Rev.Phil.Psych.
DOI 10.1007/s13164-012-0100-9

Experimenter Philosophy: the Problem of Experimenter Bias in Experimental Philosophy

Brent Strickland · Aysu Suben

Abstract It has long been known that scientists have a tendency to conduct experiments in a way that brings about the expected outcome. Here, we provide the first direct demonstration of this type of experimenter bias in experimental philosophy. Opposed to previously discovered types of experimenter bias mediated by face-to-face interactions between experimenters and participants, here we show that experimenters also have a tendency to create stimuli in a way that brings about expected outcomes. We randomly assigned undergraduate experimenters to receive two different hypotheses about folk intuitions of consciousness, and then asked them to design experiments based on their hypothesis. Specifically, experimenters generated sentences ascribing intentional and phenomenal mental states to groups, which were later rated by online participants for naturalness. We found a significant interaction between experimenter hypothesis and participant ratings indicating a general tendency for experimenters to obtain the result that they expected. These results indicate that experimenter bias is a real problem in experimental philosophy since the methods and design employed here mirror the predominant survey methods of the field as a whole. The bearing of the current results on Knobe and Prinz's (Phenomenology and Cognitive Science 7(1):67–83, 2008) group mind hypothesis is discussed, and new methods for avoiding experimenter bias are proposed.

Experimenter bias refers to experimenters' (often unconscious) disposition to obtain an empirical result that they are looking for, even when that result does not necessarily reflect the truth. The empirical work from the experimental philosophy community on consciousness and theory-of-mind (as well as many other topics) could potentially be undermined by experimenter bias effects that the community itself is unaware of.

Despite its apparent danger, experimenter bias exists in the sciences very generally (e.g. medicine: Ioannidis 2005) and more specifically in the domain of psychology (Rosenthal and Rubin 1978; Doyen et al. 2012). It has typically been studied by

B. Strickland (✉) · A. Suben
Department of Psychology, Yale University, Box 208205, New Haven, CT 06520-8205, USA
e-mail: brent.strickland@yale.edu

Published online: 21 August 2012

examining the ways in which experimenters physically or socially interact with their participants in order to induce an expected outcome. For example, when teachers expect a certain group of students to learn more over the course of a school year, they interact with those students in a way that brings about this outcome (Rosenthal and Rubin 1978; although see Barber and Silver 1968 for a critical perspective on this work).

Experimental philosophy is an emerging movement at the intersection of philosophy and psychology (Knobe et al. 2011) that is typically interested in folk intuitions about philosophically important topics like morality, causation, or the attribution of consciousness. One might think that experimental philosophy as a movement is unlikely to be strongly influenced by experimenter bias to the extent that it commonly employs questionnaire studies in which experimenters either do not directly interact with participants (in the case of online surveys) or have limited opportunities to influence the outcome of the study (in the case of text based surveys which are administered face-to-face).

However, this view is too simple. Instead, we suggest here that a different but equally dangerous type of experimenter bias lurks in the shadows of experimental philosophy: shaping one's stimuli in a way that brings about desired results. It is important that experimental philosophy as a movement be conscious of this danger and develop experimental protocols for dealing with this kind of bias unless they are willing to have their theoretical conclusions undermined. Below we provide a novel demonstration of experimenter bias using Knobe and Prinz's (2008) study examining the psychological mechanisms underlying the attribution of consciousness. While the ultimate conclusion of Knobe and Prinz's research may prove to be correct, we hope that the current demonstration of experimenter bias in this context can serve as a useful warning to researchers in the field as a whole.

1 Experimenter Bias in Psychology

The social and professional pressures in the sciences incentivize discovering positive, novel results. This incentive structure often has the desired effect of producing compelling and important research, but it also often has the unintended consequence of producing high rates of false positives (Ioannidis 2005).[1]

Many mechanisms can bring about these false positives. At one extreme is simple and straightforward academic fraud where researchers intentionally alter their data to obtain a desired result. In contrast to conscious cheating, there are also more subtle ways in which a research community with otherwise good intentions can unwittingly produce systematically unreliable results. These include "publication bias" (Dickersin 1990) which is the disposition of journals to prioritize publication of positive over null results and the related "file drawer" problem (Rosenthal 1979), which is the tendency for null results never to be submitted for publication in the first place.

Experimenter bias is slightly different in that experimenters obtain a result that they expect or desire (Rosenthal and Fode 1963). It differs from simple cheating in

[1] This is true across many domains in the natural sciences, but we will limit the discussion here to empirical psychology.

that the experimenters in question do not intend scientific misconduct. It differs from publication bias in that no findings are shelved away because they are undesirable for publication. Instead, experimenters subtly and often unconsciously manipulate certain aspects of how their experiment is conducted or analyzed in a way that makes the expected result more likely than it otherwise should be.

In one of the classic demonstrations of experimenter bias, Rosenthal and Fode (1963) told unwitting experimenters from a senior psychology course that they would train "maze-bright" or "maze-dull" rats to navigate a maze. Although the rats in the two groups did not actually differ significantly in their inherent attributes, the experimenters were told that the two groups of rats differed genetically in such a way that they should expect the maze-bright rats to effectively learn how to do the maze task, but that they should expect the maze-dull rats to show little to no learning on the task. After running what should have been the same experiment on virtually identical groups of rats, those experimenters in the maze-bright group found that their rats were more effective learners than those in the maze-dull group, thus obtaining the result that they expected.

Careful post-experiment debriefing revealed that experimenter bias may not have been operating at the level of conscious awareness. It was instead hypothesized that the expected outcome was induced by subtle and perhaps unconscious differences between the two groups in how experimenters handled and responded to their subjects. This type of experimenter bias resulting from subtle differences in face-to-face interaction has also been documented in other experimental contexts testing perception (Adair and Epstein 1968), pedagogy (Rosenthal and Jacobson 1968), and priming effects from social psychology (Doyen et al. 2012).

As mentioned in the introduction, one of the principle methods employed in experimental philosophy consists in collecting responses to text-based surveys (e.g. Cushman et al. 2006; Feltz et al. 2012; Young and Phillips 2011; Strickland et al. 2012; Knobe and Prinz 2008; see also Cullen 2010 and Scholl 2008 for interesting methodological critiques). Many, though not all, of these surveys are administered in on-line in contexts in which the experimenter has no chance to physically interact with the participant. In other cases, the survey is physically administered by the experimenter (e.g. Monroe and Malle 2010; Feltz et al. 2009). However, even in these face-to-face surveys, the stimulus and response are almost entirely text based, and the experiment requires little direct interaction between the experimenter and the participant. Given the popularity of survey methods like this, one could think that experimental philosophy would be largely immune to experimenter bias effects.

Contrary to this view however, we claim that there typically exists ample opportunity to mold and craft one's stimuli in such a way that the expected or desired outcome is more likely to be obtained than it should be. Fore example, imagine a typical experiment in which an experimenter is creating stimuli where they expect sentences from one group to be rated as more natural (or intentional/causal/morally desirable etc…) than sentences from a second group. Even if no true difference in the naturalness of the two groups of sentences exists, the experimenter could, for example, obtain the expected effect simply by pragmatically sabotaging the dispreferred group in subtle ways that may go undetected by readers, reviewers or even the experimenters themselves. Pragmatic sabotage is not the only method that exists, and different implicit strategies may apply in different contexts. The point is simply

that there is room for experimenters to craft their stimuli in a way that is likely to yield the expected result.

Of course, the above critique applies to many areas of psychology including cognitive, social, and developmental psychology. Similarly to these fields, experimental philosophy should also be aware of the dangers of experimenter bias in stimulus creation, especially given its frequent use of questionnaire studies.

Below we show that experimenter bias is not just a theoretical possibility, but that it actually does negatively influence the stimulus creation in experimental philosophy. It should thus be considered a major worry for the field.

1.1 Experiment: Experimenter Bias Effects in the Study of Folk Intuitions About Consciousness

To test for the possibility of such experimenter bias effects, we had undergraduate assistants design a version of a well-known experiment from experimental philosophy looking at folk intuitions about consciousness (described in more detail below). Half of the assistants were told to expect one outcome, and the other half were told to expect the exact opposite outcome. We then ran those experiments on Amazon's Mechanical Turk. Would experimenters who have received otherwise identical instructions, create stimuli biased on their favor?

The study we chose comes from Knobe and Prinz's (2008) recent paper in which they argued that people's naïve intuitions about mental states differ dramatically as a function of the type of mental state in question. On the one hand, there are *intentional* mental states (e.g. beliefs, desires and intentions) that are not associated with any particular feeling or conscious experience. On the other hand, there are *phenomenal* mental states (e.g. being in love, imagining something, feeling sad) that are associated with particular types of feelings or conscious experience.

Knobe and Prinz hypothesized that people's naïve intuitions about these mental states differ in that people (implicitly) believe that having a body is necessary for having phenomenal mental states but not intentional mental states. They successfully confirmed their hypothesis in a study in which on-line participants rated the naturalness of passages ascribing either intentional mental states to a group (e.g. "Acme corporation believes that its profit margin will soon increase.") or phenomenal mental states to the same group (e.g. "Acme corporation is now vividly imagining a purple square."). In line with their predictions, participants indeed rated the intentional sentences as being more natural sounding than phenomenal sentences.

We adopted this design as a test case for potential experimenter bias effects by having undergraduate assistants help us design an attempted replication of Knobe and Prinz's study. Some of the experimenters received the same hypothesis that Knobe and Prinz had in mind when they originally created their study, while others received the opposite hypothesis. M-turk participants then rated the stimuli that were created by our experimenters. If experimenter bias effects are indeed a worry for studies like these, then one might expect that the outcome of the experiment would change as a function of the hypothesis that the experimenters received.

The design of the Knobe and Prinz study has a couple of features that make it an interesting test case for experimenter bias. First, despite the fact that in this study the researchers who created the stimuli had a hypothesis about the way in which the study

should turn out, there were few (if any) checks in place to prevent unfair stimulus creation or pragmatic sabotage. Secondly, in this study, the researchers did not directly physically interact with any of their participants. So any experimenter bias effects that may be found within in a design like this would likely be purely a result of how the stimuli themselves were constructed.

2 Methods

2.1 Experimenters

Nineteen students from an introduction to cognitive science class at Yale University were recruited to help design experimental stimuli. They received extra credit for their participation and were debriefed about the purposes of the study.

2.2 Participants

Two hundred forty participants from Amazon's Mechanical Turk filled out a short survey for a small monetary reward.

2.3 Design Procedure

Experimenters stayed after class, and received a quick power point presentation in addition to written instructions. They were told that they would be helping the authors design an experiment. The authors were also the teaching assistants in the class, thus somewhat mirroring the power dynamic that occurs in real lab settings where a professor employs a graduate or undergraduate assistant to help design stimuli.

Nine participants were randomly assigned to the "feeling" condition and 10 were assigned to the "non-feeling" condition. Each group received its power point presentation separately, and all of the experimenters within a group viewed the power point presentation simultaneously.

All participants were given clear instructions on how to design their stimuli. All experimenters were told that they should create 8 sentences involving the ascription of a mental state to Acme Corp. They had to create 4 sentences that involved the ascription of a phenomenal (i.e. "feeling") mental state to Acme Corp. as well as 4 sentences that ascribed an intentional (i.e. "non-feeling") mental state to the group. For each type of mental state, experimenters were given a list of specific verbs that they must use (see below for the actual list). This list drew from the stimuli used in Knobe and Prinz (2008). Experimenters were free to choose the verb tense and anything that came after the verb itself.

They were told that the authors would then use the stimuli that they created to run their experiment on M-turk. In that experiment, participants would read each sentence and be asked to say on a scale of 1–7 how natural it sounded.

Experimenters in the feeling condition were told that the authors were interested in testing the hypothesis that people would intuitively judge the feeling mental state ascriptions to be more acceptable than the non-feeling mental state ascriptions because people intuitively believed that having a body was necessary for non-

feeling mental states (this is the opposite hypothesis from Knobe and Prinz,). And thus the authors expected that participants would judge the feeling sentences to be more natural sounding than the non-feeling sentences. Experimenters in the non-feeling condition were told that the authors were interested in testing the opposite hypothesis (i.e. that non-feeling mental state ascriptions to groups would be more natural sounding). Below is a verbatim copy of the instructions that the experimenters read in the feeling condition (the underlined parts of the instructions were switched for those participants appearing in the non-feeling conditions):

Instructions to Subject:

We would like you to help us to design an experiment. The goal of the experiment is to test whether or not people implicitly believe that groups or organizations can have minds, and if so, what kinds of minds they might have. Our hypothesis is that people naively believe that groups have minds with "feeling" mental states but not "non-feeling" mental states. Below are some examples of each type of mental state:

1) Some examples of "non-feeling" mental states
    a) to believe
    b) to intend
    c) to want
    d) to know
2) Some examples of "feeling" mental states
    a) to experience
    b) to suffer
    c) to love
    d) to feel

The difference between a "feeling" and a "non-feeling" mental state is that, "feeling" mental states have a distinct type of experience associated with them. You necessarily feel very different when you are in pain compared to when you are experiencing joy. On the other hand, "non-feeling" mental states do not have a specific, defining feeling associated with them. For example, as you started reading this instruction sheet, you presumably knew that 2+2=4, but there wasn't any particular feeling or experience you had that was associated with this knowledge.

As we stated above, our hypothesis is that people implicitly believe that groups (like corporations, teams, societies, or organizations) can have "feeling" mental states but not "non-feeling" mental states. We think this because groups do not have bodies, and other work has shown that peoples' naive beliefs make them think that having a biological body is necessary for having "non-feeling" mental states. However, people do not think that a biological body is necessary for having feelings and experiences.

We want you to help us test this hypothesis by creating stimuli with the items above. You will create full, grammatical sentences about Acme Corporation using each of the verbs listed above. Only make one sentence for each verb (this means that at the end, you will have constructed 8 sentences). In addition to the verbs in the above list, each sentence must include the noun "Acme Corporation" as the subject. You are free to use any adjectives, direct objects, or indirect objects that you see fit. Please pay attention to the length of the sentences and make sure each sentence is about the same length.

After you have created your sentences, we will show them to subjects online. Subjects will be asked to rate on a scale of 1–7 how natural each sentence sounds. If our hypothesis is correct, then subjects should rate the "non-feeling" sentences as sounding less natural than the "feeling" sentences.

If the overall experiment works, this would be really interesting and would help further scientific knowledge of how the mind works. So please try your best!

2.4 Experimental Procedure

Participants were presented with the stimuli created by the above experimenters. For each experimenter, 12 on-line participants rated their stimuli. All sentence orders were randomized. Participants were instructed to read each sentence and rate how natural it sounds on a scale of 1 (sounds weird)-7 (sounds natural). These instructions were identical to those used in Knobe and Prinz (2008).

3 Results

For each experimenter, the average ratings for their feeling and non-feeling sentences were calculated. A mixed model ANOVA (with experimenter hypothesis as a between subjects factor and sentence type as a within subjects factor) revealed a significant interaction between experimenter hypothesis (feeling vs. non-feeling) and sentence type (feeling vs. non-feeling), $F(1,17)=7.78$ $p<.05$. There was no main effect of hypothesis, $F(1,17)=.00$, $p=.99$. There was however a main effect of sentence type, with non-feeling sentences judged as more natural sounding than feeling sentences, $F(1,17)=12.976$, $p<.01$.

Post-hoc *t*-test revealed that this main effect was almost entirely carried by the experiments created by the experimenters in the "non-feeling" condition. These "non-feeling" experiments indeed revealed that participants judged the non-feeling sentences as more natural ($M=5.65$) than the feeling sentences (4.83), $t(9)=4.151$, $p<.01$. However, for the non-feeling experiments, there was no significant difference between the non-feeling ($M=5.29$) and the feeling sentences ($M=5.19$), $t(8)=.662$, $p=.527$.

Qualitatively, 10/10 of the experimenters in the non-feeling condition produced a result in which the non-feeling mean was higher (although not always statistically reliable) than the feeling mean. In contrast only, 5/9 experimenters in the feeling condition produced such a result.

4 Discussion

The above results suggest that the hypothesis that experimenters have in mind when creating their stimuli can influence the outcome of the experiment. Experimenters who were testing non-feeling hypothesis (i.e. that intentional mental state ascriptions to groups should be more natural than feeling mental state ascriptions) indeed replicated the original effect found in Knobe and Prinz (2008) whereby non-feeling sentences were rated as more natural than feeling sentences. On the other hand, those

experimenters who received the feeling hypothesis failed to replicate the original Knobe and Prinz results, and instead obtained no difference between the feeling and non-feeling sentences.

While the current design makes it impossible to say if the experiments generated by the feeling or non-feeling experimenters are closer to what would have been produced by an unbiased experimenter, the current study nevertheless supports the view that the threat of experimenter bias is real in experimental philosophy. While traditional experimenter bias effects resulting from face-to-face interaction between experimenters and participants may be less of a worry, there is nevertheless a tendency to create one's stimuli in a way that will obtain a desired or expected result.

4.1 Damning Evidence for the Group Mind Hypothesis?

The only experimenter group that was able to replicate Knobe and Prinz's original finding was the non-feeling group, who had Knobe and Prinz's original hypothesis in mind when creating their experiments. Thus in addition to providing evidence for experimenter bias, one could additionally wonder whether this pattern of results also casts doubt on the truth of Knobe and Prinz's original hypothesis. Perhaps for example, their original results were only an artifact of experimenter bias, and people do not think that having a body is necessary for the possession of phenomenal mental states.

We think that while these results may indeed cast some degree of doubt on the original findings and interpretations, any critique of the Knobe and Prinz's work based on our own should be relatively weak. It is certainly possible that the predicted result based on the non-feeling hypothesis in the original Knobe and Prinz study obtained only because the original experimenters were biased in creating their stimuli. Indeed, flaws in their designs and conclusions have been suggested elsewhere (Phelan et al. 2012; Sytsma and Machery 2009; Arico 2007).

However, upon examining the pattern of results above more closely, one could also conclude that current results do not rule out the possible truth of Knobe and Prinz's original hypothesis. The experimenters in the non-feeling group were able to replicate the original finding, while the experimenters in the feeling group were only able to obtain a null result. If both the feeling and non-feeling hypotheses were equally wrong, then one might expect the feeling experimenters to get a complete reversal of the original effect whereby the feeling sentences would be judged as more natural sounding than the non-feeling sentences. That only the non-feeling experimenters could only get the result they were looking for could thus be telling in favor of Knobe and Prinz's original hypothesis.

One cannot cite a lack of ability to create a "working" experiment as an explanation for the troublesome null result in the feeling condition. The problem for this view is that the experimenters in the non-feeling condition *were* able to create working experiments, and there is no reason to expect differences between the two groups of experimenters. Along these same lines, since there was no main effect of hypothesis on naturalness ratings, the difference in the two experiments could not be put down to one group simply making overall better or worse stimuli. It should also be noted that this null result is unlikely to be due to a low N since it has replicated in follow-up studies from our lab not reported here.

For the naysayers, there is still a way to explain the null result in the feeling condition. Perhaps it is just easier for experimenters to come up with bad examples of feeling sentences for corporations than non-feeling sentences. So on this view, there may be many sentences of both types that are totally acceptable, but it is just more difficult to pragmatically sabotage a non-feeling sentence than a feeling sentence for some uninteresting reason. This is an area where follow-up experiments could be particularly useful and informative to those interested in how these particular results bear on existing theories in this literature (see below for some further ideas about how such follow-up experiments could be constructed).

4.2 Avoiding Experimenter Bias

There are two routes that could be useful in avoiding experimenter bias in the field of experimental philosophy. The first is external and relies on the review process[2] in the hopes that external reviewers will be able to spot imperfect or biased stimuli. So long as the imperfect stimuli are accurately detected, reviewers can always prevent studies founded on biased stimuli from seeing the light of publication. It is certainly true that the review process is likely to catch at least some cases of bias, and it is important for reviewers to pay close attention to the (often very subtle) details of the stimuli themselves. Nevertheless, reviewers are unlikely to detect all such cases since (1) bias can be expressed in very subtle ways that are difficult to recognize and (2) reviewers often share the same theoretical biases as the experimenters. Therefore some of the responsibility also falls on the researchers themselves.

Thus a second, complementary way of overcoming experimenter bias (e.g. Rosenthal and Fode 1963) involves making sure that the experimenters are unaware of the relevant hypothesis. In many areas of psychology (e.g. infant research, comparative psychology and social psychology) this logic has taken root, and professional standards have developed to require that the researchers go to great lengths to avoid experimenter bias effects. For example, infancy researchers often ensure that the research assistants coding videos for looking time are blind to the hypothesis (Hamlin et al. 2007) and even that the experimenters showing the infant displays are also blind to the hypothesis (Wertz and Wynn in prep).

Based on the success of this approach, one strategy that could be applied by experimental philosophy would be "blind" stimulus creation in which the person or people designing the actual survey stimuli would be blind to hypothesis. For experimental designs involving relatively simple sentences (like the current study), it is possible to have M-Turk participants design one's stimuli anonymously. This would require clear criteria for how the stimuli should be created. Otherwise M-Turkers may produce too many unusable sentences to be of practical value. Given that this would inevitably be a sloppy process, it would also be necessary that researchers be able to exclude problematic stimuli (for example if participants did not follow instructions) in a relatively unbiased way. However, any sort of stimulus exclusion would additionally require clear exclusion criteria so that researchers do not cherry pick naturally created stimuli and unfairly exclude awkward stimuli for their hypothesis. It is worth noting that recent studies from our own lab (Strickland and Keil in prep) and others

[2] Thanks to an anonymous reviewer for this point.

(Khemlani et al. 2011) have successfully employed this method to generate interpretable findings.

However, it is also true that experiments in the experimental philosophy tradition often involve testing people's intuitions regarding longer and more complicated scenarios (e.g. Knobe 2003). For these types of scenarios, is it is less likely that untrained online communities would be able to generate stimuli that would produce anything other than statistical noise. Here, we recommend collaboration with more junior researchers (e.g. undergraduate lab assistants) so that they ultimately create all of the stimuli while being blind to the hypothesis in question. One can go to great lengths to ensure that the junior researchers in question are unaware of the more senior researcher's hypothesis while the senior researcher nevertheless guides the stimulus creation to accurately get at the theoretical question being asked in the study.

The current results make the threat of experimenter bias in experimental philosophy particularly clear. Here, we have offered some ideas on practical solutions for this problem. Regardless of whether or not these particular strategies ultimately catch on, we hope that some methods in the field will evolve to address the looming problem of experimenter bias.

Acknowledgments We would like to thank Joshua Knobe, Chaz Firestone, Alex Shaw, Brandon Liverence, Hugo Mercier and Brian Scholl for useful comments and feedback. We would also like to thank Julie Van Dyke and Laurie Santos for help finding experimenters. Finally, we thank the Yale undergraduate experimenters for designing the experiments.

References

Adair, J.G., and J.S. Epstein. 1968. Verbal cues in the mediation of experimenter bias. *Psychological Reports* 22: 1045–1053.

Arico, A. 2007. *Deregulating corporate consciousness: A critique of Knobe and Prinz's intuitions about consciousness*. Toronto: Poster presented at the Society for Philosophy and Psychology.

Barber, T.X., and M.J. Silver. 1968. Pitfalls in data analysis and interpretation: a reply to Rosenthal. *Psychological Bulletin* 70: 48–62.

Cullen, S. 2010. Survey-driven romanticism. *Review of Philosophy and Psychology* 1: 275–96.

Cushman, F.A., L. Young, and M.D. Hauser. 2006. The role of reasoning and intuition in moral judgments: testing three principles of harm. *Psychological Science* 17(12): 1082–1089.

Dickersin, K. 1990. The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association* 263: 1385–9.

Doyen, S., O. Klein, C. Pichon, and A. Cleeremans. 2012. Behavioral priming: it is all in the brain, but whose brain? *PLoS One* 7(1): e29081.

Feltz, A., E.T. Cokely, and T. Thomas Nadelhoffer. 2009. Natural compatibilism versus natural incompatibilism: back to the drawing board. *Mind and Language* 24(1): 1–23.

Feltz, A., Harris, M., and Perez, A. (2012). Perspective in intentional action attribution. *Philosophical Psychology.*

Hamlin, J., K. Wynn, and P. Bloom. 2007. Social evaluation by preverbal infants. *Nature* 450: 557–559.

Ioannidis, J.P.A. 2005. Why most published research findings are false. *Public Library of Science, Medicine* 2: e124.

Khemlani, S.S., A.B. Sussman, and D.M. Oppenheimer. 2011. *Harry Potter* and the sorcerer's scope: latent scope biases in explanatory reasoning. *Memory & Cognition* 39(1).

Knobe, J. 2003. Intentional action and side effects in ordinary language. *Analysis* 63: 190–193.

Knobe, J., and J. Prinz. 2008. Intuitions about consciousness: experimental studies. *Phenomenology and Cognitive Science* 7(1): 67–83.

Knobe, J., Buckwalter, W., Nichols, S., Robbins, P., Sarkissian, H., and Sommers, T. (2011). Experimental philosophy. *Annual Review of Psychology,* 63.

Monroe, A.E., and B.F. Malle. 2010. From uncaused will to conscious choice: the need to study, not speculate about people's folk concept of free will. *Review of Philosophy and Psychology* 1(2): 211–224.

Phelan M., Arico A., and Nichols S. (2012). Thinking things and feeling things: On an alleged discontinuity in folk metaphysics of mind.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86: 638–641.

Rosenthal, R., and K. Fode. 1963. The effect of experimenter bias on performance of the albino rat. *Behavioral Science* 8: 183–189.

Rosenthal, R., and I. Jacobson. 1968. *Pygmalion in the classroom*. New York: Holt, Rinehart and Winston.

Rosenthal, R., and D.B. Rubin. 1978. Interpersonal expectancy effects: the first 345 studies. *The Behavioral and Brain Sciences* 3: 377–386.

Scholl, B.J. 2008. *Two kinds of experimental philosophy, and their methodological dangers. Talk given at the SPP Workshop on Experimental Philosophy.* Philadelphia: University of Pennsylvania.

Strickland, B., M. Fisher, and J. Knobe. 2012. Moral structure falls out of general event structure. *Psychological Inquiry* 23(2): 198–205.

Sytsma, J., and E. Machery. 2009. How to study folk intuitions about phenomenal consciousness. *Philosophical Psychology* 22: 21–35.

Young, L., and J. Phillips. 2011. The paradox of moral focus. *Cognition* 119: 166–178.