Journal of Economic Psychology 47 (2015) 50–61

Contents lists available at ScienceDirect

Journal of Economic Psychology

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

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

Natalie Gold [a,*], Briony D. Pulford [b,1], Andrew M. Colman [b,1]

[a] Department of Philosophy, King's College London, UK
[b] School of Psychology, University of Leicester, UK

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 3 November 2014
Accepted 6 January 2015
Available online 12 January 2015

JEL classification:
D03
D63
D64
C91

PsycINFO classification:
2340
2630

Keywords:
Moral behavior
Moral decision making
Moral judgments
Trolley problems

ABSTRACT

Many people judge that it is permissible to harm one person in order to save many in some circumstances but not in others: it matters how the harm comes about. Researchers have used trolley problems to investigate this phenomenon, eliciting moral judgments or behavioral predictions about hypothetical scenarios where five people can be saved at the cost of harming one other person. We operationalized trolley problems in the laboratory, with economic incentives and real-life consequences, allowing us to observe not only judgments but actual decisions. We varied whether the five were saved by clicking a switch that diverted the harm to the one or by dragging the one in front of the harm. We found differences in moral judgments between the two tasks, but no differences in behavior. The judgments of actors and observers also differed, with observers judging it more right to act. Our results suggest that the difference between moral judgments and actions arises because participants think that doing the right action still involves doing something morally discreditable, and that the morality of taking action does not exhaust the normative reasons for acting.

1. Introduction

Trolley problems were devised by philosophers in order to investigate why it is permissible to cause a harm to one person in order to save many in some circumstances but not in others (Foot, 1967; Thomson, 1976, 1985). The paradigm trolley problem is *Side-track*: There is a runaway trolley that threatens to kill five men on the track ahead. An agent can save the five by switching a lever that will divert the trolley onto a Side-track. However, on the Side-track is one man, who would be killed. This contrasts with *Footbridge*, where the agent can save the five by pushing a large man off a footbridge in front of the trolley, stopping the trolley but killing the one. In both cases, the decision is whether to take an action that results in

* Corresponding author at: Philosophy Department, King's College London, Strand, London WC2R 2LS, UK. Tel.: +44 (0)20 7848 2750.
E-mail addresses: natalie.gold@rocketmail.com (N. Gold), bdp5@le.ac.uk (B.D. Pulford), amc@le.ac.uk (A.M. Colman).
[1] Address: School of Psychology, University of Leicester, Leicester LE1 7RH, UK. Tel.: +44 (0)116 229 7197.

http://dx.doi.org/10.1016/j.joep.2015.01.001
0167-4870/

the death of one person in order to save five. However, many people have the intuition that it is morally permissible to turn the trolley in Side-track but not to push the man in Footbridge. It matters how the harm to the one and the saving of the five come about.

There is a lot of evidence that people make different moral judgments in hypothetical Side-track and Footbridge problems (Gold, Pulford, & Colman, 2013; Greene et al., 2009; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Mikhail, 2011). However, there is little evidence about people's actual behavior. Some previous experiments on actions in moral dilemmas have asked participants to predict their own behavior (e.g. Bartels, 2008; Petrinovich & O'Neill, 1996; Schaich Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006; Tassy, Oullier, Mancini, & Wicker, 2013). Predictions of behavior have been shown to be notoriously unreliable (Osberg & Shrauger, 1986; Vallone, Griffin, Lin, & Ross, 1990). They may be especially problematic in moral dilemmas, because people's predictions may be biased toward whatever response they think is more socially desirable, even if this would not be reflected in their actual behavior (Koritzky & Yechiam, 2010). The best way to discover what people would do is to observe their actions, but we cannot operationalize life and death trolley problems in the laboratory.

Gold et al. (2013) found, using hypothetical scenarios, that the difference in moral intuitions between Side-track and Footbridge is preserved when the outcomes are economic harms. This suggests that we can study trolley problems using the methodology of experimental economics, which has already been used to investigate moral behaviors such as altruism, fairness, trust, cooperation, and reciprocity (e.g. Andreoni, Brown, & Vesterlund, 2002; Andreoni & Miller, 2002; Berg, Dickhaut, & McCabe, 1995; Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999, 2006; Houser & Kurzban, 2002). In this article, we report a laboratory experiment, comparing judgments and behavior in real-life Side-track and Footbridge scenarios, where decisions resulted in actual small economic harms to one or to five. We investigate whether there are behavioral differences between different trolley problems, what the patterns of moral judgments are in real-life trolley problems, and whether behavior corresponds to moral judgments.

1.1. Behavior in trolley problems

As far as we know, there are no previous comparisons of behavior in real-life trolley problems, but there are some relevant precursors. Navarrete, McDonald, Mott, and Asher (2012) studied behavior in Side-track in a virtual reality environment and found that 90.5% of participants turned the trolley. This is similar to the percentage of people who judge turning the trolley to be morally permissible in hypothetical scenarios. However, despite the heightened level of realism, the virtual reality environment still does not lead to actual outcomes for real people.

Côté, Piff, and Willer (2013) ran what was effectively a hypothetical version of Footbridge with small economic harms, asking participants whether they would take money from one other participant to benefit three others. This set-up is analogous to Footbridge for the following reason: An important difference between Side-track and Footbridge, according to moral philosophers, is that in Side-track the harm to the one occurs as a foreseen side effect when saving the five, whereas in Footbridge harming the one is a necessary means to saving the five. (If the one person was not there, then turning the train would still save the five in Side-track but the five in Footbridge could not be saved.) In Côté, Piff, and Willer's study, the one is harmed as a means to benefit the three, hence it is analogous to Footbridge but with small economic harms. They found that 60% of their participants said that they would take the money. However, the decisions were only hypothetical; the outcomes were not actually realized.

Our experimental design is based on a task used by Hsu, Anen, and Quartz (2008) who, while investigating the brain regions involved in trade-offs between equity and efficiency, used a computer animation where participants could flip a switch to divert a threat from one group of children to another. The outcomes were actually implemented, using small economic incentives. Gold, Colman, and Pulford (2014) adapted this methodology to study cultural variations in Side-track, finding that Chinese participants were less likely to divert the threat than British.

None of the preceding experiments compared behavior in Side-track and Footbridge. In this paper, we use the same methodology as Gold, Colman et al. (2014), supplemented with a Footbridge variant, to compare behavior and moral judgments in Side-track and Footbridge scenarios with real economic outcomes.

1.2. Relationship between behavior and moral judgments

Another reason to implement a real-life trolley task is to elicit moral judgments about a real situation, and to compare judgments with behavior. It is not obvious that moral judgments made in hypothetical situations will be the same as moral judgments made in real-life situations with actual consequences (FeldmanHall et al., 2012; Gold, Pulford, & Colman, 2014), or that behavior in trolley problems will correspond to moral judgments. People's actual moral behavior often does not live up to the moral attitudes that they express, not even that of ethicists (Schwitzgebel & Rust, 2014). People's predictions of their behavior in moral dilemmas are more utilitarian than their moral judgments, so that they are more likely to predict that they would harm one to benefit many than they are to judge that it is acceptable to do so (Kurzban, DeScioli, & Fein, 2012; Tassy et al., 2013).

Tassy et al. (2013) suggest three possible explanations for the divergence between moral judgment and predicted behavior. One possibility is that moral judgments and behavior are the outputs of separate psychological processes (Separate Processes Hypothesis). A second possibility is that differences between judgment and choice are due to *akrasia*, or