Professional Philosophers’ Susceptibility to Order Effects and Framing Effects in Evaluating Moral Dilemmas

Eric Schwitzgebel
Department of Philosophy
University of California at Riverside

Fiery Cushman
Department of Psychology
Harvard University

Authors Note

Eric Schwitzgebel, Department of Philosophy, University of California at Riverside

Fiery Cushman, Department of Psychology, Harvard University

Correspondence concerning this article should be addressed to Eric Schwitzgebel, Department of Philosophy, University of California at Riverside, Riverside, California, 92521-0201, USA. Email: eschwitz at domain-ucr.edu.

Word count (excluding title, abstract, tables, figures, and bibliography): **** words
Ethicists’ Judgments

Professional Philosophers’ Susceptibility to Order Effects and Framing Effects in Evaluating Moral Dilemmas

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 large 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 to 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.

Keywords: Doctrine of the Double Effect, experimental philosophy, expertise, framing effects, loss aversion, morality, order effects, reasoning, social cognition
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), Schwitzgebel and Cushman (2012), and Tobia, Buckwalter, and Stich (2013), 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 a certain class of optimistic views about philosophical expertise in assessing such scenarios (Ludwig, 2007; Grundmann, 2010; Williamson, 2011; though see Nado, forthcoming; De Cruz, forthcoming).

In the present study we make four attempts 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 dilemma in question; or (b) by limiting the target group to philosophers who report having stable opinions on the matter (see discussion in Wright, 2010, 2013; Rini, 2014); 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 philosophical judgment appropriately restricted. Conversely, to the extent the magnitude of the order effect is not so reduced, that deepens the skeptical challenge.

Prior research

Our prior 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; Thomson, 1985; McIntyre, 2004/2011), cases involving moral luck (e.g., degree of blameworthiness when identical conduct such as drunk driving is either harmless or fatal; Nagel, 1979; Williams, 1981; Nelkin, 2004/2013), and cases that contrast active harm and passive harm (e.g., snatching a life preserver away from a drowning person versus failing to offer that person your own life preserver; Quinn, 1989; Bennett, 1998; Howard-Snyder, 2002/2011). Aggregating across all three types of case we found no evidence that order effects were weaker for philosophers; in fact, we observed a non-significant trend towards stronger order effects among philosophers. Moreover, in most cases, one case in a matched pair was 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, such as the Doctrine of Double Effect, which purports to justify discrepant judgments between these cases. We hypothesized that participants—both philosophers and non-philosophers—might choose to endorse moral principles in a manner that would match their patterns of judgment. Because order of presentation influenced the likelihood of the cases being judged equivalently, we hypothesized that this influence might carry over to influence participants’ endorsement of moral principles. For philosophers, we found the hypothesized effects 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, and Stich (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 the presence of a “clean” Lysol odor (though again in different directions).
The present study

In the present study we aimed to provide several “best case” tests of the hypothesis that philosophical expertise will diminish the influence of biasing factors on moral judgment and justification – that is, factors that we assume philosophers would not endorse as legitimate upon reflection. We chose to solicit judgments of cases that are widely discussed in the philosophical and psychological literatures, and thus that most philosophers would be familiar with: the trolley problem, and the “Asian disease” case introduced by Tversky and Kahneman (1981) to illustrate the effect of “save” versus “die” framing on judgment. We also asked whether participants were familiar with these cases so that we could limit our analyses to the subset of participants reporting familiarity. Also, for half of our participants, we explicitly instructed them to reflect before making a judgment, imposing a time delay to ensure at least a minimal level of reflection and specifically encouraging them to consider potential alternative phrasings and variants of each case before submitting their response. Finally, at the end of the test we asked participants whether they endorsed two putative moral principles, including the doctrine of double effect.

Our design allows for several tests of biased response under these conditions. We can assess (1) susceptibility to the effect of die versus save framings for Asian disease-type cases. We can ask whether order of presentation affects the judgment of (2) trolley-type problems (replicating Schwitzgebel and Cushman 2012) as well as (3) Asian disease-type problems. For trolley-type cases, we can ask whether (4) these order effects carry over to affect the endorsement of the doctrine of double effect.

Additionally, we varied the specific contrast of trolley-type cases between participants in order to provide several additional tests. Half of participants viewed the traditional footbridge
and sidetrack switch cases, while the other half of participants viewed a modified “drop” case in place of the traditional footbridge case. In the traditional footbridge case, the protagonist pushes his victim off a footbridge with his hands; in the modified “drop” case, the protagonist instead drops his victim via a lever-operated trap door. Thus, the footbridge and switch cases differ in terms of both the Doctrine of the Double Effect and also in the presence of a direct “push” involving physical contact. Both factors have been found to influence moral judgments among ordinary participants (Cushman, Young, and Hauser, 2006). In contrast, the footbridge and trapdoor cases differ in terms of the presence of a direct push but not in terms of the Doctrine of the Double Effect. This allows us to ask whether (5) order of presentation affects the endorsement of a principle distinguishing cases according to the degree to which harm is caused in a “personal” manner (e.g. a direct push). It also allows us to ask whether (6) the presence of a direct push increases the likelihood of endorsing the Doctrine of the Double Effect, despite its irrelevance to that doctrine.

Methods

Participants

We obtained email addresses of potential participants from the websites of philosophy departments and comparison departments in the United States, excluding departments that had been contacted in Schwitzgebel and Cushman (2012). An email invited recipients to participate in a study of philosophers’ and similarly educated non-philosophers’ judgments about moral dilemmas, linking to a website containing our questionnaire and encouraging recipients to forward the message to academic colleagues.
Near the end of the questionnaire we asked participants’ age, nationality, highest degree, highest degree (if any) in philosophy, and “Are you a professor of philosophy?” with response options “yes, and ethics is my area of primary specialization”, “yes, and ethics is an area of competence for me”, “yes, but not in the area of ethics”, or “no”. We excluded any participant who did not report having a graduate degree, leaving 497 respondents reporting graduate degrees in philosophy (“philosophers”), 469 (94%) with philosophy PhD’s, and 921 respondents reporting graduate degrees, but not in philosophy (“non-philosophers”), 799 (87%) with PhD’s. Among the philosophers, 104 (21%) reported being professors of philosophy with specialization in ethics and 167 (34%) reported competence but not specialization. 98% of participants reported U.S. nationality, and 33% reported being female.

*Questionnaire design*

*Reflection vs. control condition.* Half of participants were randomly assigned to a reflection condition. Before seeing any scenarios, participants in the reflection condition were told:

Over the course of the five\(^1\) questions that follow, we are particularly interested in your reflective, considered responses. After each case, please take some time to consider the different moral dimensions at issue, including potential arguments for and against the position to which you are initially attracted. Also please consider how you might respond to different variants of the scenario or to different ways of describing the case. After you finish reading each of the five cases, there will be a 15-second delay to encourage careful reflection before you

---

\(^1\) The description of “five” questions to follow was a typo that we did not notice until after conducting the survey; in fact, participants were asked to respond to six questions, not five.
are asked a question about the case. You needn’t answer immediately after the question appears. Please feel free to take as much time as you like.

Also, after each scenario, participants in the reflection condition were told:

Please take some time to consider the different moral dimensions of the scenario, including potential arguments both for and against [the action described]. Please also consider how you might respond to different variants of the scenario or different ways of describing the case. In fifteen seconds, you will be asked a question about the scenario. You needn’t answer immediately after the question appears. We want you to reflect carefully about it, so please take as much time as you like.

When you are ready to BEGIN the reflection period, hit the advance button (>>) below. The text of the scenario will remain on the screen. After 15 seconds you will be permitted to make a response, but take as much time as you would like.

Participants in the control condition were given no special instructions either to answer reflectively or to answer non-reflectively.

Trolley problems. Participants then saw two of three “trolley”-type problems, in random order. One was a Switch scenario, involving saving five people in the path of a runaway boxcar by flipping a switch to divert the boxcar onto a sidetrack where it will kill one person. The other was randomly selected to be either a Push scenario, involving saving five people by pushing a hiker with a heavy backpack into the path of a runaway boxcar, or a Drop scenario, involving saving five people by pulling a lever to drop one person into the path of a runaway boxcar.

Respondents rated each scenario on a 1-7 scale from “extremely morally good” (1) through
“neither good nor bad” (4) to “extremely morally bad” (7). The exact text of these scenarios and the rest of the questionnaire is available online in the Supplementary Online Materials. We excluded any scenario response that was produced in fewer than 4 seconds (<1% of responses in the control condition, and by design none in the reflection condition).

We used runaway boxcar scenarios to maximize philosophers’ sense of familiarity with the scenario type. Most philosophers, we believe, upon seeing any of the boxcar scenarios, would be swiftly reminded of the famous “trolley problems” of Foot (1967) and Thomson (1985). We hypothesized that philosophers seeing any one of these scenarios might be able to anticipate the type of scenario that would come next – perhaps especially in the reflection condition, in which we explicitly asked participants to “consider how you might respond to different variants of the scenario”. This design thus gave expert participants, we thought, an excellent chance to reduce the magnitude of any order effect by accurately anticipating the type of scenario that might come next.

We varied Drop and Push with the thought that participants in Drop might differ less than participants in Push in their endorsements of two abstract principles later in the questionnaire, as we will soon explain.

Framing effect scenarios. Participants then saw two of four loss aversion or framing effect scenarios of the sort made famous by Tversky and Kahneman (1981). In Save Disease, an unusual disease is expected to kill 800 people and participants chose between Program A in which 200 people would be saved and Program B in which there was a 1/4 probability that 800 people would be saved and a 3/4 probability that no people would be saved. Kill Disease was identical except that the programs were phrased in terms of how many people would die rather than how many would be saved. Save Nuclear and Kill Nuclear involved a nuclear meltdown
expected to kill 600 and probabilities of 1/3 and 2/3. Participants saw either Save Disease and Kill Nuclear or Kill Disease and Save Nuclear, in random order.

The general finding in the literature on loss aversion is that respondents tend to prefer the risky choice (Program B) when the scenario is framed in terms of how many will die and the safe choice (Program A) when the scenario is framed in terms of how many will be saved (Tversky and Kahneman 1981; Kühberger, 1998). We were interested to see whether professional philosophers, including professional philosophers explicitly encouraged to consider “different ways of describing the case”, and including professional philosophers who regard themselves as experts on framing effects and loss aversion, would show the same size framing effects as a comparison group. As with the trolley cases, we chose phrasings and cases close to the classic formulations of Tversky and Kahneman so as to give expert participants, especially in the reflection condition, an excellent opportunity to reduce the effects by trying to avoid being excessively swayed by the “saved” vs. “die” phrasing.

_Doctrine of the Double Effect and the Personal Principle._ Next we asked two questions about moral principles. First, we asked about the famous _Doctrine of the Double Effect:_ whether using one person’s death as a means of saving others is morally better, worse, or the same as killing one person as a side effect of saving others. Second we asked about a “Personal Principle”: whether helping several people by harming one person in a personal, face-to-face way is morally better, worse, or the same as helping others by harming one person in a less immediately personal way.

We predicted that philosopher participants who saw Push/Drop before Switch would be more likely to rate the scenarios equivalently and then reject the Doctrine of the Double Effect, as in Schwitzgebel and Cushman (2012). For similar reasons, we predicted that participants
would also be more likely to say it’s bad to harm in a personal way if they saw Switch before Push than if they saw Push before Switch. Given the generally lower ratings for Push than for Switch, we also predicted that participants in the Push condition would be less likely than those in Drop to say it’s better to harm in a personal way. Also, we suspected that participants in the Push condition, if they were less likely to rate the scenario pairs equivalently, might therefore be more likely than those in the Drop condition to endorse a principle that treats the scenarios inequivalently (Doctrine of the Double Effect), despite the apparent irrelevance of the Push-Drop difference to the Doctrine of the Double Effect.

**Familiarity, stability, and expertise.** Next were the demographic questions already described. Finally, we asked a few questions about familiarity, stability, and expertise. We asked four prior familiarity questions: one concerning trolley problems, one concerning loss aversion/framing effects, one concerning the Doctrine of the Double Effect, and one concerning previous empirical research on philosophers’ responses to trolley problems. Respondents who claimed familiarity both with trolley problems and with the Doctrine of the Double Effect were then asked if they regarded themselves as having expertise on those issues and if they regarded themselves as “having had a stable opinion about the trolley problem and Doctrine of Double Effect before participating in this experiment”. We asked similar expertise and stability questions for those reporting familiarity with loss aversion/framing effects. Again, see the Supplementary Online Material for exact wording.

**Results**

*Double Effect scenarios*
Means. Figure 1 displays the mean results for the Double Effect scenarios. As expected, Push was rated worse than Drop (5.3 vs. 4.5, t(1398) = 9.3, p < .001), which was rated worse than Switch (4.5 vs. 3.7, t(2094) = 11.1, p < .001). Also as expected, order effects were present for all cases and largest for Switch (Push 5.5 vs. 5.2, t(697) = 2.4, p = .02; Drop 4.8 vs. 4.3, t(699) = 4.2, p < .001; Switch 3.2 vs. 4.2, t(1393) = 12.1, p < .001).

Other predictions were tested with one multiple regression model for each scenario, predicting response from philosopher, reflection condition, presentation in second position, and all interaction variables. (Here, as in all of the linear and logistic regression models reported in this paper, we code categorical predictor variables as 1 = feature present, -1 = feature absent, and calculate interactions as the product of predictors. This allows us to interpret “main effects” of predictors and their interactions in the manner of a classic analysis of variance).

If philosophers are less subject to order effects than are non-philosophers, we would expect to see an interaction effect of philosopher by position. If philosophers are less subject to order effects specifically in the reflection condition, we would expect to see a three-way interaction between philosopher, condition, and position. Neither interaction was found for any of the three scenarios. For Push, the statistically significant predictors were presentation in the second position ($\beta = -0.09$, t(698) = -2.4, p = .02), reflection condition ($\beta = 0.09$, t(698) = 2.2, p = .03), and philosopher respondent ($\beta = -0.08$, t(698) = -2.0, p = .046). For Drop, the only significant predictors were reflection condition ($\beta = 0.21$, t(700) = 5.5, p < .001) and position ($\beta = -0.15$, t(700) = -3.9, p < .001). For Switch, the significant predictors were position ($\beta = 0.30$, t(1394) = 11.2, p < .001) and philosopher respondent ($\beta = -0.06$, t(1394) = -2.4, p = .02).

As a manipulation check, we confirmed that response time in the reflection condition exceeded that in the control condition. In the control condition the median response time was 49
seconds for the first scenario and 34 seconds for the second scenario. In the reflection condition, median response times were about double: 98 and 66 seconds respectively.
Figure 1: Trolley problem mean moral ratings by order of presentation, non-reflection (control) vs. reflection conditions, and professional philosophers vs. non-philosophers. Asterisks indicate one-tailed statistical significance at $p < .05$ for each pair of adjacent bars. For aggregated statistics see the main text.
Equivalency ratings. We also analyzed equivalency ratings. Participants were coded as having rated the scenario pairs *equivalently* if they gave both scenarios the same 1-7 rating, *inequivalently* if they rated the scenarios differently in the predicted direction (that is, Push or Drop worse than Switch), and they were excluded if they rated the scenarios differently in the unpredicted direction (that is, Switch worse than Push or Drop: 2% of participants). Equivalency ratings are less subject to scaling concerns and correspond more closely to canonical statements of the Doctrine of the Double Effect, which is generally expressed in terms of the inequivalency of harm as a means vs. harm as a side-effect.

Figure 2 shows the equivalency results. As predicted from results in Schwitzgebel and Cushman (2012), respondents were more likely to rate the scenarios equivalently if Push or Drop was presented before Switch, since ratings in the Switch scenario tend to be labile and matched to the first-presented scenario if Switch is presented second. Push and Switch were rated equivalently by 24% of respondents when Switch was presented first and 45% of respondents when Switch was presented second (Fisher’s exact, N = 689, p < .001). Drop and Switch were rated equivalently by 46% of respondents when Switch was presented first and 70% of respondents when Switch was presented second (Fisher’s exact, N = 671, p < .001).

Other predictions were tested by a binary logistic regression model, predicting response from Drop condition, philosopher, reflection condition, Switch-first condition, and all interaction variables. If philosophers are less subject to order effects than are non-philosophers, we would expect to see an interaction effect of philosopher by Switch-first. If philosophers are less subject to order effects specifically in the reflection condition, we would expect to see a three-way interaction between philosopher, reflection condition, and Switch-first. No interaction effects proved statistically significant. Significant predictors were Switch first ($e^\beta = -0.63, p < .001$),
Drop condition \( (e^\beta = 1.7, p < .001) \), and reflection condition \( (e^\beta = 0.72, p < .001) \). Higher equivalency in the Drop condition and lower equivalency in the Switch-first condition are as predicted. The model also finds lower rates of equivalency in the reflection condition, which we had not predicted (overall 54\% vs. 39\%, Fisher’s exact, \( N = 1360 \), \( p < .001 \)). A model with the three predictive variables, philosopher, and philosopher-by-Switch-first yields a non-significant trend toward smaller order effects for philosophers \( (e^\beta = 1.1, p = .08, CI 0.99 to 1.3) \).
Figure 2: Percentage of participants rating the two trolley problems equivalently, by order of presentation, non-reflection (control) vs. reflection conditions, and professional philosophers vs. non-philosophers. Asterisks indicate one-tailed statistical significance at $p < .05$ for each pair of adjacent bars.
**Framing effect scenarios**

The disease and nuclear scenarios differed only slightly in overall percentage favoring the risky choice (57% vs. 62%, N = 2732, p = .007) and did not detectably differ in the size of the framing or order effects, so the two scenario types were merged for analysis. Figure 3 displays the results.

Median response time in the control condition was 45 seconds for the first-presented scenario and 33 seconds for the second-presented scenario. In the reflection condition, median response times were a bit less than double: 77 seconds and 59 seconds, respectively.

**Framing effects.** In the first-presented scenario, as Tversky and Kahneman (1981) and most subsequent studies have found, participants were much more likely to select the risky option (Program B) when the options were expressed in terms of how many of the disaster victims expected to die “will die” than when an otherwise equivalent pair of options was presented in terms of how many “will be saved”. (To see the traditional framing effect in the figure, look only at the dark bars in the graphs which represent the first-presented scenarios compared between participants.) The effect was large both for philosophers (79% vs. 32%, Fisher’s exact, N = 475, p < .001) and for non-philosophers (83% vs. 43%, Fisher’s exact, N = 903, p < .001).

Other predictions were tested by a binary logistic regression model, predicting response from “die” frame, philosopher, reflection condition, and all interaction variables. Significant predictors were frame ($e^\beta = 2.7, p < .001$) and philosopher respondent ($e^\beta = 0.84, p = .008$). In a model with frame, philosopher, and frame-by-philosopher, philosophers showed nominally larger framing effects, but this effect did not approach significance: interaction $e^\beta = 1.1, p = .41$, CI 0.93 to 1.2).
Order effects. To see the order effects in Figure 3, compare the pairs of adjacent dark and light bars. In every case, preference for the risky option is significantly closer to 50% when the scenario is presented second, having followed presentation of a very similar scenario with the opposite framing, than when the scenario was presented first. Order manipulation substantially moderates the framing effect: In the second-presented scenarios, 60% of participants selected the risky option in the “die” frame vs. 56% in the “saved” frame (Fisher’s exact, N = 1368, p = .10).

Other predictions were tested by two binary logistic regression models, predicting “die”-frame response and “save”-frame response from philosopher, reflection condition, second-position presentation, and all interaction variables. If philosophers are less subject to order effects than are non-philosophers, we would expect to see an interaction effect of philosopher by position. If philosophers are less subject to order effects specifically in the reflection condition, we would expect to see a three-way interaction between philosopher, reflection condition, and position. No interaction variable was significant in either model. In both analyses, responses were closer to 50% in the second position (“die” frame: $e^{β} = 0.59$, p < .001; “save” frame: $e^{β} = 1.5$, p < .001) and philosophers were less likely to favor the risky choice (“die” frame $e^{β} = 0.84$, p = .007; “save” frame $e^{β} = 0.85$, p = .005). Models with position, philosopher, and position-by-philosopher show nominally larger interaction order effects for philosophers that do not approach statistical significance: (“die”: $e^{β} = 0.95$, p = .38, CI 0.83 to 1.07; “saved”: $e^{β} = 1.1$, p = .22, CI 0.96 to 1.20).
Figure 3: Percentage of participants favoring the risky choice in “die” vs. “saved” framing scenarios, by order of presentation, non-reflection (control) vs. reflection conditions, and professional philosophers vs. non-philosophers. Asterisks indicate one-tailed statistical significance at p < .05 for each pair of adjacent bars.
Endorsement of Principles

We did not see the expected order effects on endorsement for either the Doctrine of the Double Effect or the Personal Principle. Philosophers’ DDE endorsement (worse to harm as means than as side effect) was 59% with Switch first vs. 62% with Push or Drop first (Fisher’s exact, N = 462, p = .57, CI for diff -12% to +6%). Non-philosophers were actually more likely to endorse DDE if they received Push or Drop before Switch: 51% vs. 58% (Fisher’s exact, N = 845, p = .045) – a result for which we have no explanation, but which we also found in Schwitzgebel and Cushman (2012). For the Personal Principle philosophers split 26% better, 63% same, 11% worse when Switch was first vs. 27%-59%-14% when Push or Drop was first ($\chi^2 = 1.4, p = .49$); for non-philosophers it was 26%-61%-13% vs. 30%-61%-10% ($\chi^2 = 3.3, p = .19$). Only 66 philosopher participants unfamiliar with previous research on philosophers’ responses to trolley problems were in the Push version of the control condition – the condition closest to the original Schwitzgebel and Cushman (2012) – unfortunately too few to allow a meaningful direct comparison with that study’s finding of a 62% vs. 46% order effect on endorsement of DDE. (The confidence interval for the effect size in this subgroup does include the observed 16% effect size in our earlier work: 22/40 vs. 14/26, CI for difference -26% to +23%, $Z = 0.1, p = .93$.)

We did find that philosophers’ endorsements of the Personal Principle were substantially influenced in the predicted direction by whether they had been assigned to the Push or Drop condition. In the Drop condition, 32% of philosophers said harm done in a personal manner was morally better than harm done impersonally, compared to 22% in the Push condition (58% vs. 63% said “same”, and 10% vs. 16% said worse, 3x2 $\chi^2 = 8.4, p = .02$). In contrast, non-philosophers showed no detectable effect (30%-59%-10% vs. 25%-62%-12%, $\chi^2 = 3.2, p = .20$).
consistent with Schwitzgebel and Cushman’s (2012) finding that philosophers were more likely than non-philosophers to shift their endorsements of principles to match their experimentally manipulated judgments about scenarios.

Finally, we found some evidence that endorsements of the DDE were also influenced by whether participants were assigned to the Push or Drop condition. As we saw above, participants who viewed Drop were more likely to rate it equivalent to Switch than were participants who viewed Push. Both pairs of cases differ along a dimension captured by the DDE; participants’ higher likelihood of rating the Push-Switch pair inequivalently than the Drop-Switch pair is likely a consequence of the additional presence of an up-close, personal harm in Push. We reasoned that participants might exhibit greater endorsement of the DDE as a convenient explanation for their discrepant judgments of Push and Switch than for their more weakly discrepant judgments of Drop and Switch cases. This effect is of particular interest because it involves the misapplication of an intuition driven by one stimulus feature (an up-close, personal harm) to the endorsement of another stimulus feature (harm caused as a means to an end). Consistent with this predicted effect we found that non-philosophers were more likely to endorse the DDE when they had seen Push (60%) than if they had seen Drop (50%; Fisher’s exact, N = 845, p = .005). We found a non-significant trend in the same direction for philosophers (63% vs. 57%; Fisher’s exact, N = 462, p = .18 [p = .099 one-tailed]).

**Familiarity, Stability, and Expertise**

A substantial majority of philosophers, 77%, claimed to be familiar with both the trolley problem and the Doctrine of the Double Effect, while 20% claimed expertise and 40% claimed to have had a stable opinion about the trolley problem and the Doctrine of the Double Effect before
participating in the study. For non-philosophers, the corresponding percentages were 9%, 1%, and 3%. A substantial minority of philosophers, 33%, claimed familiarity with empirical studies of philosophers’ responses to trolley problems, compared to 4% of other respondents. For framing effects / loss aversion, the corresponding percentages for philosophers were 62% familiarity, 13% expertise, 26% stability; for non-philosophers 24%, 4%, and 9%.

Philosophers self-reporting familiarity, expertise, stability, and specialization in ethics appeared to be just as subject to order effects as did philosophers reporting unfamiliarity, lack of expertise, lack of stability, and lack of specialization in ethics. See Figures 4 and 5. As is evident from the figures, philosophers reporting familiarity, expertise, stability, and specialization in ethics trended toward showing larger order effects than the remaining philosophers. For example, among philosopher respondents reporting being philosophy professors with an area of specialization in ethics, 26% rated the scenarios equivalently when Switch was first vs. 56% when Push or Drop was first (Fisher’s exact, N = 99, p = .004), compared to a 42%-52% shift for all remaining philosopher respondents (Fisher’s exact, N = 376, p = .06 [one-tailed p = .03]), a difference in order effect size marginally significant in a binary logistic regression (interaction effect of order and specialization $e^\beta = 1.3$, p = .06).
Figure 4: Percentage of philosophers rating the two trolley problems equivalently, by order of presentation, broken down by types of expertise. Asterisks indicate one-tailed statistical significance at p < .05 for each pair of adjacent bars.
Philosophers reporting familiarity, expertise, and stability also trended toward showing larger die-vs.-saved framing and order effects rather than did philosophers reporting lack of familiarity, expertise, and stability. This trend was statistically significant in the die frame, where philosopher participants reporting familiarity showed significantly larger order effects ($\beta = 1.3, p = .02$).
Figure 5: Percentage of philosophy participants favoring the risky choice in “die” vs. “saved” framing scenarios, by order of presentation and level of familiarity or expertise. Asterisks indicate one-tailed statistical significance at $p < .05$ for each pair of adjacent bars.
Ethicists’ Judgments

Even combining expertise and stability does not appear to reduce the estimated effect size, though the numbers are too small for the comparisons to be anything other than suggestive. Among philosophers reporting both expertise and stability regarding the Doctrine of the Double Effect, 12/34 (35%) rated the scenarios equivalently when Switch was presented first, vs. 19/39 (49%) when Switch was presented second (Fisher’s exact, \( p = .34 \), CI for difference -9% to 35%), a 14% shift comparable to the 16% shift among the remaining philosophers. Among philosophers reporting both expertise and stability regarding loss aversion, 24/28 (86%) selected the risky option in the first scenario in the “die” frame, vs. 8/19 (42%) in the “saved” frame (Fisher’s exact, \( p = .003 \), CI for difference 18%-69%), a 44% shift comparable to the 50% shift among the remaining philosophers.

Discussion

Replicating prior research, we found substantial order effects on participants’ judgments about trolley-type moral dilemmas, substantial order effects on their judgments about making risky choices in loss-aversion-type scenarios, and substantial framing effects on their judgments about making risky choices in loss-aversion-type scenarios.

Moreover, we could find no level of philosophical expertise that reduced the size of the order effects or the framing effects on judgments of specific cases. Across the board, professional philosophers (94% with PhD’s) showed about the same size order and framing effects as similarly educated non-philosophers. Nor were order effects and framing effects reduced by assignment to a condition enforcing a delay before responding and encouraging
participants to reflect on “different variants of the scenario or different ways of describing the case”. Nor were order effects any smaller for the majority of philosopher participants reporting antecedent familiarity with the issues. Nor were order effects any smaller for the minority of philosopher participants reporting expertise on the very issues under investigation. Nor were order effects any smaller for the minority of philosopher participants reporting that before participating in our experiment they had stable views about the issues under investigation.

Previous research has found substantial loss-aversion framing effects even among fairly sophisticated participants (reviewed in Kühberger 1998; see also García-Retamero & Dhami, 2013). The present study confirms and extends these results to very high levels of expertise. That the effect is present in participants with very high levels of expertise raises the question of whether those experts might in fact be responding rationally to relevant factors, contrary to our initial assumptions in experimental design. For example, Mandel (2014) argues that participants might naturally read “200 people will be saved” as meaning something like at least 200 people will be saved (and maybe more), and comparably “400 people will die” as meaning something like at least 400 people will die – in which case it might be rational to prefer the risky choice in the die frame and the safe choice in the save frame. If Mandel’s explanation were correct in the present case, however, we would expect to see the same frame-driven pattern in the second-presented scenarios as in the first-presented scenarios, since the wording is the same; and we would expect to see smaller framing effects among expert participants who were presumably aware that the intended interpretation of the options is exact numbers saved and dying, not minimum numbers. It remains open, however, that there are other ways of interpreting the framing effects and order effects as rational, contra existing psychological orthodoxy.
Our findings on the effect of contextual factors on philosophers’ endorsement of moral principles were more equivocal. We found that assignment to different pairs of trolley cases (Drop-Switch, not differing in degree of personal contact between agent and victim, vs. Push-Switch, very different in degree of personal contact between agent and victim) substantially influenced philosophers’ endorsements of a principle regarding the value or disvalue of harming in a personal face-to-face way. We found a predicted but non-significant trend towards an effect of the case presented on the endorsement of the Doctrine of the Double Effect. However, we did not replicate Schwitzgebel and Cushman’s (2012) finding that the order of presentation of trolley-type dilemmas affects philosophers’ subsequent endorsement of the Doctrine of the Double Effect. This is consistent with Schwitzgebel and Cushman’s finding that philosophers’ endorsements of abstract principles are substantially influenced by contextual factors, such as perhaps in this case the presence of the loss-aversion cases, the absence of the moral luck and action-omission cases, and the assignment of half of the participants into a reflection condition. On the whole, further research would help to clarify the effect of order, case, and similar spurious factors on philosophers’ endorsement of moral principles.

We confess that we find our main result surprising: that is, our across-the-board failure to find evidence for philosophical expertise and reflection in moderating biased moral judgment. We would have thought – we are still inclined to think – that at a high-enough level of expertise people won’t be much swayed by order and framing, at least when they report having stable opinions and are encouraged to answer reflectively. We wouldn’t expect, for example, that Judith Jarvis Thomson (1985) or John Martin Fischer (Fischer & Ravizza, 1992) would rate Push and Switch equivalently if the scenarios are presented in one order and inequivalently if they are presented in another order. However, if there is a level of philosophical expertise that reduces
the influence of factors such as order and frame upon one’s moral judgments, we have yet to find empirical evidence of it.

Acknowledgements

We thank Daniel Guevara, Regina Rini, and Walter Sinnott-Armstrong for their valuable suggestions regarding this manuscript. We gratefully acknowledge the University of California, Riverside Academic Senate and **** for their financial support.


Nado, J. forthcoming: Philosophical expertise and scientific expertise. Philosophical Psychology.


Tobia, K., Chapman, G., & Stich, S. 2013: Cleanliness is next to morality, even for philosophers. Journal of Consciousness Studies, 20 (no. 11-12), 195-204.


