaphysics to Ethics: A Defense of Conceptual Analysis*, Oxford University Press.
- _____. 2001. Responses, *Philosophy and Phenomenological Research* 62.3: 653 – 64.
- Kauppinen, Antti 2007. The Rise and Fall of Experimental Philosophy, *Philosophical Explorations* 10/2: 95-118.
- Knobe, Joshua 2003a. Intentional Action and Side-Effects in Ordinary Language, *Analysis* 63/3: 190-4.
- _____. 2003b. Intentional Action in Folk Psychology: An Experimental Investigation, *Philosophical Psychology* 16/2: 309–23.
- _____. 2004. What is Experimental Philosophy?, *The Philosophers’ Magazine* 28:37-39.

- _____. 2006a. *Folk Psychology, Folk Morality*, unpublished PhD dissertation, Princeton University.
- _____. 2006b. The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology, *Philosophical Studies* 130: 203-231.
- Knobe, Joshua and Ben Fraser 2008. Causal Judgment and Moral Judgment: Two Experiments, in *Moral Psychology*, ed. Walter Sinnott-Armstrong: 441-448.
- Knobe, Joshua and Shaun Nichols 2007. An Experimental Philosophy Manifesto, in Joshua Knobe and Shaun Nichols (eds.), *Experimental Philosophy*. Oxford University Press.
- Livengood, Jonathan and Edouard Machery 2007. The Folk Probably Don't Think What You Think They Think: Experiments on Causation by Absence, *Midwest Studies in Philosophy* XXXI/1: 107-127.
- McCann, H. 2005. Intentional Action and Intending: Recent Empirical Studies, *Philosophical Psychology* 18/6: 737- 48.
- Mele, Alfred 2001. Acting Intentionally: Probing Folk Notions, in B. Malle, L. Moses, and D. Baldwin, eds. *Intentions and Intentionality: Foundations of Social Cognition*. Cambridge, MA: MIT/Bradford.
- _____. 2003. Intentional Action: Controversies, Data, and Core Hypotheses, *Philosophical Psychology* 16: 325- 40.
- Miller, George and N. Chomsky 1963. Finitary Models of Language Users, in Luce, R.; Bush, R. and Galanter, E. (eds.) *Handbook of Mathematical Psychology*, Vol 2. New York: Wiley: 419-93.
- Nadelhoffer, Thomas 2004. The Butler Problem Revisited, *Analysis* 64/3: 277-84.
- _____. 2006. Bad Acts, Blameworthy Agents, and Intentional Actions: Some Problems for Juror Impartiality, *Philosophical Explorations* 9/2, 203-219.
- Nadelhoffer, Thomas and Nahmias, E. 2007. The Past and Future of Experimental Philosophy, *Philosophical Explorations* 10/2: 123-149.
- Nahmias, Eddy 2006. Folk Fears about Freedom and Responsibility: Determinism vs. Reductionism, *Journal of Cognition and Culture* 6/1-2: 215-37.
- Nahmias, Eddy, S. Morris, T. Nadelhoffer and J. Turner 2005. Surveying Freedom, *Philosophical Psychology* 18/5: 561-584.
- Nahmias, Eddy, Stephen Morris, Thomas Nadelhoffer, and Jason Turner 2006. Is Incompatibilism Intuitive? *Philosophy and Phenomenological Research* 73/1: 28-53.

- Nichols, Shaun 2006. Folk Intuitions on Free Will, *Journal of Cognition and Culture* 6/1-2: 57-86.
- Nichols, Shaun and Joshua Knobe 2007. Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions, *Nous* 41/4: 663-685.
- Nichols, Shaun and J. Ulatowski 2007. Intuitions and Individual Differences: The Knobe Effect Revisited, *Mind and Language* 22/4: 346-365.
- Pinillos, N. Ángel, N. Smith, G. Shyam Nair, P. Marchetto, C. Mun 2011. Philosophy's New Challenge: Experiments and Intentional Action, *Mind and Language* 26/1: 115-139.
- Rose, D., J. Livengood, J. Sytsma and E. Machery, 2012. Deep Trouble for the Deep Self, *Philosophical Psychology* 25:5: 629-646.
- Schuman, Howard and Stanley Presser 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
- Sosa, Ernest 2007. Experimental Philosophy and Philosophical Intuition, *Philosophical Studies* 132: 99-107.
- Sripada, Chandra and S. Konrath 2011. Telling More Than We Can Know about Intentional Action, *Mind and Language* 26/3: 353-380.
- Sprouse, Jon and Diogo Almeida 2012. Assessing the Reliability of Textbook Data in Syntax: Adger's *Core Syntax*, *Journal of Linguistics*: 1-44.
- Sprouse, Jon, Carson Schütze and Diogo Almeida *m.s.*. Assessing the Reliability of Journal Data in Syntax: Linguistic Inquiry 2001-2010.
- Stich, Stephen and J. Weinberg 2001. Jackson's Empirical Assumptions, *Philosophy and Phenomenological Research* 62/3: 637-643.
- Sverdlik, S. 2004. Intentionality and Moral Judgments in Commonsense Thought About Action, *Journal of Theoretical and Philosophical Psychology* 24/2: 224-236.
- Turner, Jason 2005. Folk Intuitions, Asymmetry, and Intentional Side Effects, *Journal of Theoretical & Philosophical Psychology* 24/2: 214-219.
- Williamson, Timothy 2004. Philosophical 'Intuitions' and Scepticism about Judgments, *Dialectica* 58/1: 109-55.