Commentary

What Is the Right Question for Moral Psychology to Answer? Commentary on Bostyn, Sevenhant, and Roets (2018)

Michał Białek[1,2] ⓘ, Martin Harry Turpin[1], and Jonathan A. Fugelsang[1]
[1]Department of Psychology, University of Waterloo, and [2]Centre for Economic Psychology and Decision Sciences, Koźmiński University

We know ourselves only as far as we've been tested.

–Wisława Szymborska (2016, p. 35)

Imagine that a mad philosopher has tied five people to some trolley tracks, and a speeding trolley is barreling toward them. You see that there is a track switch that if thrown would redirect the trolley onto a set of side tracks, where it would hit a single worker instead. How likely are we to ever find ourselves in this situation? This is the question most often asked of moral psychologists by researchers from unrelated fields. What they are really asking, of course, is what could we possibly learn from such hypothetical, unlikely, and artificially constructed scenarios? The hidden assumption is that they already know the answer, and that answer is "not much." We disagree.

Providing a result that would appear consistent with these antitrolley arguments, Bostyn, Sevenhant, and Roets (2018) have shown that hypothetical scenarios inform little to nothing of our understanding of real-life moral decision making. In their study, 208 participants reported on their willingness to sacrifice one person to save five people in hypothetical versions of the classic trolley dilemma. Afterward, they completed either a hypothetical or a real-life dilemma in which they had to choose whether to shock a single mouse to save five mice from an electric shock. The essential finding of this experiment was that responses to hypothetical trolley dilemmas failed to predict people's future behavior. Participants tended to cause harm to one innocent mouse more often than they reported they would. Similar findings from experiments in virtual reality seem to be consistent with this: People are actually more willing to cause harm than they report they would in a hypothetical situation (Francis et al., 2016; Francis et al., 2017; Patil, Cogoni, Zangrando, Chittaro, & Silani, 2014). In other words, people's stated preferences more strongly fall in line with deontological principles (e.g., do no harm) than do their revealed preferences. As the argument goes, asking them what they would do elicits only their stated preferences and confounds our understanding of moral decision making.

Decisions on whom to kill or save are scarcely made by the average person, and participants in psychological experiments only rarely, if ever, have faced such problems in their life. Therefore, when asked about their predicted behavior, people are forced to use their intuitions instead of relying on past experiences (Beshears, Choi, Laibson, & Madrian, 2008). Such intuitions are, however, not a perfect predictor of what people will really do, as their behavior is often affected by the context in which such choices are made (Bartels, 2008). Adding to the raft of research that has been conducted in psychology and behavioral economics identifying the disconnect between what people say they want and how they actually behave, Bostyn et al.'s finding that our stated preferences do not accurately predict our revealed preferences in the moral dimension would have been a natural extension of the existing work on thought–action disconnect (Blake, McAuliffe, & Warneken, 2014). Yet still, to some researchers, failure to accurately predict behavior about a future case invalidates years of trolley studies.

Corresponding Author:
Michał Białek, University of Waterloo, Department of Psychology, PAS Building, 200 University Ave. West, Waterloo, Ontario N2L 3G1 Canada
E-mail: mbialek@uwaterloo.ca

Although the mismatch between moral thoughts and action would make for an interesting topic on its own, we wish to focus on other, broader issues relevant to Bostyn et al.'s article. Specifically, we ask whether the ultimate goal of moral psychology should be to predict human behavior in such scenarios. To some extent, yes, we may be interested in extreme cases where one's moral behavior is truly tested, and if this is the case, hypothetical moral scenarios would seem to be suboptimal. However, we think that intuitions discovered with hypothetical trolley scenarios can answer several far more consequential questions: What people believe the current moral rules of their communities are, what those rules should be (similar views have been expressed in the popular press; see, e.g., Engber, 2018), and how they judge the actions of others. For instance, we use the perception of our own morality as a way to sort ourselves into "moral communities": deciding whom to vote for, selecting social partners, etc. Understanding these intuitions is essential, as the moral dimension is dominant in forming impressions of other people (Goodwin, Piazza, & Rozin, 2014; Uhlmann, Pizarro, & Diermeier, 2015). For example, when judging other people's trustworthiness on the basis of their responses to moral hypotheticals, people heavily consider the types of decisions that they themselves would have made (Bostyn & Roets, 2017; Everett, Pizarro, & Crockett, 2016). Investigating moral intuitions even if they do not align with real behavior is also important because of the sheer volume of such community-forming judgments. Most of us will only rarely (if ever) face problems as consequential as a real-life trolley dilemma. However, we share the planet with more than seven billion other decision makers, so there will never be a shortage of people making such decisions that will be available for us to judge. In other words, despite almost never finding ourselves in the driver's seat for such moral decisions, we are almost never shy about being moral backseat drivers.

The finding that sacrificial decisions in hypothetical moral dilemmas are associated with psychopathy (Bartels & Pizarro, 2011; Koenigs, Kruepke, Zeier, & Newman, 2012) tells us a lot about forming impressions of other people. We know that people will judge those who opt for sacrificial options in such moral dilemmas harshly. Relatedly, we also understand why people might feel negatively toward health economists, who must regularly assign monetary and material value to individual human lives and occasionally trade those lives for other benefits (Tinghög & Västfjäll, 2018). The dislike of both health economists and people who choose to sacrifice the few for the benefit of the many may stem from the perception that these decision makers are lacking in empathy or are indifferent to causing harm to other people (Uhlmann, Zhu, & Tannenbaum,

2013). But as we learn from Bostyn et al., if put in their place, people would likely make the exact same utilitarian decision.

Considering the above, we posit that the critical question for moral psychology is often what affects peoples' judgments about what they would do, what others should be doing, and what the hypothesized motives of people who make such decisions are. And to this end, hypothetical sacrificial moral dilemmas are just fine.

Action Editor

D. Stephen Lindsay served as action editor for this article.

Author Contributions

M. Białek developed the concept for the Commentary and drafted the manuscript. M. H. Turpin and J. A. Fugelsang provided critical revisions. All authors approved the final version of the manuscript for submission.

ORCID iD

Michał Białek https://orcid.org/0000-0002-5062-5733

Declaration of Conflicting Interests

References

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, *108*, 381–417.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*, 154–161. doi:10.1016/j.cognition.2011.05.010

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed? *Journal of Public Economics*, *92*, 1787–1794.

Blake, P. R., McAuliffe, K., & Warneken, F. (2014). The developmental origins of fairness: The knowledge–behavior gap. *Trends in Cognitive Sciences*, *18*, 559–561.

Bostyn, D. H., & Roets, A. (2017). Trust, trolleys and social dilemmas: A replication study. *Journal of Experimental Psychology: General*, *146*(5), e1–e7.

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*, 1084–1093. doi:10.1177/0956797617752640

Engber, D. (2018, June 18). Does the trolley problem have a problem? *Slate*. Retrieved from https://slate.com/technology/2018/06/psychologys-trolley-problem-might-have-a-problem.html

Everett, J. A., Pizarro, D. A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*, 772–787.

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality:

Transitioning from moral judgment to moral action? *PLOS ONE*, *11*(10), Article e0164374. doi:10.1371/journal .pone.0164374

Francis, K. B., Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports*, *7*, Article 13954. doi:10.1038/ s41598-017-13909-9

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148–168.

Koenigs, M., Kruepke, M., Zeier, J., & Newman, J. P. (2012). Utilitarian moral judgment in psychopathy. *Social Cognitive and Affective Neuroscience*, *7*, 708–714.

Patil, I., Cogoni, C., Zangrando, N., Chittaro, L., & Silani, G. (2014). Affective basis of judgment-behavior discrepancy in virtual experiences of moral dilemmas. *Social Neuro-science*, *9*, 94–107.

Szymborska, W. (2016). *Map: Collected and last poems* (S. Barańczak & C. Cavanagh, Trans.). Boston, MA: Mariner Books.

Tinghög, G., & Västfjäll, D. (2018). *Why people hate health economics – Two psychological explanations* (LiU Working Papers in Economics, No. 6). Linköping, Sweden: Division of Economics, Department of Management and Engineering, Linköping University.

Uhlmann, E. L., Pizarro, D. A., & Diermeier, D. (2015). A person-centered approach to moral judgment. *Perspectives on Psychological Science*, *10*, 72–81.

Uhlmann, E. L., Zhu, L. L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*, 326–334.