Economics and Philosophy, **29** (2013) 213–233 © Cambridge University Press
doi:10.1017/S0266267113000205

YOUR MONEY OR YOUR LIFE: COMPARING JUDGEMENTS IN TROLLEY PROBLEMS INVOLVING ECONOMIC AND EMOTIONAL HARMS, INJURY AND DEATH

NATALIE GOLD
Kings College, London, UK
natalie.gold@rocketmail.com

BRIONY D. PULFORD
University of Leicester, UK

ANDREW M. COLMAN
University of Leicester, UK

There is a long-standing debate in philosophy about whether it is morally permissible to harm one person in order to prevent a greater harm to others and, if not, what is the moral principle underlying the prohibition. Hypothetical moral dilemmas are used in order to probe moral intuitions. Philosophers use them to achieve a reflective equilibrium between intuitions and principles, psychologists to investigate moral decision-making processes. In the dilemmas, the harms that are traded off are almost always deaths. However, the moral principles and psychological processes are supposed to be broader than this, encompassing harms other than death. Further, if the standard pattern of intuitions is preserved in the domain of economic harm, then that would open up the possibility

The authors gratefully acknowledge support from the Arts and Humanities Research Council grant AH/H001158/1, and from the University of Leicester for granting study leave to the second author. We thank the editors of this special issue and three anonymous referees for helpful input, and our research assistant Manisha Chauhan.

of studying behaviour in trolley problems using the tools of experimental economics. We report the results of two studies designed to test whether the standard patterns of intuitions are preserved when the domain and severity of harm are varied. Our findings show that the difference in moral intuitions between bystander and footbridge scenarios is replicated across different domains and levels of physical and non-physical harm, including economic harms.

1. INTRODUCTION

There is a long-standing debate in philosophy about whether it is morally permissible to harm one person in order to prevent a greater harm to others and, if not, what is the moral principle underlying the prohibition. A standard methodology involves devising hypothetical scenarios describing moral dilemmas in order to probe moral intuitions, and then coming to a reflective equilibrium between intuitions and principles. For this purpose, Phillipa Foot (1967) introduced the trolley problem, which grew into a family of scenarios extended and named by Judith Thomson (1985).[1] The common basis of the scenarios is as follows: A runaway trolley threatens to kill five men on the track ahead. It is possible to save the five but, in doing so, one other person would be killed. The question is whether it is morally permissible to kill one person in order to save five. In one scenario, *bystander*, the agent can save the five by switching a lever that will divert the trolley onto a side-track. On the side-track is one man, who would be killed. In another scenario, *footbridge*, the agent can save the five by pushing a large man off a footbridge in front of the train, stopping the train and saving the five, but killing the one. Thomson (1985) surmised that most people would have the intuition that it is permissible to switch the lever but not to push the large man. More recent experimental research shows that Thomson was correct (Petrinovich *et al*. 1993; Greene *et al*. 2001; Cushman *et al*. 2006).

The harms that are traded off in trolley problems are deaths. We present two experimental studies designed to examine whether the differences in moral judgements between *footbridge* and *bystander* will generalize across other types of harm than death, including other bodily harms (both permanent and impermanent), emotional harm and economic harms.

Death is ubiquitous in the examples discussed in the philosophical literature, such as: whether to manufacture a gas that saves five, at the cost of releasing lethal fumes that kill one (Foot, 1967); the permissibility of strategic versus terror bombing (Quinn, 1989*b*); splitting a life-saving

[1] Thomson named the puzzle of finding a moral principle that could explain the difference in intuitions between scenarios ‘the trolley problem’. But nowadays one is just as likely to find the individual scenarios being called ‘trolley problems’.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

medicine between five people who need it versus giving it to one person who needs it all (Kamm, 1985); allowing people to die of starvation in Africa versus sending them poisoned food (Unger, 1996); allowing people to die of starvation in Africa versus walking past someone who is dying in the street when it would be possible to stop and help (Unger, 1996). However, the moral principles that are derived from these scenarios are not supposed to be solely about the permissibility of killing. Rather, they are about trade-offs between harms more generally conceived; death is merely an illustrative example. As Warren Quinn (1989*a*) says, 'Harm here is meant to include any evil that can be the upshot of choice, for example, the loss of privacy, property, or control. But to keep matters simple, my examples will generally involve physical harm, and the harm in question will generally be death' (287, n 2). Quinn (1989*b*: 334, n 1) makes the same caveat.

There is an assumption that intuitions about cases involving deaths will generalize to cases involving other harms. But this is not obviously true. Loss of privacy, property and control are very different from physical harm and death. Even if we could place the different harms on a spectrum according to their severity, death would be at the extreme end of the spectrum and extreme cases may be treated very differently from intermediate ones. For instance, in the area of decision-making under uncertainty, it is well known that the ends of the probability spectrum, certainty and impossibility, are treated very differently from the possible but uncertain outcomes in the middle (Kahneman and Tversky 1979).

It is not just philosophers who extrapolate general theories from intuitions associated with life-threatening situations. In his theory of moral cognition, John Mikhail (2007, 2011) postulates that our moral intuitions are produced by a 'Universal Moral Grammar', a complex set of rules, concepts and principles that generate structured mental representations of actions, which underpin our judgements of the normative status of those actions. He argues that intuitions in trolley problems 'can be explained by postulating tacit knowledge of two specific legal rules: the prohibition of intentional battery and the principle of double effect' (Mikhail 2007: 145). The theory is supposed to apply to instances of battery but, in Mikhail's examples and evidence, the harm invoked is virtually always death. The sole exception is a hybrid case where there is a trade-off between death and property damage.[2] In a version of *bystander*, where there is five million dollars of new railway

[2] Peter Unger (1996) also discusses hybrid cases, although not ones involving trolleys, in the context of an argument that our intuitive responses are not a good guide to moral values. For instance, we would think it seriously morally wrong if a driver refused to give someone with a bleeding leg a lift to hospital, causing the loss of the leg, because the blood would do $5000 of damage to the car's upholstery. One the other hand, we do not find it intuitively wrong to ignore a request for $100 from UNICEF that would save 30 children's

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

equipment on the main track and the bystander can throw the switch, diverting the train onto a side-track and killing a man, Mikhail found that none of the participants in an experiment judged it morally permissible to turn the train. This evidence was given in support of an auxiliary principle, that it is not permissible to throw the switch when the costs outweigh the benefits. Mikhail's aim was not to test whether differences in intuitions between bystander and footbridge cases extend to other types of harm and, although his theory is framed in terms of battery, the evidence adduced never involves other bodily harms.

Mikhail's (2007, 2011) approach is typical of the experimental literature on trolley problems, where the harmful outcomes of the actions being judged are virtually always deaths, although the scenario has sometimes been modified so that it does not involve trolleys (e.g. Cushman *et al*. 2006) or extended to include moral dilemmas other than killing one to save five (e.g. Greene *et al*. 2004, 2008, 2009; Moore *et al*. 2008). From these scenarios, broader conclusions are drawn about, for example, the role of emotion in moral judgement (Greene *et al*., 2001; Koenigs *et al*., 2007), or the dependence of moral judgement on whether the locus of intervention is the agent of harm itself or the victim who is moved into its path (Waldmann and Dieterich 2007).

One exception is a pair of probes used by Michael Waldmann and Jörn Dieterich (2007), where the harm is paraplegia rather than death. The scenario does not involve trolleys, but it preserves the key features of trolleys: a threat to many that, in one case, can be prevented by being redirected toward one person and, in another case, can be prevented by the physical use of the one.

Virus bystander case:

> A virus causing paraplegia threatens four patients. Through the ventilation system, the virus could be redirected into a room with one patient.

Virus footbridge case:

> A virus causing paraplegia threatens four patients. The bone marrow of one patient could save them. However, the required procedure would lead to paraplegia in this patient.

They asked participants to rate whether or not the actor in the scenario should perform the action, on a scale from 1 (definitely not) to 6 (definitely yes), and found that the *bystander* case (M = 4.0, SE = 0.25) was rated higher than *footbridge* (M = 2.1, SE = 0.18).

lives. The car upholstery case has been included in experiments by Joshua Greene *et al*. (2004, 2008). Moore *et al*. (2008) give subjects a hybrid footbridge-style case, where a man's life is saved by pushing a valuable sculpture into the path of a lawnmower. However, it was a filler task and no results were reported.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

Another exception is found in Shaun Nichols and Ron Mallon (2006), where participants judged a dilemma based on Foot's (1967) trolley scenario but with teacups substituted for people. In their scenarios, a mother forbids her child from breaking any teacups and then leaves the house. The teacup version of *bystander* has the five teacups on a toy train track, with one teacup on a side-track. In the teacup version of *footbridge*, the child can stop a toy truck that is about to smash into five teacups by throwing and smashing one other cup. They found that in their version of *footbridge* involving teacups, 92% of participants said that it was 'all-in permissible' for the protagonist to throw the teacup (they do not report the results for *bystander*).

Teacup bystander case:

> Billy's mother leaves the house one day; she says 'you are forbidden from breaking any of the teacups that are on the counter'. Later that morning, Billy starts up his train set and goes to make a snack. When he returns, he finds that his 18-month-old sister Ann has taken several of the teacups and placed them on the train tracks. Billy sees that if the train continues on its present course, it will run through and break five cups. Billy cannot get to the cups or to the off-switch in time, but he can reach a lever, which will divert the train to a side track. There is only one cup on the side track. He knows that the only way to save the five cups is to divert the train to the side track, which will break the cup on the side track. Billy proceeds to pull the lever and the train is diverted down the side track, breaking one of the cups.

Teacup footbridge case:

> When Susie's mother leaves the house one day, she says 'you are forbidden from breaking any of the teacups that are on the counter'. While Susie is playing in her bedroom, her 18-month-old brother Fred has taken down several of the teacups and he has also turned on a mechanical toy truck, which is about to crush five of the cups. As Fred leaves the room, Susie walks in and sees that the truck is about to wreck the cups. She is standing next to the counter with the remaining teacups and she realizes that the only way to stop the truck in time is by throwing one of the teacups at the truck, which will break the cup she throws. Susie is in fact an excellent thrower and knows that if she throws the teacup at the truck she will save the five cups. Susie proceeds to throw the teacup, which breaks that cup, but it stops the truck and saves the five other teacups.

Nichols and Mallon's results suggest that intuitions may not generalize across all types of harm. However, the breaking of teacups is not a paradigm case of moral decision making, and, to the extent that it is a trade-off between harms (presumably to the mother who owns the tea cups), (a) the harm of having one's teacups smashed is a very minor damage to property, and (b) all the harms in the scenario are done to only one person, the mother.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

It is an open question whether intuitions about life-and-death outcomes correspond to intuitions about other harms. Our main aim is to provide an answer to this question. In doing so, we test an assumption of domain generality that is often made in philosophy and psychology. If differences were to emerge from experimental research, then they might provide information that would be useful in understanding the origins, basis and nature of those intuitions.

If differences do not emerge and trolley intuitions generalize to economic harms, then that would open up interesting new avenues of research. In contrast to the trolley literature, which focuses on life–death hypothetical cases, the paradigm economics experiment involves participants actually experiencing small economic benefits and harms. This approach has been used to investigate moral behaviours such as altruism, fairness, trust, cooperation and reciprocity (e.g. Berg *et al*. 1995; Fehr and Schmidt 1999, 2006; Bolton and Ockenfels 2000; Andreoni and Miller 2002; Andreoni *et al*. 2002). It has also been used to complement and extend research on the Knobe effect, the finding that people say that harmful foreseen side effects are caused intentionally but helpful ones are not (Utikal and Fischbacher 2009). For obvious reasons, the original trolley problem cannot be replicated with real incentives. However, if the standard pattern of intuitions is preserved in hypothetical scenarios involving only property damage, then that would raise the possibility of investigating behaviour in trolley problems using economic incentives.

2. STUDY 1

This study was designed to examine whether intuitions in trolley problems differ depending on the degree of bodily harm suffered by the victims and to compare death, other bodily harms, and property damage.

2.1 Method

Participants. Participants were voluntary respondents to an online survey. There were 546 participants: 192 men, 319 women, and 35 who did not answer the question about their gender. They were not paid for their participation. The survey was promoted online and through UK university e-mail lists. The majority of the participants (69%) were British, the rest came from all over the world. Participants were aged between 15 and 67 years (M = 27.59, SD = 10.38).

Design and materials. We compared 10 scenarios in a 2 (Scenario: *footbridge* vs. *bystander*) × 5 (Type of harm: *death*, *limb loss*, *broken bones*, *bruising*, *property*) between-subjects design. Scenarios were based on the trolley problem, but the setting was located down a mine to increase the plausibility of some of the outcomes.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

The *bystander-death* scenario:

> Peter is visiting a mine when he sees an out-of-control trolley that is full of coal approaching. There are five miners on the trolley track who do not have time to get out of the way. If nothing is done, they will be killed.
>
> The track has a side-track leading off to the left. Peter is standing next to a lever. If he pulls the lever, that will turn the trolley onto the side track and the five miners on the main track will not die. But there is one miner on the side track who does not have time to get out of the way. If the trolley goes onto the side track, then the miner on the side track will die.
>
> Peter is aware of all these facts.
>
> Thus, Peter can pull the lever, in which case the one miner will die but the five miners will not; or Peter can refrain from pulling the lever, in which case the five miners will die but the one miner will not.
>
> Is it morally wrong for Peter to pull the lever?

The *footbridge-death* scenario:

> Peter is visiting a mine when he sees an out-of-control trolley that is full of coal approaching. There are five miners on the trolley track who do not have time to get out of the way. If nothing is done, they will be killed.
>
> Peter is standing on a bridge over the trolley tracks, next to a miner who is twice as large as Peter. If Peter pushes the miner off the footbridge and onto the tracks, the trolley will collide with the miner, who weighs just enough to make it stop, and the five miners on the main track will not die. But the collision will cause the death of the large miner.
>
> Peter is aware of all these facts.
>
> Thus, Peter can push the miner, in which case the one miner will die but the five miners will not; or Peter can refrain from pushing the miner, in which case the five miners will die but the one miner will not.
>
> Is it morally wrong for Peter to push the man?

In the other harm scenarios, we replaced the last sentence of the first paragraph as follows:

limb loss: If nothing is done, the trolley will roll over their legs and they will all lose their legs.

broken bones: If nothing is done, the trolley will roll over their legs and their legs will all be broken.

bruising: If nothing is done, they will be severely bruised and will be confined to bed for some weeks.

property: There are five rucksacks on the tracks, each containing personal items including phones and laptops, through no fault of their owners.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

> It is not possible for Peter to get the rucksacks off the tracks in time. The speed of the trolley is such that, if nothing is done, the phones and the laptops will be crushed.

In each scenario, we made equivalent substitutions for 'die' and 'death' in other paragraphs, and removed all references to 'miners' in the *property* scenario, just referring to 'rucksacks'. So, in every scenario, the harm that threatened the five was the same as the harm that befalls the one, and in the property scenario the choice was to save five rucksacks by destroying one. The full text of all the scenarios is available in the online supplementary materials.

Procedure. Participants completed the study in their own time, online, after following a link to a SurveyGizmo site. Participants read a consent form and were assured of the anonymity of their data. After granting consent they were randomly allocated to read only one of the 10 scenarios. After reading the scenario, they were asked (a) Is it morally wrong for Peter to *push the man/push the rucksack/pull the lever*? (Yes/No), (b) Should Peter *push the man/push the rucksack/pull the lever*? (Yes/No), (c) Now please indicate how wrong or how right you think it would be to *push the man/push the rucksack/pull the lever*? (0 *Definitely wrong* to 10 *Definitely right*).

2.2 Results

For all five levels of harm, there was a significant effect of the *bystander* versus *footbridge* condition on judgements of whether or not the action was morally wrong: death: χ^2 (1, 111) = 50.525, $p < 0.001$; loss of limb: χ^2 (1, 76) = 28.070, $p < 0.001$; broken bones: χ^2 (1, 99) = 38.500, $p < 0.001$; bruising: χ^2 (1, 131) = 62.662, $p < 0.001$; property: χ^2 (1, 129) = 10.601, $p < 0.001$. The overall pattern is that more participants judged that taking the action was morally wrong in the *footbridge* condition than in the *bystander* condition. Figure 1 shows the percentage of participants in each condition who indicated that 'No', it was not morally wrong for Peter to do the action. We display the data this way in order to make it more easily comparable with studies that have asked whether the action is permissible or acceptable.[3]

[3] In deontic logic, anything that is not morally wrong is morally permissible so, in theory, our results are directly comparable with studies that ask if the action is 'morally permissible'. In practice, the conversational usage of 'morally wrong' and 'morally permissible' may differ from that prescribed by formal logic. In particular one might worry that 'morally permissible' is not often used in everyday speech, hence our asking if the action was 'morally wrong'. Sinnott-Armstrong *et al.* (2008) use the wording 'morally wrong' for similar reasons. Note that O'Hara *et al.* (2010) tested wording effects in moral judgements and concluded that studies with different wordings can legitimately be compared.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

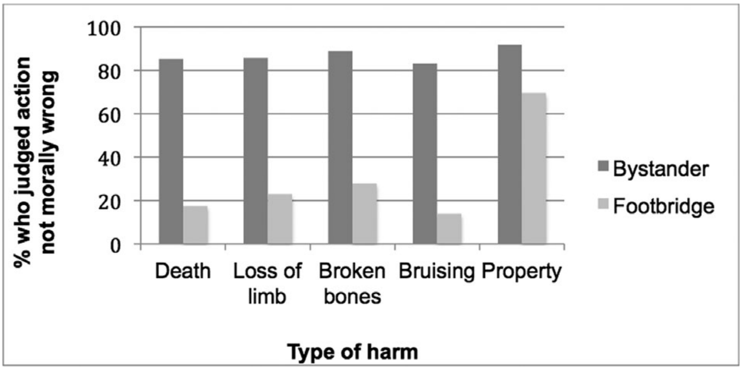

FIGURE 1. Results of Study 1: Percentage of participants in each condition who said 'No' it was not morally wrong for Peter to do the action.

Within the *bystander* condition, there was no effect of type of harm, χ^2 (4, 271) = 2.665, p = 0.615, as there was exceptionally high agreement among participants that it was not morally wrong to pull a lever to sacrifice one person to save five people from injury or death, or to damage one rucksack to save five. Within the *footbridge* conditions there was a significant effect of type of harm, χ^2 (4, 275) = 57.321, $p < 0.001$. Pairwise contrasts reveal that this is driven by differences in judgements in the *property-footbridge* condition, where significantly more people said that pushing the rucksack was not morally wrong compared with pushing the man in all four other *footbridge* conditions (all $p < 0.001$, see Table 1). In the other *footbridge* conditions, the majority of participants judged the action to be morally wrong whereas, in the *property-footbridge* condition, the majority of participants judged pushing the rucksack not to be morally wrong. Hence the difference between the *bystander* and *footbridge* scenarios was smallest in the *property* conditions, with 70% of people in the *property-footbridge* condition judging it not to be morally wrong to destroy property in order to save larger amounts of property, compared with 92% in *property-bystander* (although, as noted above, this difference is still highly significant).

In the follow-up question 'Should Peter *push the man/push the rucksack/pull the lever*?' there is no significant difference in the *bystander* condition between the proportion of people in each of the five different harm conditions who judge that 'Yes', Peter should take the action: χ^2 (4, 264) = 2.877, p = 0.58, most people indicate 'Yes'. However, there is a significant effect of scenario in the *footbridge* condition: χ^2 (4, 268) = 37.869, $p < 0.001$. Here, most people judge that 'No', Peter should not take the action, except in the property condition where 62% of people indicated that 'Yes', he should push the rucksack onto the track to save the

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

measure	harm				
	death	loss of limbs	broken bones	bruising	property
bystander					
morally wrong, % No	85% $_{a}$	86% $_{a}$	89% $_{a}$	83% $_{a}$	92% $_{a}$
should, % Yes	80% $_{a}$	85% $_{a}$	81% $_{a}$	76% $_{a}$	87% $_{a}$
right–wrong, mean	6.53 $_{a}$	6.52 $_{a}$	6.66 $_{a}$	6.34 $_{a}$	7.31 $_{a}$
(*SD*)	(2.51)	(1.83)	(1.92)	(2.30)	(2.47)
footbridge					
morally wrong, % No	18% $_{a}$	23% $_{a}$	28% $_{a}$	14% $_{a}$	70% $_{b}$
should, % Yes	16% $_{a}$	28% $_{a}$	33% $_{a}$	17% $_{a}$	62% $_{b}$
right–wrong, mean	3.00 $_{a}$	2.75 $_{a}$	3.75 $_{a}$	2.75 $_{a}$	5.79 $_{b}$
(*SD*)	(3.10)	(2.38)	(3.26)	(2.16)	(2.64)

Note: Percentages on the same row that do not share a common subscript differ at $p < 0.05$ according to a χ^2 test. Means on the same row that do not share a common subscript differ at $p < 0.05$ according to a post-hoc Tukey test.

TABLE 1. Results of Study 1: Percentages of 'no' responses to the question 'Is it morally wrong to do the action?', percentage of 'yes' responses to the question 'Should Peter do the action?', and mean right–wrong ratings.

other property from being damaged. The pattern of the effects matches the 'is it morally wrong?' results already described (see Table 1). Comparing the answers to the question about whether it is morally wrong to take the action with the answers about whether Peter should actually take the action, using McNemar tests for each of the 10 scenarios, we found no significant differences at all between these two questions. So if people thought the action was morally wrong, then they subsequently judged that Peter should not do it.

The results of a two-way ANOVA on the 0 to 10 ratings of how wrong or right taking the action would be show that there was a significant difference between the five harms, $F(4, 523) = 12.902$, $p < 0.001$, partial $\eta^2 = 0.09$, and between the *bystander* and *footbridge* scenarios, $F(1, 523) = 186.467$, $p < 0.001$, partial $\eta^2 = 0.26$. There was also a significant interaction between the type of harm and the *bystander* and *footbridge* conditions, $F(4, 523) = 3.852$, $p = 0.004$, partial $\eta^2 = 0.029$. As is clear from Figure 2, there is no significant difference in the *bystander* ratings across the types of harm, the ratings are consistently high, and the ratings in *footbridge* are all significantly lower (see also Table 1). According to a test for simple effects, the difference in ratings between *bystander* and *footbridge* conditions was highly statistically significant for all the harms tested (all $p < 0.001$). The difference between *bystander* and *footbridge* (although still significant)

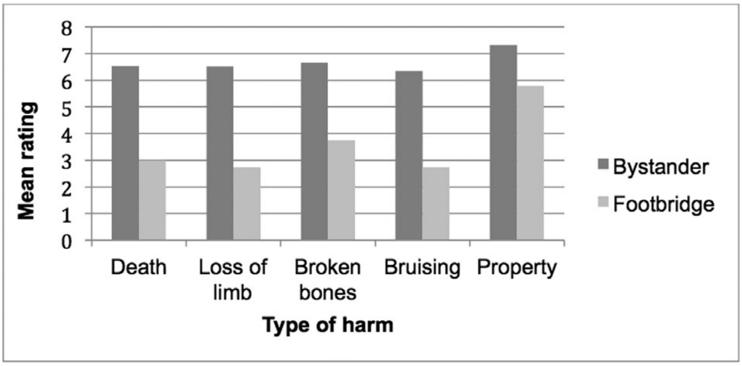

FIGURE 2. Results of Study 1: Mean ratings of how wrong or right it would be to take the action.

decreases in the *property* condition because, in the *footbridge* condition, it is judged less wrong to harm one in order to save five when the harm is property damage, rather than bodily harm (see Table 1). In fact, with a mean right–wrong rating of 5.79, the *property-footbridge* is the only one of the *footbridge* scenarios where the mean rating indicates that it is right to take the action. The results of the right–wrong ratings corroborate the results of the binary 'is it morally wrong?' question.

3. STUDY 2

Study 1 examined degrees of physical harm and damage to property in trolley dilemmas. We also wanted to examine consequences such as emotional distress, and economic consequences such as losing one's job or losing significant amounts of money – harms that were not obviously possible to investigate with the trolley scenario. Thus we designed a new scenario, the *sauna* dilemma. The *sauna* dilemma preserves the key feature of trolley problems, namely a harm that threatens five people, which could be diverted onto one by pressing a switch (*bystander*) or stopped by pushing one person into its path (*footbridge*). This allowed us to investigate a broader range of harms.

3.1 Method

Participants. Participants were voluntary respondents to an online survey. There were 386 participants: 152 men, 239 women and 37 people who did not answer the question about gender. They were not paid for their participation. The majority of the participants (68%) were British, the rest came from all over the world. Participants were aged between 17 and 72 years (M = 28.29, SD = 11.64). Study 1 and Study 2 were promoted

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

together and accessed via the same web link, to reduce the likelihood that the same people would participate in both studies (where it was clear that a person had taken part in a study more than once we used only their first set of data).

Design and materials. We compared eight scenarios in a 2 (Scenario: *footbridge* vs. *bystander*) × 4 (Type of harm: *death*, *job loss*, *financial loss*, *emotional distress*) between-subjects design.

The *bystander-emotional distress* scenario:

> Five strangers are having a nude sauna in a spa in Finland. They all have strict principles of modesty but have each made an exception to their usual rule because nude saunas are the norm in Finland. Unknown to them the spa keeps surveillance cameras in the sauna. The camera has malfunctioned and is about to feed live to the internet. There will be five seconds of internet feed before the camera automatically switches off. The nudity of those in the sauna will be exposed in public and people that they interact with in daily life will be able to see the pictures. The five will discover that they were nude in public. That will breach their principles of modesty and they will each suffer extreme emotional distress.
>
> Peter is outside the sauna and knows what is about to happen. He cannot turn off the camera or the internet feed any faster than it will turn off by itself. But by pressing a switch, he can make the five seconds of internet feed come from a surveillance camera in the spa's other sauna and therefore the five people will not suffer the extreme emotional distress that would result from their being nude in public. However, there is one person in the other sauna, who also has a strict principle of modesty and will suffer extreme emotional distress because he will be nude in public and people that he interacts with in daily life will see the pictures.
>
> None of the people will get any compensation because the sauna will go bankrupt due to the negative publicity.
>
> Peter is aware of all these facts.
>
> Thus, Peter can press the switch, in which case the one person will suffer extreme emotional distress but the five people will not; or Peter can refrain from pressing the switch, in which case the five people will each suffer extreme emotional distress but the one person will not.
>
> Is it morally wrong for Peter to press the switch?

The *footbridge-emotional distress* scenario:

> Five strangers are having a nude sauna in a spa in Finland. They all have strict principles of modesty but have each made an exception to their usual rule because nude saunas are the norm in Finland. Unknown to them the spa keeps surveillance cameras in the sauna. The camera has malfunctioned

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

> and is about to feed live to the internet. There will be five seconds of internet feed before the camera automatically switches off. The nudity of those in the sauna will be exposed in public and people that they interact with in daily life will be able to see the pictures. The five will discover that they were nude in public. That will breach their principles of modesty and they will each suffer extreme emotional distress.
>
> Peter is outside the sauna and knows what is about to happen. He cannot turn off the camera or the internet feed any faster than it will turn off by itself. But he can push a sixth person, who is nude and waiting to go into the sauna, into the room. This person will be pushed right in front of the camera, obscuring the other five for the five seconds of internet feed and therefore the five people will not suffer the extreme emotional distress that would result from their being nude in public. However, the sixth person also has a strict principle of modesty and will suffer extreme emotional distress because he will be nude in public and people that he interacts with in daily life will see the pictures.
>
> None of the people will get any compensation because the sauna will go bankrupt due to the negative publicity.
>
> Peter is aware of all these facts.
>
> Thus, Peter can push the man, in which case the one person will suffer extreme emotional distress but the five people will not; or Peter can refrain from pushing the man, in which case the five people will each suffer extreme emotional distress but the one person will not.
>
> Is it morally wrong for Peter to push the man?

In the other harm scenarios, we replaced the last sentence of the first paragraph as follows:

job loss: However their jobs require that they be modest in public, so they will each lose their jobs as a result. Given their age and the economic conditions, they will not find new jobs.

financial loss: The five in the sauna do not care about modesty, or being nude on the internet per se, but each one will lose a contract worth £10 000 as a result of having been nude on the internet.

death: However the five are each members of a religious sect and extremists from the sect will object to the nudity and assassinate them as a result.

We made equivalent substitutions for ‘emotional distress’ in other paragraphs. In every scenario, the harm that threatened the five was the same as the harm that befalls the one. The full text of all the scenarios is available in the online supplementary materials.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

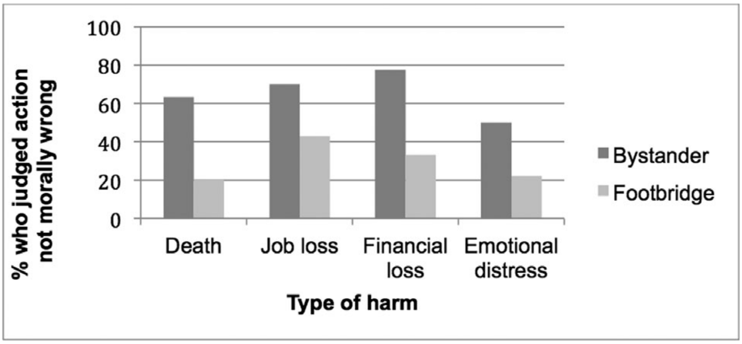

FIGURE 3. Results of Study 2: Percentage of participants in each condition who said 'No' it was not morally wrong for Peter to do the action.

Procedure. As in Study 1, participants completed the study in their own time, online, after following a link to a SurveyGizmo site. Participants read a consent form and were assured of the anonymity of their data. After granting consent they were randomly allocated to read only one of the eight scenarios. After reading the scenarios they were asked (a) Is it morally wrong for Peter to *push the man/press the switch*? (Yes/No), and (b) Should Peter *push the man/press the switch?* (Yes/No), (c) Now please indicate how wrong or how right you think it would be to *push the man/press the switch* (0 *Definitely wrong* to 10 *Definitely right*).

3.2 Results

Figure 3 shows the percentage of participants in each condition who said that 'No', it was not morally wrong for Peter to do the action. Our reasons for displaying the data this way are the same as those in Study 1.

For all four types of harm, there was a significant effect of the *bystander* versus *footbridge* scenarios on judgements of whether or not the action was morally wrong: *death*: χ^2 (1, 122) = 23.199, p < 0.001; *job loss*: χ^2 (1, 95) = 6.786, p = 0.009; *financial loss*: χ^2 (1, 103) = 20.405, p < 0.001; *emotional distress*: χ^2 (1, 108) = 9.030, p = 0.003. The overall pattern is that more participants in the *footbridge* than in the *bystander* conditions judge the action to be morally wrong.

Within the *bystander* condition, there was a significant effect of type of harm, χ^2 (3, 235) = 10.145, p = 0.017. This was because only 50% of participants judged that it was not morally wrong to push the switch and cause emotional distress, which was significantly fewer than the number of people in the *job loss* and *financial loss* scenarios who judged that it was not morally wrong to push the switch (see Table 2). Within the *footbridge*

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

measure	harm			
	death	job loss	financial loss	emotional distress
bystander				
morally wrong, % No	$64\%_{a,b}$	$70\%_{b}$	$78\%_{b}$	$50\%_{a}$
should, % Yes	$63\%_{a}$	$67\%_{a}$	$68\%_{a}$	$40\%_{b}$
right–wrong, mean	$5.02_{a,b}$	5.64_{a}	5.66_{a}	4.20_{b}
(*SD*)	(2.40)	(2.89)	(2.55)	(2.84)
footbridge				
morally wrong, % No	$20\%_{a,c}$	$43\%_{b}$	$33\%_{b,c}$	$22\%_{a,c}$
should, % Yes	$32\%_{a,b}$	$50\%_{a}$	$36\%_{a,b}$	$22\%_{b}$
right–wrong, mean	3.00_{a}	2.75_{a}	3.75_{a}	2.75_{a}
(*SD*)	(3.10)	(2.38)	(3.26)	(2.16)

Note: Percentages on the same row that do not share a common subscript differ at $p < 0.05$ according to a χ^2 test. Means on the same row that do not share a common subscript differ at $p < 0.05$ according to a post-hoc Tukey test.

TABLE 2. Results of Study 2: Percentages of 'no' responses to the question 'Is it morally wrong to do the action?', percentage of 'yes' responses to the question 'Should Peter do the action?' and mean right–wrong ratings.

condition there was a marginally significant effect of type of harm, χ^2 (3, 193) = 7.082, p = 0.069. This is because the percentage of participants who judged that it was not morally wrong to push the man into the room and cause him to lose his job was higher than the percentage of participants who judged that it was not morally wrong when pushing him would result in his death or emotional distress (although not significantly more than if it would just cause financial loss). The percentage of participants who judged that it was not morally wrong to harm one to save five when the harm was *emotional distress* was not significantly different from that when the outcome was *death*, in both the *footbridge* and *bystander* scenarios. The full results are shown in Table 2.

Again, the pattern of judgements about whether Peter should push the man closely followed that of the judgements about whether it would be morally wrong to do so. Within the *bystander* condition, there was a significant effect of type of harm, χ^2 (3, 235) = 10.145, p = 0.017. Pairwise comparisons show that this is because participants were less likely to judge that Peter should push the man when it caused emotional distress, which was also the only scenario in which the majority of participants judged that he should not push the man (see Table 2). Within the *footbridge* condition, there was a marginally significant effect of type of harm, χ^2 (3, 193) = 7.082, p = 0.069. In this condition, the majority of participants

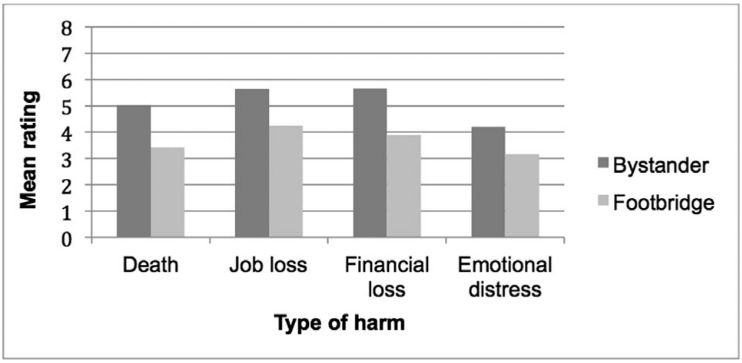

FIGURE 4. Results of Study 2: Mean ratings of how wrong or right it would be to take the action.

judged that Peter should push the man, but more people judged that he shouldn't push when it resulted in emotional distress rather than job loss (p = 0.008). Comparing the answers to the question about whether it is morally wrong to do the action with the answers about whether Peter should actually take the action, using McNemar tests for each of the eight scenarios, found no significant differences at all between these two questions, as in Study 1. So if people thought that the action was morally wrong then they subsequently said that Peter should not do it.

A two-way ANOVA on the 0 to 10 ratings of how wrong or how right taking the action would be shows that there was a significant difference between the four types of harm, F(3, 403) = 4.620, p = 0.003, partial η^2 = 0.03. Post hoc pairwise comparisons show that the ratings of how right or wrong it is to take the action in the *emotional distress* condition (M = 3.66) is significantly lower (it is seen as more wrong) than in the *financial* loss (M = 4.88) and *job loss* (M = 5.15) conditions, but is not rated significantly differently from ratings in the *death* condition (M = 4.23) – see Table 2. There is a significant difference between the ratings in the *bystander* and *footbridge* conditions overall, the action being rated as more right in the *bystander* condition, F(1, 403) = 30.215, $p < 0.001$, partial η^2 = 0.07. A test for simple effects shows that the differences between *bystander* and *footbridge* were statistically significant for all types of harm (all $p < 0.05$). There was no significant interaction between the type of harm and scenario, F(3, 403) = 0.377, p = 0.769. As is clear from Figure 4 there is a significant difference in the *bystander* ratings across type of harm as the *emotional distress* scenario is rated as wrong (M = 4.20), the *death* scenario is rated at the neutral point between right and wrong (M = 5.02) and both *death* and *emotional distress* are rated as significantly more wrong than *financial loss* (M = 5.66) or *job loss* (M = 5.64). The ratings across harms in

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

the *footbridge* condition do not differ significantly (F(3, 187) = 1.482, p = 0.22): they are all low and the actions are all rated as being morally wrong (see Table 2).

4. DISCUSSION OF THE STUDIES

The difference in moral intuitions between *bystander* and *footbridge* scenarios does not depend on the outcome being death or serious injury. The difference is robust across different domains and levels of physical and non-physical harm, and also across our two different dilemmas, the *trolley* and the *sauna*.

In Study 1, the difference in intuitions is sharply attenuated but not eliminated (it is still highly significant) when the loss is merely of property; the results for death and other physical injuries are largely similar to one another, and the results for death are comparable to those from other trolley studies (e.g. Mikhail 2007; Waldmann and Dieterich 2007). The attenuation is driven by the increased percentage of participants who judged that it would not be wrong to push the rucksack in *property-footbridge*, compared with the number who judged that it would not be wrong to push the man in the physical injuries scenarios.

In Study 2, the difference was replicated across a range of non-physical forms of harm. The percentage of participants who judged that it would not be wrong to switch the lever in *bystander* varied across the different harms, as did the percentage who said that it would not be wrong to push the man in *footbridge*, but the difference in responses between *bystander* and *footbridge* that is found in the standard trolley problem was always preserved.

In philosophy, the difference in moral intuitions between *bystander* and *footbridge* is deployed most often in support of the *doctrine of double effect*, that it is sometimes worse to intend a harm than to produce that same harm as a foreseen side effect, when intending something else. In *bystander* the agent's aim is to divert the train and it is an unfortunate side effect that there is a person on the side-track. In contrast, in *footbridge* the death of the man on the bridge is intended as a necessary part of the plan to stop the train: if the man on the bridge does not die then the train will not stop. The cases have also been discussed with respect to the *doctrine of doing and allowing*, that 'it is often more acceptable to let a certain harm befall someone than actively to bring the harm about' (Quinn 1989*b*: 334). Although there is a presumption that intuitions about deaths will generalize, in both these doctrines there is a qualifier, that they are 'sometimes' or 'often' operative. Hence, even if the difference in intuitions did not generalize across domains and

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

levels of harm, that would not decisively refute the principles. However, intuitions about trolley problems with different domains and levels of harm are still of interest because they are relevant for determining the scope of the principles and the extent to which they are operative. Our data is encouraging for people who deploy these intuitions because it suggests that they are fairly general across domains and levels of harm, with the difference between *bystander* and *footbridge* being preserved.

Our experiment was not intended to discriminate between different psychological theories and, because of the design of our scenarios, our results are consistent with theories that depend on the difference between pressing switches to divert trains and pushing people into their path, such as Greene *et al.*'s (2009) theory of the interaction between intention and personal force, Waldmann and Dieterich's (2007) theory of agent versus patient intervention, and Mikhail's (2007, 2011) theory of a Universal Moral Grammar. However, the theories may differ in how easily they can accommodate the results from the non-lethal dilemmas, especially the data on property damage.

The theories of Greene *et al*. (2009) and Waldmann and Dieterich (2007) were both developed to explain results from life or death scenarios. Waldman and Dieterich include the paraplegia case but Greene *et al*. give no indication as to whether or not their theory is meant to be domain specific. Neither of these theories addresses the case where it is an item of property that is being acted on, rather than a person. However, it is easy to see how the theories could be extended to include bodily harm and even property damage, by allowing pushes and interventions to be performed on inanimate objects.

In contrast, Mikhail's (2011) theory of a Universal Moral Grammar is concerned with explaining near-universal prohibitions on homicide and causing bodily harm (133). Hence, according to the Universal Moral Grammar theory, it should not make a difference to people's moral intuitions if death is replaced with bodily injury. Our results support that claim. In the *property* scenarios, there is no battery or contact with a person, so Mikhail's theory makes no predictions and, unlike the other psychological theories discussed above, the emphasis on battery (rather than just the principle of double effect) means that it is hard to see how the Universal Moral Grammar theory could be extended to explain intuitions in the *property* scenario.

Of course, this does not necessarily count against the theory. One can think of principled reasons why we might wish to distinguish the *property* case from the other *trolley* scenarios. For instance a standard objection to utilitarianism is that it ignores the 'separateness of persons' (Rawls, 1971), that is, it allows one person to be made worse off in order to provide benefits for others. Arguably, this objection is more plausible in

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

the domain of bodily injury than in the domain of relatively small losses of property.[4]

Further, although the *property* scenarios did exhibit the standard difference between *bystander* and *footbridge*, participants were much less likely to judge it morally wrong to push the rucksack in the *property-footbridge* scenario than in the other *footbridge* scenarios, and it was also the only *footbridge* scenario where the action was rated on the 'right' side of the right–wrong scale. So our results could support an argument that property damage belongs in a different moral domain from bodily harm and death.

In the sauna scenarios, *job loss* and *financial loss* are treated equivalently, and they are treated differently from *emotional distress.* A job's main function is arguably to bring in money so, in theory, a job loss could be compensated for by a sum of money equal to lifetime expected earnings. Hence it may not be surprising that our participants treated the *job loss* and the *financial loss* scenarios equivalently. This idea places both the scenarios firmly in the economic domain. In contrast, emotional distress seems like a very personal sort of harm. It is plausible that moral judgements differ in the economic and the personal domains, as moral thinking may be different when it involves economic rather than personal relationships (Rai and Fiske 2011).[5]

Death and *emotional distress* are also treated equivalently in the *sauna* scenarios. In the *trolley* study, we found that bodily harms are treated equivalently to death. It may be that emotional distress is perceived as being a type of bodily harm. After all, emotions typically have physical effects. Thus, causing emotional distress may be perceived as a very negative thing to do, even worse than causing someone to lose their job or a large amount of money.

Alternatively, it may be that underlying our moral judgements in trolley problems there is a trade-off between the magnitude of the harm prevented and the strength of the one person's right not to be harmed, so the wrongness of acting depends on their relative magnitudes. Acting in a trolley problem might be considered wrong if the right of the one outweighed the magnitude of the harm prevented. When the harm is death, the harm prevented is very great but the right of the victim not to be harmed may be very strong. When the harm is emotional distress, both the harm and the right may be correspondingly weaker. Then, in order to produce the pattern of moral intuitions that we found when the outcome was job loss or financial loss, in those two scenarios the decrease in harm (that surely occurs compared with when the outcome is death), must be relatively small compared to the weakening of the

[4] We thank Christian List for bringing this point to our attention.

[5] We thank Armin Schulz for suggesting the two points in this paragraph to us.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

right not to be harmed, which must be relatively large. Furthermore, compared with the emotional distress conditions, the increase in harm must be relatively large but the strengthening of the right relatively small. Some philosophers have seen trolley problems as embodying a conflict between rights (Foot, 1967; Quinn, 1989*a*, 1989*b*). However, whether philosophical theories can allow trade-offs between harms and rights is another question; at least some theories do not (e.g. Nozick, 1974).

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/doi:10.1017/S0266267113000205

REFERENCES

Andreoni, J., P. M. Brown and L. Vesterlund. 2002. What makes an allocation fair? Some experimental evidence. *Games and Economic Behavior* 40: 1–24.

Andreoni, J. and J. Miller. 2002. Giving according to GARP: an experimental test of the consistency of preferences for altruism. *Econometrica* 70: 737–753.

Berg, J., J. Dickhaut and K. McCabe. 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10: 122–142.

Bolton, G. and A. Ockenfels. 2000. ERC: a theory of equity, reciprocity, and competition. *American Economic Review* 90: 166–193.

Cushman, F., L. Young and M. Hauser. 2006. The role of conscious reasoning and intuition in moral judgment: testing three principles of harm. *Psychological Science* 17: 1082–1089.

Fehr, E. and K. Schmidt. 1999. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics* 114: 817–868.

Fehr, E. and K. Schmidt. 2006. The economics of fairness, reciprocity and altruism: experimental evidence. *Handbook of the Economics of Giving, Altruism and Reciprocity, Vol. 1*. Amsterdam: North-Holland/Elsevier.

Foot, P. 1967. The problem of abortion and the doctrine of double effect. *Oxford Review* 5: 5–15.

Greene, J. D., R. B. Sommerville, L. E. Nystrom, J. M. Darley and J. D. Cohen. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537): 2105–2108.

Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley and J. D. Cohen. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* 44: 389–400.

Greene, J. D., S. A. Morelli, K. Lowenberg, L. W. Nystrom and J. D. Cohen. 2008. Cognitive load selectively interferes with utilitarian moral judgment. *Cognition* 107: 1144–1154.

Greene, J. D., F. A. Cushman, L. E. Stewart, K. Lowenberg, L. E. Nystrom and J. D. Cohen. 2009. Pushing moral buttons: the interaction between personal force and intention in moral judgment. *Cognition*, 111: 364–371.

Kahneman, D. and A. Tversky. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47: 263–292.

Kamm, F. M. 1985. Equal treatment and equal chances. *Philosophy and Public Affairs* 14: 177–194.

Koenigs, M., L. Young, R. Adolphs, D. Tranel, F. Cushman, M. Hauser and A. Damasio. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature* 446: 908–911.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press

Mikhail, J. 2007. Universal moral grammar: theory, evidence and the future. *Trends in Cognitive Sciences* 11: 143–152.

Mikhail, J. 2011. *Elements of Moral Cognition.* Cambridge: Cambridge University Press.

Moore, A. B., B. A. Clark and M. J. Kane. 2008. Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science* 19: 549–557.

Nichols, S. and R. Mallon. 2006. Moral dilemmas and moral rules. *Cognition* 100: 530–542.

Nozick, R. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

O'Hara, R., W. Sinnott-Armstrong and N. A. Sinnott-Armstrong. 2010. Wording effects in moral judgments. *Judgment and Decision Making* 5: 547–554.

Petrinovich, L., P. O'Neill and M. J. Jorgensen. 1993. An empirical study of moral intuitions: toward an evolutionary ethics. *Journal of Personality and Social Psychology* 64: 467–478.

Rai, T. and A. Fiske. 2011. Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review* 118: 57–75.

Quinn, W. S. 1989*a*. Actions, intentions, and consequences: the doctrine of doing and allowing. *The Philosophical Review* 98: 287–312.

Quinn, W. S. 1989*b*. Actions, intentions, and consequences: the doctrine of double effect. *Philosophy and Public Affairs* 18: 334–351.

Rawls, J. 1971. *A Theory of Justice.* Cambridge, MA: Harvard University Press.

Sinnott-Armstrong, W., R. Mallon, T. McCoy and J. G. Hull. 2008. Intention, temporal order, and moral judgments. *Mind and Language* 23: 90–106.

Thomson, J. J. 1985. The trolley problem. *Yale Law Journal* 94: 1395–1415.

Unger, P. 1996. *Living High and Letting Die: Our Illusions of Innocence*. Oxford: Oxford University Press.

Utikal, V. and Fischbacher, U. 2009. On the attribution of externalities. *TWI Research Paper Series 46, Thurgau Institute of Economics*.

Waldmann, M. R. and J. H. Dieterich. 2007. Throwing a bomb on a person versus throwing a person on a bomb. *Psychological Science* 18: 247–253.

https://doi.org/10.1017/S0266267113000205 Published online by Cambridge University Press