

In press *Psychological Science*.

Comparing Hypothetical and Real-Life Trolley Problems

Andrew M. Colman¹, Natalie Gold², and Briony D. Pulford¹

¹ School of Psychology, University of Leicester, UK

² Faculty of Philosophy, University of Oxford, UK

Author Note

Correspondence concerning this article should be addressed to Andrew M. Colman, School of Psychology, University of Leicester, Leicester LE1 7RH, United Kingdom. E-mail: amc@le.ac.uk

The study by Bostyn, Sevenhant, and Roets (2018) was not the first to compare hypothetical and real-life trolley problems (see Gold, Colman, & Pulford, 2014; Gold, Pulford, & Colman, 2014, 2015). Bostyn et al. changed the victims to mice and the harm to electric shock. Some of the participants decided whether to press a button to redirect an electric shock from a cage containing five mice to a cage containing one; others imagined that they were faced with the same decision and were asked: “Would you press the button?” The researchers found that 84% of participants pressed the button, compared to only 66% who predicted that they would.

This finding is consistent with studies showing that people are bad at predicting their own behavior, including moral decisions. Loewenstein (1996, 2000, 2005) has argued that there is a hot-cold empathy gap: In affectively “cold” states, people fail to appreciate fully how “hot” states will affect their preferences and behavior. People also overestimate how bad a negative visceral feeling can be (Kang & Camerer, 2013). This could explain why participants say that they will sacrifice more money to spare others from mild electrical shocks than they actually do when the shocks are real (FeldmanHall et al., 2012), and it could also explain Bostyn et al.’s (2018) results.

We should be careful about generalizing from decisions about causing pain to decisions involving other types of harm. Gold, Colman, and Pulford (2014) found the opposite result to Bostyn et al. (2018) in a trolley problem in which the victims were children in a Ugandan orphanage and the harm was losing a pre-allocated meal: In a real decision, 80% of British and 49% of Chinese participants clicked a switch to save five children from losing meals, with one other child losing a meal instead, whereas in a hypothetical version of the task, 91% of British and 73% of Chinese participants predicted that that they would click the switch. One obvious difference between the two studies is that Bostyn et al.’s study involved harming mice, whereas Gold et al.’s involved harming people. Bostyn et al. state that “recent research suggests that there is a symmetry between how people tend to treat animals and other humans” (p. 1091), but this may not apply to trolley-type dilemmas. Preferences for redistributing pain are different from preferences for redistributing money (Story et al., 2015), but there is evidence that moral judgments in hypothetical trolley problems involving economic harms follow the same patterns as moral judgments involving life-or-death decisions and other major physical harms (Gold, Pulford, & Colman, 2013).

Bostyn et al. (2018) did not answer the research question that motivated their article: “whether subjects’ hypothetical moral judgments are predictive of the actual behavior they would display in a dilemma-like situation in real life” (p. 1085). Most of the literature using hypothetical moral dilemmas takes moral judgments as the dependent variable of interest, but Bostyn et al. did not elicit moral judgments in their hypothetical and real scenarios. They argued that “hypothetical-dilemma research, while valuable for understanding moral cognition, has little predictive value for actual behavior and that future studies should investigate actual moral behavior along with the hypothetical scenarios dominating the field” (p. 1084). This fails to acknowledge that investigating moral judgments in and of themselves—the approach taken in most trolley problem research—is a valid and theoretically important approach to the study of moral reasoning, irrespective of the judgment–behavior discrepancy. Further, had Bostyn et al. elicited moral appropriateness judgments in hypothetical and real mouse dilemmas, then they would have been on firmer ground in making claims about a judgment-behavior discrepancy than they were by showing that behavior was not correlated with a measure of “moral preferences” for consequentialist reasoning derived from a hypothetical dilemma battery.

Using the paradigm involving meals for Ugandan children described above, Gold, Pulford, and Colman (2015) elicited both actions and moral judgments about real-life trolley decisions. Participants could intervene in an on-screen animation to change which children lost their meals. As well as a condition that corresponded to Bostyn et al.’s (2018) version of the trolley problem, we also tested the well-known footbridge variation, in which the five could be saved by dragging a photo of the single child into the path of the threat. Consistent with previous research, we found a difference in moral judgment between the standard and footbridge versions, with significantly more participants judging the action permissible in the standard version than in the footbridge version. However, we found no significant difference between the two versions in actions actually performed—clicking the switch in the standard version or dragging the photo in the footbridge version—although this nonsignificant difference cannot be interpreted as evidence for the null hypothesis in actions actually performed. Nevertheless, actions were significantly different from permissibility judgments and, in a regression model, rightness judgments were related to behavior. It seems possible—even likely—that hypothetical moral judgments have predictive power for actual behavior.

The relationship between moral judgments and moral decisions appears to be complex. To add to the complexity, moral judgments may differ depending on whether they are made on the basis of reading a vignette about a dilemma or actually viewing the dilemma in real life. In an experiment designed to compare moral judgments in a hypothetical trolley problem to moral judgments in a corresponding real-life scenario (where, in both, the victims were quiz participants who could lose their winnings), Gold, Pulford, and Colman (2014) found differences in moral judgments between the hypothetical and the real scenarios.

Some researchers have suggested that differences between moral judgments and moral actions indicate that judgment and action are underpinned by two separate processes (Tassy et al., 2013). Another obvious alternative is that what people consider morally permissible does not exhaust the factors that they actually consider when making moral decisions (Gold et al., 2015).

The reasons for these discrepancies need to be investigated further, but preferably without deception of participants. In Bostyn et al.'s (2018) experiment, mice were not shocked, even in the "real-life" version, although the participants were told that they would be. Deception is increasingly avoided in judgment and decision-making research, and in behavioral and experimental economics it has been prohibited since the 1990s (e.g., Davis & Holt, 1993). Experimental rigor relies on participants believing what they are told by experimenters, and the avoidance of deception in experimental economics has been cited as one of the principal reasons why economic experiments replicate more successfully than psychological experiments (Camerer et al., 2016). Bardsley et al. (2010) explained the potential effects of deception:

If deceptive practices are used by other researchers, knowledge of that fact might spread among a local subject pool (e.g., by word of mouth). More worryingly, it could spread more widely as knowledge of experimental method is disseminated, for example through journals, and teaching in which experimental research is discussed. The importance of this transmission route should not be underestimated, given that experimental subjects tend to be drawn from among university students. (p. 283)

Epley and Huff (1998) provided compelling evidence that participants who are deceived become suspicious, and that their suspicion remains elevated for several months. Reviewing all the evidence a decade later, Hertwig and Ortmann (2008) reported: "We found evidence that suspicion has the potential to adversely impact research outcomes, both in the experiment at hand and in subsequent studies" (p. 81).

Another reason to avoid attempting to deceive experimental participants is that such attempts can easily fail, without the experimenters realizing it. The Milgram experiment is now so well known that one wonders whether participants in Bostyn et al.'s (2018) experiment really believed that shocks would be delivered in the real-life condition. They reported in post-experimental debriefing sessions that they felt uncomfortable (p. 1089), but that is hardly convincing evidence. In deceptive experiments we always need to consider who was deceived: the participants or the experimenters?

Author Contributions

All three authors contributed to the commentary. A. M. Colman drafted the manuscript, N. Gold and B. D. Pulford made revisions, and all authors approved the final version.

Orcid iD

Andrew M. Colman <http://orcid.org/0000-0003-0499-3479>

Briony D. Pulford <http://orcid.org/0000-0001-8636-9957>

Natalie Gold <https://orcid.org/0000-0003-0706-1618>

Acknowledgements

Preparation of this article was supported by an award from the Arts and Humanities Research Council of the UK (Grant No. AH/H001158/1).

Declaration of Conflicting Interests

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

References

- Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton, NJ: Princeton University Press.
doi.org/10.1515/9781400831432
- Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, 29(7), 1084–1093. doi:10.1177/0956797617752640
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmeld, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280), 1433–1436. doi:10.1126/science.aaf0918
- Davis, D. D., & Holt, C. A. (1993). *Experimental economics*. Princeton, NJ: Princeton University Press.
- Epley, N., & Huff, C. (1998). Suspicion, affective response, and educational benefit as a result of deception in psychology research. *Personality and Social Psychology Bulletin*, 24(7), 759–768. doi:10.1177/0146167298247008
- FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, 123(3), 434–441. doi:10.1016/j.cognition.2012.02.001
- Gold, N., Colman, A. M., & Pulford, B. D. (2014). Cultural differences in responses to real-life and hypothetical trolley problems. *Judgment and Decision Making*, 9(1), 65–76.
- Gold, N., Pulford, B. D., & Colman A. M. (2013). Your money or your life: Comparing judgements in trolley problems involving economic and emotional harms, injury and death. *Economics and Philosophy*, 29, 213–233. doi:10.1017/S0266267113000205
- Gold, N., Pulford, B. D., & Colman, A. M. (2014). The outlandish, the realistic, and the real: Contextual manipulation and agent role effects in trolley problems. *Frontiers in Psychology: Cognitive Science*, 5, 35, 1–9. doi:10.3389/fpsyg.2014.00035
- Gold, N., Pulford, B. D., & Colman, A. M. (2015). Do as I say, don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *Journal of Economic Psychology*, 47, 50–61. doi:10.1016/j.joep.2015.01.001
- Hertwig, R., & Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1), 59–92. doi:10.1080/10508420701712990
- Kang, M. J., & Camerer, C. F. (2013). fMRI evidence of a hot-cold empathy gap in hypothetical and real aversive choices. *Frontiers in Neuroscience*, 7, 104. doi:10.3389/fnins.2013.00104
- Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, 65(3), 272–292. doi:10.1006/obhd.1996.0028
- Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2), 426–432. doi:10.1257/aer.90.2.426
- Loewenstein, G. (2005). Hot–cold empathy gaps and medical decision making. *Health Psychology*, 24(4), S49–S56. doi:10.1037/0278-6133.24.4.S49
- Story, G. W., Vlaev, I., Metcalfe, R. D., Crockett, M. J., Kurth-Nelson, Z., Darzi, A., & Dolan, R. J. (2015). Social redistribution of pain and money. *Scientific Reports*, 5(1), 15389. doi:10.1038/srep15389
- Tassy, S., Oullier, O., Mancini, J., & Wicker, B. (2013). Discrepancies between judgment and choice of action in moral dilemmas. *Frontiers in Psychology*, 4, 250. doi:10.3389/fpsyg.2013.00250