

Comment on “How not to test for philosophical expertise”

Wesley Buckwalter

wesleybuckwalter@gmail.com

Abstract: Rini 2015 [*Synthese* 192, (2): 431-452] claims to have identified a methodological flaw that invalidates the results of two experimental studies [Schwitzgebel & Cushman (2012) *Mind and Language* 27, (2): 135-153; Tobia, Buckwalter & Stich (2013) *Philosophical Psychology* 26, (5): 629–638] demonstrating order effects in professional philosophical intuition. This conclusion is reached on the basis of unsupported empirical premises for which no evidence is given. Subsequent findings in experimental cognitive science further reveal this as unsupported speculation.

Does professional philosophical expertise reside in a superior ability to assess philosophical thought experiments and generate intuitions about them? This question has dominated recent metaphilosophical discussions in philosophy about the nature of philosophical activity. One way that researchers have investigated this question is by comparing the performance of professional philosophers and laypeople on these tasks. Several teams of independent researchers demonstrate that professionals are susceptible to cognitive biases, environmental and framing effects, and personality influences when evaluating thought experiments (Swain, Alexander, and Weinberg 2008; Schultz, Cokely, and Feltz 2011; Machery 2012; Schwitzgebel and Cushman 2012; Tobia, Chapman, and Stich 2013; Tobia, Buckwalter, and Stich 2013; Schwitzgebel and Cushman 2015). These

results suggest that professional philosophical expertise does not manifest itself in terms of generating superior intuitions to those of others.

One group of studies demonstrates that the intuitions of both professional philosophers and laypeople are sensitive to the order in which thought experiments are presented (Schwitzgebel and Cushman 2012; Tobia, Buckwalter, and Stich 2013; Schwitzgebel and Cushman 2015). In one study, for example, researchers recruited hundred of professional philosophers from philosophy departments across the United States (Schwitzgebel and Cushman 2012). Participants were asked to evaluate several hypothetical philosophical scenarios involving moral luck, the action and omission distinction, and doctrine of double effect cases reminiscent of classic “push” and “switch” trolley problems. Researchers discovered that professional philosophical intuitions varied depending on the order in which the thought experiments were presented. These order effects in professional philosophical intuition were just as large or larger than order effects there were observed among laypeople. Researchers discovered that order also influenced which abstract ethical principles professional philosophers tended to reflectively endorse. In some cases, for example, researchers found that order of presentation shifted “rates of endorsement of the doctrine of the double effect from 28% to 70%” and that this was “a very large change considering how familiar and widely discussed the doctrine is within professional philosophy” (Schwitzgebel and Cushman 2012: 149).

In a recent paper, Regina Rini (2015) claims to identify a methodological flaw that invalidates the findings by Schwitzgebel and Cushman, as well as subsequent order effect findings using a similar experimental paradigm by Tobia, Buckwalter, and Stich

(2013). This challenge is based on the idea that familiarity with hypothetical scenarios may confound performance comparisons between philosophers and laypeople:

“It is problematic to compare philosophers and non-philosophers directly on these tasks, because the two groups do not engage with the stimuli in the same way. For non-philosopher subjects, the cases and principles employed are likely to be quite unfamiliar; intuitive judgments and reasoning about them are likely to be experienced as relatively novel cognitive projects. But for philosophers, this is quite unlikely to be true. The case-types and principles employed as stimuli are well-known to philosophers. Philosophers will have memories of previously encountering the stimuli, and how they have responded in the past. So when philosophers react to these stimuli, they are doing something different from non-philosopher subjects, a different sort of cognitive task.” (Rini 2015: 440)

One might have thought that whatever differences there might be in this task afforded by familiarity gives professional philosophers an advantage in evaluating scenarios. However, the assumption about familiarity is presently used to generate an argument dismissing prior experimental findings of order effects.

The argument from familiarity goes as follows. Thought experiments used in Schwitzgebel and Cushman’s study are more familiar to professional philosophers than they are to laypeople. Because of this familiarity, professional philosophers respond to them by relying only on their memory to recall a time when they first worked out their opinion about them in the past. The process of recalling a memory is not susceptible to cognitive bias like order effects. As a result of these premises, we should not expect professional philosophers to display order effects in the experiment. Contrary to that

expectation however, philosophers do display order effects. It is inferred from this chain of reasoning that the order effects observed among professional philosophers are an “anomaly” that cannot be satisfactorily explained. As a result of this, Rini concludes that prior findings and methods associated with testing expertise in philosophy should be abandoned.

This conclusion is reached on the basis of several unsupported empirical premises for which no evidence is given. First, no evidence is given for levels of familiarity with the hypothetical scenarios administered between or within groups. It is plausible that professional philosophers may be more familiar with certain thought experiments than laypeople. Trolley problems, for example, are easily recognizable by professionals as doctrine of double effect cases perhaps. But the degree to which professional philosophers have previously worked out opinions about all seventeen of Schwitzgebel and Cushman’s novel hypothetical scenarios is doubtful. For example, consider an actual scenario administered by Schwitzgebel and Cushman:

Ralph is working at a construction site. His immediate task is to clear concrete bags off the roof of the building. The regulations require that Ralph lower the bags down in groups using a lift, rather than simply throwing the bags over the side. The reason for that rule is clear: Workers cannot see the ground from the roof of the building, so items thrown over the side might land on and kill someone. Ralph knows this, but he is feeling lazy, no one is watching, and he rightly thinks there’s only about a one-tenth of one percent (1 in 1000) chance that anybody would be near the side of the building. So Ralph violates the rule and

throws the concrete bags over the side. As it turns out, someone happens to be walking near the side of the building and is killed by a falling bag.

While “trolley” cases have become famous philosophical thought experiments, “Ralph the lazy construction worker” has not. It is highly unlikely that any professional philosopher has memories of previously encountering Ralph the construction worker stimuli or harbors a previously worked out opinion about his concrete bags. Contrary to the present claims, Schwitzgebel and Cushman’s order effects seem to have persisted over a range of cases that appear to vary widely in degree of familiarity. No evidence is offered one way or the other.

Second, no evidence is given for the manner in which philosophers processed and evaluated scenarios. Specifically, no evidence is presented that professional philosophers rely solely or predominately on memory of previous worked out opinions to answer questions in Schwitzgebel and Cushman’s experiments. The observation that a situation or general idea is familiar to us in some sense entails very little in the way of specifics about the underlying psychological processing of individual stimuli. It certainly does not establish that responses to stimuli rely on memory. Moreover, it is highly unlikely that professional philosophers could rely predominately on memory to complete the actual tasks in the experiment. To illustrate this, consider another scenario used to test order effects on intuitions in doctrine of double effect cases, arguably the cases most familiar to professional philosophers:

Mike is a firefighter inside a deadly blaze in an orphanage. He is in a room with five children, and they must be evacuated immediately or the smoke will choke them. The only way to evacuate the children is through the window, and the only

possible way to open the window is to smash it hard with a beam of wood. However, another firefighter put a toddler strapped to a hospital crib on a large platform outside the window, waiting to be rescued. If Mike smashes the window with the beam, the beam is sure to knock the crib and that one toddler off the platform, and the one toddler will die, but the five children will be safely evacuated. If Mike does not smash the window with the beam, the five children will die.

While philosophers may be able to answer standard trolley cases by relying predominately on memory, the same is not true when evaluating various cover stories featuring novel situations and additional surface details. Relying on memory to recall a previously worked out opinion about a case requires you to have actually encountered that case before. The cases involved in these experiments were not encountered before. Without positive experimental evidence demonstrating the role of memory of processing in these cases, one cannot assume that process explains the judgments in question.

Third, no evidence is offered for the claim that the process of recalling previous worked out opinions is not susceptible to cognitive bias associated with order effects. Little or no work has been done investigating the connections between memory and philosophical opinion. Perhaps relying on memory is less likely to result in errors on these tasks. But it's equally possible to hypothesize about ways in which doing this would be more likely to result in errors on these tasks. For example, perhaps relying on memory detracts attention away from the details of the scenarios presented, which results in less accurate judgments. This suggestion, like the suggestion that familiarity with the philosophical thought experiments presented increases performance, amounts to pure

speculation. Without evidence to support any of these claims about the familiarity, psychological processing, or nature of memory and bias, Rini's call to reject prior methods and findings on philosophical order effects evaporates.

It is an interesting question how professional philosophers process thought experiments and if this is substantially different than laypeople less familiar with philosophical scenarios. This question might potentially serve as the motivation for a research project testing whether familiarity is a confound in order effect research. The conclusion that familiarity is a confound in a particular experiment cannot be reached through a chain of unsubstantiated assumptions however, it must be shown. Philosophers interested in critiquing findings or methodology in experimental cognitive science should bear the requirement of evidence in mind when making sweeping conclusions about individual research papers or approaches. It is not responsible to reject findings of controlled experimental studies because the results do not cohere with expectations of an outcome based entirely on speculation. Doing so risks corrupting the research record about expertise in professional philosophical judgment, familiarity, and the reliability of methods used for testing susceptibility to bias.

In this case, the risk of this is made apparent by subsequent research in experimental cognitive science testing the role of familiarity in order effect research. Subsequent work by Schwitzgebel and Cushman (2015) has since replicated order effects in professional philosophical judgment about trolley and other doctrine of double effect cases. However the researchers also included a measure asking participants to indicate whether they were familiar with the trolley problem. They found that the majority of professional philosophers indicated they were familiar with the trolley problem. However

when researchers compared responses between those professional philosophers indicating they were familiar and unfamiliar, they found that those philosophers displayed just as large or larger order effects in their judgments. These findings further reveal the challenge to prior order effect findings and methods concerning familiarity as unsupported speculation.

References

- Machery, Edouard (2012), 'Expertise and Intuitions about Reference', *Theoria*, 27 (1), 37-54.
- Rini, Regina A. (2015), 'How not to test for philosophical expertise', *Synthese*, 192 (2), 431-52.
- Schultz, E., Cokely, E. T., and Feltz, A. (2011), 'Persistent bias in expert judgments about free will and moral responsibility: A test of the Expertise Defense', *Consciousness and Cognition*, 20 (4), 1722-31.
- Schwitzgebel, Eric and Cushman, Fiery (2012), 'Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers', *Mind & Language*, 27 (2), 135-53.
- (2015), 'Philosophers' Biased Judgments Persist Despite Training, Expertise and Reflection', *Cognition*, 141, 127-37.
- Swain, S, Alexander, J, and Weinberg, Jonathan M (2008), 'The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp', *Philosophy and Phenomenological Research*, 84 (1), 138-55.

Tobia, Kevin, Chapman, Gretchen B, and Stich, Stephen (2013), 'Cleanliness is Next to Morality, Even for Philosophers', *Journal of Consciousness Studies*, 20 (11-12).

Tobia, Kevin, Buckwalter, Wesley, and Stich, Stephen (2013), 'Moral intuitions: Are philosophers experts?', *Philosophical Psychology*, 26 (5), 629-38.