Cognition 141 (2015) 127–137

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Philosophers' biased judgments persist despite training, expertise and reflection

Eric Schwitzgebel [a,*], Fiery Cushman [b]

[a] Department of Philosophy, University of California at Riverside, USA
[b] Department of Psychology, Harvard University, USA

ARTICLE INFO

Article history:
Received 7 January 2015
Revised 23 April 2015
Accepted 24 April 2015
Available online 14 May 2015

Keywords:
Expertise
Framing effects
Loss aversion
Order effects
Reasoning
Social cognition

ABSTRACT

We examined the effects of framing and order of presentation on professional philosophers' judgments about a moral puzzle case (the "trolley problem") and a version of the Tversky & Kahneman "Asian disease" scenario. Professional philosophers exhibited substantial framing effects and order effects, and were no less subject to such effects than was a comparison group of non-philosopher academic participants. Framing and order effects were not reduced by a forced delay during which participants were encouraged to consider "different variants of the scenario or different ways of describing the case". Nor were framing and order effects lower among participants reporting familiarity with the trolley problem or with loss-aversion framing effects, nor among those reporting having had a stable opinion on the issues before participating the experiment, nor among those reporting expertise on the very issues in question. Thus, for these scenario types, neither framing effects nor order effects appear to be reduced even by high levels of academic expertise.

1. Introduction

Schwitzgebel and Cushman (2012) report that professional philosophers are no less subject to order effects on their judgments about familiar types of moral dilemmas (such as the famous "trolley problem") than are non-philosophers: When scenario pairs were presented in order AB, participants responded differently than when the same scenario pairs were presented in order BA, and the philosophers showed no less of a shift than did the comparison groups, across several types of scenario. As suggested by Sinnott-Armstrong (2008), Weinberg, Gonnerman, Buckner, and Alexander (2010), Liao, Wiegmann, Alexander, and Vong (2012), Schwitzgebel and Cushman (2012), Tobia, Buckwalter, and Stich (2013), and Mizrahi (2015), if philosophers' judgments about puzzle cases in their area of expertise are highly influenced by presumably irrelevant factors such as order of presentation or superficial differences in phrasing, that creates a prima facie challenge to certain optimistic views about philosophical expertise in assessing such scenarios – views of the sort expressed in Ludwig (2007), Grundmann (2010), and Williamson (2011; though see Buckwalter, in press; Nado, in press). It would also suggest a striking persistence of biased decision-making despite extensive training both in logical reasoning in general and in closely related task types in particular.

In the present study we attempt to establish boundary conditions on this effect. Specifically, we attempted to replicate our original effect, but then to reduce its magnitude in four ways: by (a) limiting the target group to philosophers with *expertise specifically on the types of*

* Corresponding author at: Department of Philosophy, University of California at Riverside, Riverside, CA 92521-0201, USA.
E-mail address: eschwitz@ucr.edu (E. Schwitzgebel).

http://dx.doi.org/10.1016/j.cognition.2015.04.015
0010-0277/

dilemma in question; or (b) by limiting the target group to philosophers who report *having stable opinions on the matter* (see discussion in Rini, 2015; Wright, 2010, 2013); or (c) by encouraging participants to give *reflective responses, and enforcing a delay for reflection before response*; or (d) by presenting pairs of *scenarios that differ primarily in phrasing* rather than in the relevant content of the scenario. To the extent the magnitude of the order effect is reduced by any of factors (a)–(d), that might encourage optimism about expert philosophical judgment appropriately restricted. Conversely, to the extent the magnitude of the order effect is not so reduced, that deepens the skeptical challenge.

Beyond its application to philosophical methods, our study of philosophical decision-making has a broader application to cognitive science. Over the past decades researchers have extensively documented the role of heuristics and biases in human judgment and decision-making. Often they have also argued that we would be better off if we could effectively substitute unbiased procedures (Baron, 2000; Greene, 2014; Kahneman, 2011; Thaler & Sunstein, 2008). Fewer studies address how this might be accomplished, especially in complex domains without clear feedback procedures. Here, we test some likely possibilities: Slow people down, have them think reflectively and counterfactually; familiarize them with the specific types of decisions in question; provide them extensive instruction and practice in general logical reasoning. Which, if any, of these approaches reliably reduce cognitive bias?

1.1. Prior research

Our previous study yielded two main findings. First, and receiving the most straightforward empirical support, we found that professional academic philosophers' and academic non-philosophers' moral judgments were similarly influenced by order of presentation. We tested three categories of moral judgments: several versions of the trolley problem (e.g., the footbridge and switch variants; Foot, 1967; McIntyre, 2004/2011; Thomson, 1985), cases involving moral luck (e.g., degree of blameworthiness when identical conduct such as drunk driving is either harmless or fatal; Nagel, 1979; Nelkin, 2004/2013; Williams, 1981), and cases that contrast active harm and passive harm (e.g., snatching a life preserver away from a drowning person vs. failing to offer that person your own life preserver; Bennett, 1998; Howard-Snyder, 2002/2011; Quinn, 1989). Aggregating across all three types of case we found no evidence that order effects were weaker for philosophers. Moreover, one case in a matched pair was typically more influenced by order than another. For instance, judgments of the switch version of the trolley problem were more strongly influenced by order than judgments of the footbridge version. Consequently, order had an effect on the likelihood that pairs of cases were judged to be *morally equivalent*. For instance, the switch and footbridge cases were more likely to be judged equivalently when presented in the footbridge/switch order than when presented in the switch/footbridge order.

Our second finding concerned the relationship between the judgment of specific vignettes (e.g., the switch and footbridge variants of the trolley problem) and the endorsement of abstract moral principles (e.g., the Doctrine of Double Effect, which purports to justify discrepant judgments between these cases). We hypothesized that participants – both philosophers and non-philosophers – would tend to endorse moral principles in a manner that matches their patterns of judgment. Because order of presentation influenced the likelihood of the cases being judged equivalently, this influence might carry over to influence participants' endorsement of moral principles. For philosophers, we found such an effect for the Doctrine of Double Effect and for a principle asserting the non-equivalency of moral luck cases, but not for a principle asserting the non-equivalency of action/omission cases. For non-philosophers we found precisely the opposite pattern of effects. Moreover, we identified several non-predicted effects of vignette order on endorsement among philosophers (e.g., the order of presentation of moral luck cases affected the endorsement of the Doctrine of the Double Effect). Overall, these results provided tentative evidence for an effect of order-of-judgment on the endorsement of abstract moral principles, but also suggested that such effects are highly contextually dependent.

Two other empirical studies have explored the relationship between philosophical expertise and bias in moral judgment. Tobia, Buckwalter, et al. (2013) found that professional philosophers considering moral scenarios were subject to actor–observer biases of about the same magnitude as non-philosophers' (though the groups' biases went in different directions). Tobia, Chapman, and Stich (2013) replicated this result and also found philosophers influenced about as much as were non-philosophers by the presence of a "clean" Lysol odor (though again in different directions). Relatedly, Schulz, Cokely, and Feltz (2011) find personality-related differences in philosophical experts' judgments about free will, and Machery (2011) finds subfield-related differences in judgments about linguistic reference.

There is also some research that focuses on the broader question of how expertise affects susceptibility to judgment and decision biases. Reyna, Chick, Corbin, and Hsia (2014) find that intelligence analysts are, in fact, *more* likely than college students and non-expert adults to exhibit framing effects in the Asian disease problem, and a more comprehensive meta-analysis reveals no significant effects of participant group on the magnitude of framing effects (Kühberger, 1998). There is also a substantial literature on the fairly limited effects of education on other reasoning tasks, such as the conjunction fallacy and the Wason selection task (Cheng, Holyoak, Nisbett, & Oliver, 1986; Heijltjes, van Gog, Leppink, & Paas, 2014; Lehman, Lempert, & Nisbett, 1988; Ritchhart & Perkins, 2005; Tversky & Kahneman, 1983). On the other hand, some evidence suggests that philosophers in particular might be unusually skilled at reasoning. Livengood, Sytsma, Feltz, Scheines, and Machery (2010) found that philosophers exhibited superior performance on the Cognitive Reflection Test, a series of simple math problems prone to incorrect intuitive responding (Frederick, 2005), and