Judgment and Decision Making, Vol. 6, No. 7, October 2011, pp. 593–601

Trolley problems in context

Christopher Shallow* Rumen Iliev† Douglas Medin*

Abstract

Would you redirect a trolley to save five people even if it means that the trolley will run over a person on the side track? Most people say they would. Would you push that same person into the path of the trolley in order to save the five? Most people say they would not. These sorts of intuitive moral judgments are made rapidly and seem almost automatic. Now imagine a combined choice context where one can redirect a trolley, push a person in its path or do nothing. The number of lives lost from intervening can be varied. The most straightforward interpretations of current theories of moral judgment predict either no effect or that the combined context will lead to greater focus on lives lost. In contrast, we observe a similarity effect such that utilitarian choice may become less attractive in the combined choice context.

Keywords: context effects, trolley problem, moral dilemma, moral decision making, similarity effect.

1 Introduction

Moral judgments are often contrasted with mundane or secular decision making, such as consumer choice. The latter predominantly involves weighing costs and benefits, which often may not be the preferred strategy in moral situations. In the now classic trolley problem (Foot, 1978; Thomson, 1985; Spranca et al., 1991; Petrinovich et al., 1993; Greene et al., 2001) a runaway trolley threatens to run over and kill five people trapped on a railroad track. If a switch is flipped, the trolley will move onto another track, saving the five but killing a person trapped on that track. In the footbridge version the only way to save the five is to push a large person off a bridge to stop the trolley. Most people approve flipping the switch, but disapprove of pushing the person, patterns so consistent that some have claimed they are universal (Hauser, 2006; Hauser et al., 2007, Banerjee et al., 2011).

The reluctance to push the person is often described as an example of deontological decision making, where the focus is on intrinsic goodness or badness of actions per se rather than on instrumental outcomes. The asymmetry between switch and pushing scenarios has been used as an effective tool for developing and evaluating theories of moral judgment (e.g., Hauser, 2006; Mikhail, 2007, 2009). One popular distinction is the doctrine of double effect, which differentiates between harm caused as means and harm caused as a side effect. Harm as means, such as when the actor uses the body of a single victim to prevent the death of a larger number of people, is considered impermissible. Harm as a side effect, however, such as when the trolley kills the single victim after redirected on the sidetrack, is considered permissible. Other relevant distinctions are between direct and indirect harm (Royzman & Baron, 2002), and intervention on the victim versus intervention on the harmful object (Waldmann & Diterich, 2007). A number of other studies have varied a plethora of other factors associated with the action, such as spatial distance (Greene et al., 2009), physical contact between actor and victim (Cushman et al., 2006), temporal order of events Sinnott-Armstrong et al., 2008), further advancing our understanding of the role of the properties of the intervention on moral choice.

Although most of the research on trolley problems has focused on factors associated with the intervention, others have looked at the role of utility of the outcomes. For example, sacrificing a stranger is approved more than sacrificing a relative, but sacrificing an obnoxious person (a NAZI soldier) is approved more than sacrificing the stranger (Petrinovic et al., 1993). Similarly, the lives of in-group members are valued more than the lives of outgroup members (Swann et al., 2010). An additional indicator of sensitivity to outcomes is that, when the number of lives at stake increases, the approval for intervention to save these lives increases too (Bartels, 2008; Mikhail, 2009). Finally, trolley dilemmas have been shown to be influenced by framing of the experimental question. Positive framing, focusing on attention to number of lives saved by the intervention, leads to greater approval for action than negative framing, which focuses it on number of lives lost as a result of the intervention (Petrinovic & O'Neill, 1996).

This work has been supported by Northwestern University, AFOSR grant FA9550–09-C-0119 and NSF grants SES-0527396 and SES-0962185 to the third author. Some corrections in p-levels were made on Jan. 23, 2014. These did not affect the conclusions.

*Northwestern University

†Corresponding author: Department of Psychology, Northwestern University, 2029 Sheridan Road – Swift Hall 219, Evanston, Il 60208. Email: r-iliev@northwestern.edu.

Our study is concerned with context effects based on combining the two classic scenarios into one where the actor can throw a switch to redirect the trolley, push a person off the footbridge or do nothing. As in the standard case omission leads to five deaths, pushing causes one death and we vary the number of deaths associated with redirecting the trolley. Current moral theories typically do not make explicit predictions about extended context comparisons though the most straightforward interpretations suggest that a combined contexts increase attention to outcomes. For example, Mikhail (2009) proposed a switch version with two side-tracks, on one of which there is a person, and the other is empty. Under this condition redirecting the trolley onto the track with the single person on it becomes morally prohibited. In our study, however we varied both the action and the outcomes in such a way that no choice transparently dominated any other choice.

Specifically, in one of our combined contexts the actor could push a large person off a footbridge, sacrificing one person to save five, throw a switch to redirect the trolley away from the five but putting it on a side track where two people will die, or do nothing, in which case five people die. Pushing a person is less desirable than throwing a switch (all else equal) but killing two people is less desirable than killing one. If the combined choice context increases attention to outcomes, then by Mikhail's analysis the action of throwing the switch should be rated more negatively, because there is an option available leading to fewer deaths. For the same reason the act of pushing one person to save five should, if anything, become relatively more desirable. Alternatively, the action of pushing may be seen as always wrong, so the disapproval of pushing would not be affected by context.

We collected pilot data on combined contexts that appeared to undermine these seemingly straightforward predictions. In particular, the presence of the switch option led to less approval of pushing the person, even when throwing the switch led to more deaths. This led us to consider a different form of choice context effect, a *similarity effect*. We know of no theory of moral judgment that predicts similarity effects. In what follows we first describe similarity effects and closely related context effects and then turn to a study aimed at examining choice context effects.

Choice context effects. The challenges that moral psychology faces when extrapolating results from binary moral dilemmas to broader contexts shows parallels with the development of choice theory (Luce, 1959). Suppose that we are interested in how much consumers like 10 prospective car models. One way is to ask them how much they are willing to pay for a given model. A shortcoming of this method is that it lacks a reference point to which a person compares the item, so in many cases such answers will not be very informative. Another way to try to capture consumer choices is simply to use binary comparisons where consumers have to choose the more preferred car from all possible pairwise comparisons. We can transform the choices into some preference metric which will tell us the relative ranking of each car compared to the rest. Further, since such scale will assume transitivity, we can even rely on partial information, where if we do not have data from a direct comparison between cars A and C, we still can infer what people will choose based on the other comparisons in which these two cars participated (but see Tversky, 1969 for possible violations of transitivity).

These early promising attempts to model choice behavior were quickly undermined by new findings from a wider range of paradigms (Huber & Puto, 1983). Depending on the particular configuration of the choice set the addition of a new option has been found to systematically change the preferences. Several context effects have been identified (see Busemeyer et al., 2007 for a recent review); for present purposes, we focus on similarity and compromise effects. In addition, we consider a context effect associated with separate versus joint evaluation, the so-called Evaluability Hypothesis (Hsee, 1996).

Similarity, compromise and evaluability. Imagine a buyer has to choose between a Toyota Prius and a Ford Focus, and all that she cares about is price and fuel economy. The Toyota Prius gets better gas mileage, but the Ford Focus compensates by having a lower price. Further, imagine that in our hypothetical scenario she finds herself roughly indifferent (50–50) between the two options. Early work on rational choice theory recognized that, when additional options become available, the new options might well attract choices, but the theories did assume that the relative preference between the initial set, Prius and the Focus in our example, would remain the same.

Many empirical studies, however, have established that people often violate this independence principle. With the "similarity effect" (Tversky, 1972; Roe et al., 2001), adding a non-dominated option close to one of the alternatives tends to increase the relative share of its competitor. For example, if we add a Honda Insight to the choice set, slightly worse than the Prius on fuel efficiency but slightly better on price, the buyer who was previously indifferent about the Prius and Focus options typically shifts to preferring the Focus over the Prius. Informally, we could say that the Insight steals more choices from the Prius than from the Focus. Alternatively (see Figure

1), imagine adding a Honda Civic to the choice set rather than the Honda Insight and assume that the Civic is more similar to the Focus, being slightly better on gas mileage, but slightly more expensive. Now the relative choice between Prius and Focus should shift towards the Prius as the Civic competes more with the Focus than with the Prius. We refer to this choice context effects as a similarity effect.

In addition to the similarity effect, two other types of context effect are potentially relevant to the current work. The first is the compromise effect, whereby introducing a third option that results in one of the first two being seen as a compromise leads to an increase in choices of the compromise option (Simonson, 1989; Simonson & Tversky, 1992; Tversky & Simonson, 1993). Continuing with the car example, if we add as a third option the Nissan Leaf, for example, which has a much better gas mileage than the Prius, but also is significantly more expensive, then the Prius will become a compromise, and studies show that being in the middle often increases consumer preferences (e.g., Tversky & Simonson, 1993). Conversely, adding a cheaper, less fuel efficient option to the Prius versus Focus choice set may make the Focus a compromise and lead to it being favored.

How might compromise effects work in our hypothetical scenario? Let's begin with our example of throwing the switch to save five people but leading to the death of two people versus doing nothing. Adding the option of pushing one person to save five could make throwing the switch a compromise between saving more versus fewer lives and, all else equal avoiding actions leading to any deaths. This leads to the counter-intuitive notion that adding the push option may make throwing the switch more desirable.

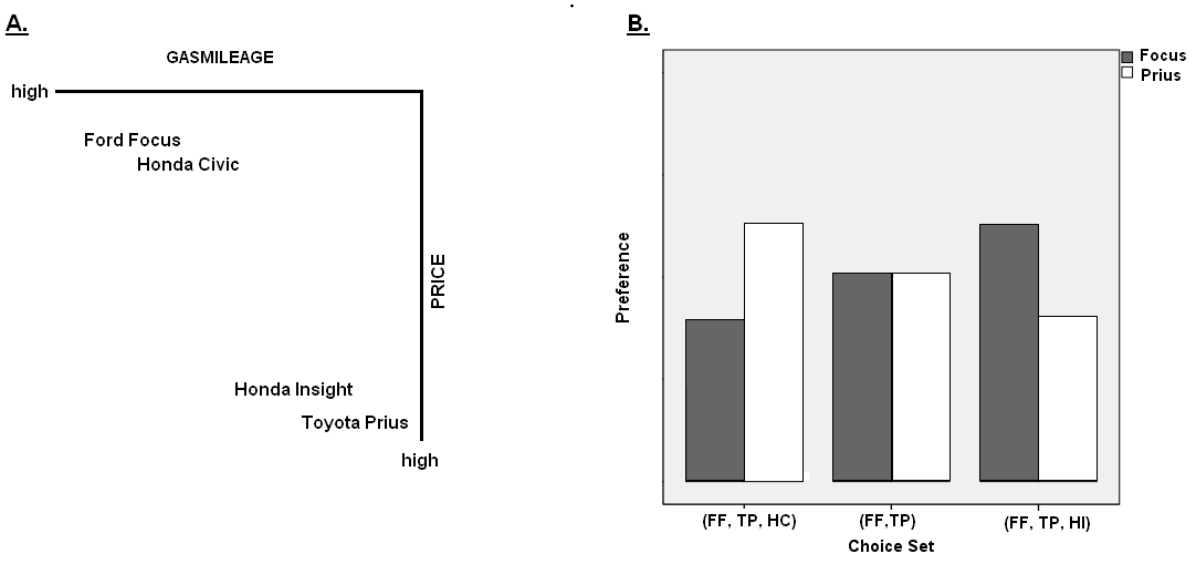
The final context effect we will mention stems from the work on separate versus joint evaluation. Imagine you have to assign a value to two dictionaries, one in very good condition and containing 10,000 words, and the other in fair condition but with 20,000 words. Hsee (1996) found that the smaller dictionary is priced higher than the larger one when each is evaluated in isolation (separate evaluation), but the smaller dictionary received a lower pricing than the larger one in joint evaluation. According to the evaluability hypothesis, number of words is not very meaningful piece of information in separate evaluation because there is no reference point for comparison. As a result, number of words becomes a more important dimension in joint, rather than in separate evaluation (see also Bazerman et al., 1998 and Paharia et al., 2009). In the same way in our scenario where throwing a switch leads to two people dying but five being saved, the evaluability may shift in the direction of outcome utility when the option of pushing a single person to save five is added. This would lead to the throwing switch option becoming less desirable (again see Mikhail, 2009).

Are moral choices equally susceptible to these choice-set context effects? Our initial hunch was that instrumental choices would be based on evaluability. But we also considered the possibility that options evoking deontological rules might not be susceptible to context effects. For example, an option saving three lives might be seen as quite positive by itself, but lose its value when there is an option available that would save four lives. But an inappropriate or immoral act, like pushing a person off a bridge, may remain immoral or equally bad, even when instrumentally less desirable options are added.

The trolley and footbridge problems can be conceptualized as comprised of a choice set having two dimensions, one representing the value or utility of the number of people saved or lost and the other representing the value or cost intrinsic to the actions themselves, including inaction. This latter dimension is useful for the basic distinction or difference between throwing the switch (acceptable) versus pushing a person (unacceptable) when instrumental outcomes are the same. In addition, Baron and his collaborators (Ritov & Baron, 1990; Spranca, et al., 1991; Baron & Ritov, 1994) have shown that, when harmful outcomes are at stake, people often choose to do nothing, showing an omission bias. For example, one would expect that, given a choice between doing nothing and one person dies versus throwing a switch and one person dies, people will show a strong preference to do nothing. Combining these observations, on the dimension of forms of action leading to harm in this context omission is preferable to throwing the switch which in turn is greatly preferable to pushing a person. And on the dimension of lives lost, less is better. These two hypothetical dimensions set of stage for examining context effects.

In the next section we present an empirical study which compares preferences for action and omission in five different choice sets. The key comparison is between two three-alternative choice-sets (trilemmas), which have the same omission and footbridge-intervention options but differ in the number of victims for the switch intervention. If the approvals for the two shared options (pushing and omission) differ as a function of the properties of the third option, this will demonstrate a context effect, which may take the form of a similarity effect, a compromise effect or an evaluability effect. To determine the form of choice context effect ratings will be compared with two-alternative versions of these scenarios. The predictions of the different theories of context effects are summarized after the description of the design.

Figure 1: Representation of a hypothetical similarity effect. In panel A, cars are ranked on their fuel efficiency and price. In panel B, the relative preference between two cars depends on the presence of a third option. The novel option reduces the share of the more similar alternative



2 Empirical study

2.1 Method

2.1.1 Participants

One hundreds sixty eight Northwestern University undergraduate students participated in the experiment in exchange for a course credit.

2.1.2 Stimuli

As in the original trolley problem, we used a scenario which described a runaway trolley which is about to kill five innocent people. Instead of directly asking if a person should intervene, we asked participants (in a between participant design) to evaluate subsets of 4 different alternatives (the number of the option also indicates the number of victims dying as a result):

1. Push intervention: The five can be saved by pushing a single person off a bridge (but 1 person dies).

2. Switch intervention: The five can be saved by redirecting the trolley onto a different track (but 2 people die).

4. Switch intervention: The five can be saved by redirecting the trolley onto a different track (but 4 people die).

5. Omission: Doing nothing. (No deaths are directly caused but 5 people die).

Notice that in terms of lives lost option 1 (push) is the best, and option 5 (omission) is the worst. In terms of negative utility associated with intervening, however, option 5 is best, while option 1 is worst and 2 and 4 are intermediate. By manipulating the numbers of lives lost in the switch intervention we can create settings for different context effects. For example, in this two-dimensional space option 2 is more similar to option 1, while option 4 is more similar to option 5, allowing us to evaluate similarity effects.

Table 1: The five different choice sets with the number of victims associated with each alternative. The first digit in the notation shows the number of victims from the push intervention, the second from flipping the switch, and the last one from doing nothing.

Notation	Push	Switch	Omission
1_5	1	n/a	5
_25	n/a	2	5
_45	n/a	4	5
145	1	4	5
125	1	2	5

2.1.3 Design and procedure

In five between-subject conditions participants rated the acceptability of different combinations of push-intervention and/or switch-intervention and omission. The push intervention always resulted in one person dying, the omission option always resulted in five people

dying, and there were two switch options, one leading to two deaths and one to four deaths. These options were combined in five choice sets, three two-alternative sets contrasting intervention and omission, and two three-alternative sets, combining two different interventions and omission. The five sets are summarized in Table 1.

For each alternative the subjects were asked to answer the following question: "Do you approve or disapprove of John's action?", indicating their approval on a 6-point scale, where −3 was "Disapprove", and +3 was "Approve". We omitted the zero from this scale in an effort to force participants to choose between approval and disapproval. For purposes of analysis we assume that we had a 7-point scale but that no one gave a neutral rating (none of our results hinge on this assumption). The critical question is whether the context of a third option affects approval or disapproval of other options (including inaction) and the form of any observed context effects.

2.2 Summary of predictions

First, the notion of evaluability suggests that throwing the switch will be much less desirable when an option that leads to fewer deaths is available (see also Mikhail, 2009). Evaluability, by itself, predicts that pushing the person will become more desirable in the combined contexts because it saves more lives. In short, in the combined contexts number of lives saved should have a larger effect on judgment than in the two choice contexts.[1] A strong deontological position, however, would suggest that pushing a person is wrong, no matter what its utilitarian value, so we might not expect different ratings for the push intervention in the different contexts.

Although the deontological position and the evaluability hypothesis predict either no difference, or consistent differences favoring one dimension, the similarity effect predicts that the relative preference for the omission intervention will depend not only on the presence of the switch option, but also on the particular number of people sacrificed by flipping the switch. More specifically, when the switch intervention sacrifices two people it is more similar to push intervention and approval for pushing relative to omission should decrease. However, when the switch scenarios sacrifices four people, it is more similar to omission, and preference for omission relative to pushing should decrease. The same logic holds for valuation of throwing the switch versus doing nothing when the pushing option is added.

[1]Technically this prediction hinges on which dimension is made salient in joint evaluation. For example, it is possible that directly juxtaposing switch and push interventions will serve to make the properties of the action itself more salient, leading to stronger aversion to the push intervention in both trilemmas than in dilemmas. Given that the aversion to pushing a person in the standard footbridge case has been claimed to be a universal, this hypothesis seems implausible.

Lastly consider the compromise effect. For the "middle alternative" we vary only the number of lives lost, such that either of the switch options may be perceived as a compromise in the three-alternative context. If so, compromise effects would predict more approval for throwing the switch in trilemmas than in dilemmas.

2.3 Results

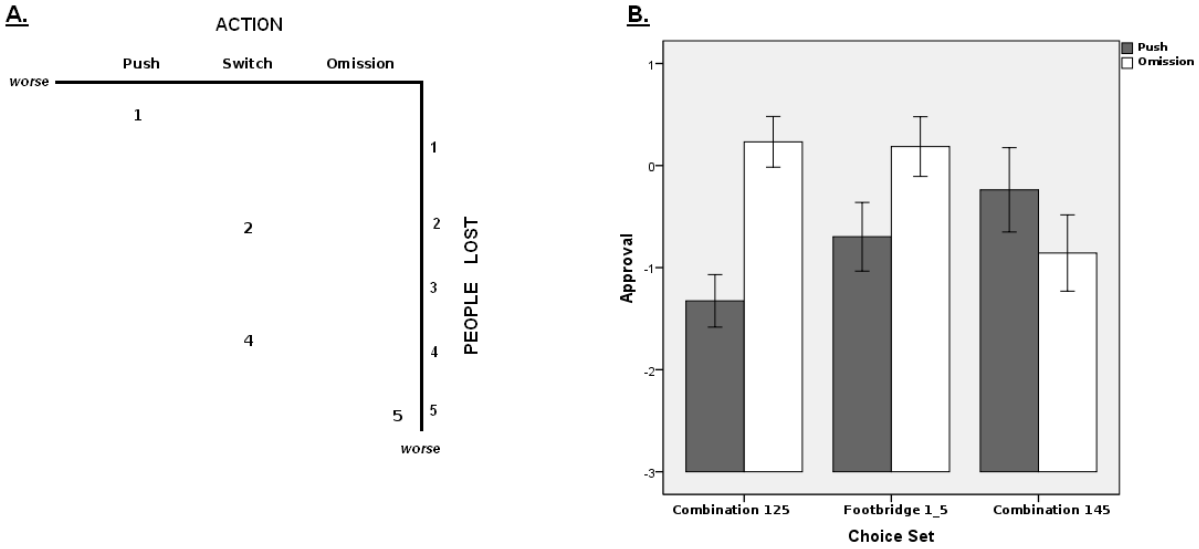
Two option ratings. In the two-alternative versions (1_5) and (_25) we replicated the classical trolley effect, where the intervention of throwing the switch was preferred in the trolley version (_25) more than pushing the person in the footbridge version (1_5) ($t(83) = 4.03$, $p<.0001$, one tailed,[2] $\eta^2 = .16$). Also, as expected, the switch intervention received positive ratings ($m_{switch2} = 1.07$) and the pushing in the footbridge case was evaluated negatively ($m_{push} = -.70$). The approval for the switch intervention in the four-person trolley version (_45) was also higher than for the footbridge version ($m_{switch4} = .42$, $t(60) = 1.98$, $p = .025$, $\eta^2 = .06$), and was not significantly different ($p = 0.07$) from the two-person switch intervention (_25). In other words, in the two choice scenarios the nature of the action affected ratings more than the number of lives saved.

Context effects. The critical comparisons involve the approval ratings in the trilemma sets. First we analyzed the relationship between approval ratings for action (pushing) and omission in the footbridge scenario. The two approval ratings were used as a within-subject factor in a 2x3 mixed-design ANOVA, where the three between-subject conditions were the choice context (1_5, 125, 145). There was a main effect of type of choice, approval being higher for omission than for action ($F(1,104) = 3.23$, $p = .04$, $\eta^2 = .03$). More relevant, however, was a significant interaction effect ($F(2,104) = 3.06$, $p = .026$, $\eta^2 = .06$). This reflected changes for both the omission and the pushing.

Although pushing was always disapproved in an absolute sense, relative to the two choice context, it was disapproved of more in the 125 context and disapproved of less in the 145 context. The higher approval for omission than for pushing actually was reversed in the 125 condition, with the change in approval being statistically reliable both for pushing ($t(62) = 2.33$, $p = .02$, two-tailed, $\eta^2 = .08$) and for omission ($t(62) = -2.47$, $p=.02$, $\eta^2 = .09$, two-tailed, see Figure 2B) This pattern of ratings is consistent with a similarity effect, where adding a third alternative (a switch intervention) decreases the approval for the option that it is most similar to. Note that this context effect includes the pushing option, suggesting that if

[2]All statistical tests are one-tailed unless noted otherwise.

Figure 2: Spatial representation of the choice alternatives and approval ratings for pushing and omission choices. Panel A shows the hypothetical utilities associated with each of the four choices, where the y-axis represents the cost of a particular action, while the x-axis represent the utility associated with the outcome. Panel B shows the approval ratings for the Push and Omission alternatives in three different types of choice sets. In the three-alternative sets, the relationship between pushing and omission was dependent on the positioning of the third option, consistent with similarity effect.



pushing triggers a deontological rule, the disapproval associated with that rule is itself subject to a context effect.

A further prediction from the similarity effect is that the switch-alternative will be ranked lower in the three alternative (125) than in the two alternative (_25) scenario, because the footbridge intervention is more similar to the switch intervention. This prediction was also supported by the data, the approval for the switch intervention in the trilemma dropped well into the disapproval range: m = $-.95$ (t(83) = 5.57, p<.001, η^2 = .27). The same drop in approval was observed for approval for the omission options. While approval for omission in the footbridge version was m = .19, adding a switch-option that was more similar to the omission decreased the approval, m = $-.86$, t(62) = 2.11, p=.02, η^2 = .07. The detailed results are presented in Table 2.

Overall the data are consistent with a similarity effect. In contrast, there was no consistent evidence for either compromise effects or the evaluability hypothesis. They predict only a main effect of context, rather than an interaction of type of context. First, with respect to compromise effects, in both cases approval for the middle option actually decreased. There was also no evidence for the evaluablity hypothesis. Evaluating the switch and push options together should highlight the utility advantage of the push version over the switch in both trillemmas, so disapproval for push intervention should be higher in 1_5 relative to both 125 and 145.[3] The results, however, showed that that the approval for pushing depends on how close is the switch options in terms of number of victims, not just on the presence or absence of another alternative.

To summarize, we found reliable context effects, where approval for different options is influenced by the properties of the choice set. The pattern observed was most consistent with similarity effect, where adding a non-dominated alternative decreases the approval of the most similar competitor. There was no evidence for compromise effect or the evaluability hypothesis.

3 Discussion

The central question of this study is whether choice context effects are observed for moral dilemmas and, if they are, what form they take. Our reading of current moral choice theories is that the only form of context effect they address are evaluability effects, where the presence of better options can reduce or even reverse approval for actions that would otherwise be approved. In contrast we found a different form of context effect. Adding a third option to a binary choice set selectively interferes with

[3] As we noted in Footnote 1 we think it is implausible that the combined context should highlight the valuation of the actions themselves. In any event, however, this idea also leads only to a main effect, not the interaction we observed.

Table 2: The mean approval and standard error (in parenthesis) for each of the alternatives in the five conditions. The percentage represents the proportion of rankings that were on the positive side of the scale. An asterisk signifies that the mean or the percentage is significantly different at (p<.05) from the mean in the cell below.

Condition	N	Push	Switch	Omission
Footbridge 1_5	43	−0.70 (0.34) 26%		0.19 (0.29) 58%
Switch _25	42		1.07 (0.28) 79%	−0.40 (0.31) 43%
Switch _45	19		0.42 (0.37) 68%	0.37* (0.36) 63%
Combination 145	21	−0.24* (0.41) 38%*	−0.57 (0.39) 38%	−0.86* (0.37) 33%*
Combination 125	43	−1.33 (0.26) 12%	−0.95 (0.23) 23%	0.23 (0.25) 74%

the approval rating of the closest alternative, a pattern consistent with similarity effect.

Although we are not aware about other research focused on choice-set context effects in moral dilemmas, the pattern we observed is consistent with a study by Huebner and Hauser (2010). In one of their experiments the authors presented two groups of subjects with either two or three alternative switch versions of the trolley problem. The two-alternative version was a variation of the classical problem, where the choice is between flipping a switch and doing nothing. In the three-alternative setting there were two identical sidetracks with one person on each of them. Adding another sidetrack to the classic version led to fewer people choosing action (they used choice rather than approval ratings). Since the two switch alternatives are highly similar to each other, based on a similarity effect we might expect that approval for either of the actions would decrease.[4]

A topic for future analysis is the "utility of the action" assumption that was essential for our experimental design. In its simple form the consequentialism-deontology debate the actions or means to achieve an outcome, and the consequences, or the value of the particular outcome, are kept as separate factors. Morally relevant actions or interventions are classified in basic categories such as obligatory, permissible or forbidden. Flipping the switch in the trolley scenario, for example is often considered permissible, while pushing the person forbidden. Notice, however, that the action by itself is ascribed no utility; instead typically the utility is associated with the magnitude and valence of the outcome which the action produces. Although this might be normatively defensible, we find such a strict distinctions of actions and outcomes to be psychologically implausible. First, even though participants might consider two harmful actions as morally forbidden, they can easily compare their badness (Wolfgang et al., 1985). Second, presenting a graphical description of an aversive action (such as smothering a crying baby) leads to stronger disapproval for the action when the outcome (number of lives saved) is kept the same. Similarly, when the action is kept the same, but the positive outcome of the action is increased participants are more likely to intervene. Instead of treating actions as categorical variables, people appear to take into account factors such as "effort and potential regret" (Baron, 1994) and compare them to the other properties of the final outcome. From this perspective, the degree of commensurability between action and outcome may be a useful factor when trying to explain preferences in extended choice sets. At a minimum, our data on context effects for deontogically forbidden actions (pushing a person to their death) undermine the idea that actions and outcomes are segregated in judgment.

On a broader theoretical level, our results add to a series of recent findings which emphasize the dynamic nature of moral judgments (e.g., Iliev et al., 2009). An number of recent findings demonstrate that moral judgments are complex and sensitive to subtle factors that are outside of the scope of abstract moral principles The large body of work on non-moral decisions and, for example, consumer choice, may be able to provide valuable insights for future work on the dynamics of moral decision making.

4 Conclusion

The majority of previous studies on moral dilemmas have been focused on isolating the factors that guide our distinction between right and wrong. Yet, how these factors are combined in a more complex situation is little understood. In this paper we suggest that previous work on context effects in consumer choice provides a useful methodology to study how alternative interventions are evaluated in a single scenario. Adding a new alternative to the choice set interfered with the evaluation of the available options in a pattern largely consistent with sim-

[4]Of course, there may be other interpretations of this effect and some would argue that the switch options are identical rather than highly similar.

ilarity effects. These results have implications for current theories of moral decision judgments which largely make their predictions based on a between-subject or between-scenario studies. What is unique about moral cognition, and the contexts in which it is unique is in need of further investigation (see Rai & Holyoak, 2010; Bennis et al., 2010; Iliev et al., 2009). Finally, our findings also have implications for mathematical models of context effects (Usher & McClelland, 2004; Busemeyer et al., 2005), revealing that not only outcomes, but also properties of the action can have utilities associated with them.

To our knowledge, theories of moral reasoning do not address these sorts of context effects. The present findings suggest that if trolley type judgments are based on a moral grammar, that grammar must be sensitive to context in a way that goes beyond evaluability considerations. The observed pattern of ratings closely followed previous results on context effects in consumer choice.

References

Banerjee, K., B. Huebner, & M. D. Hauser (2011). Intuitive moral judgments are robust across demographic variation in gender, education, politics, and religion: A large-scale web-based study. *Journal of Cognition and Culture, 37*, 151–187.

Baron, J. (1994). Nonconsequentialist decisions. *Behavioral and Brain Sciences* 17(1): 1–10.

Baron, J., & Ritov, I. (1994). Reference points and omission bias. *Organizational Behavior and Human Decision Processes*, *59*, 475–498.

Bartels, D. (2008). Principled Moral Sentiment and the Flexibility of Moral Judgment and Decision Making. *Cognition*, *108*, 381–417.

Bazerman, M. H., Moore, D., Tenbrunsel, A. E., & Wade-Benzoni, K. A. (1999). Explaining how preferences change across joint versus separate evaluations. *Journal of Economic Behavior and Organization*, *39,* 41–58.

Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science*, *5*, 187–202.

Busemeyer, J. R., Barkan, R., Mehta, S., & Chatervedi, A. (2007) Context models and models of preferential choice: Implications for Consumer Behavior. *Marketing Theory, 7,* 39–58.

Busemeyer, J. R., Townsend, J. T., Diederich, A., & Barkan, R. (2005). Contrast effects or loss aversion? Comment on Usher and McClelland (2004). *Psychological Review, 112*, 253–255.

Cushman, F, Young L., & Hauser M. (2006). The role of reasoning in moral judgments: Testing three principles of harm. *Psychological Science, 17*, 1082–1089.

Foot, P. (1978). The Problem of abortion and the doctrine of the double effect, in *Virtues and Vices.* Oxford: Basil Blackwell.

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

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

Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. New York, NY: Ecco/HarperCollins Publishers.

Hauser, M. D., Cushman, F. A., Young, L., Jin, R., & Mikhail, J. M. (2007). A dissociation between moral judgment and justification. *Mind and Language*, *22*, 1–21.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*, 247–257.

Huber, J., & Puto, C. (1983). Market boundaries and product choice: Illustrating attraction and substitution effects. *Journal of Consumer Research*, *10*, 31–44.

Iliev, R., Sachdeva, S., Bartels, D., Joseph, C., Suzuki, S. & Medin, D. (2009). Attending to moral values. In B. H. Ross (Series Ed.) & D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of Learning and Motivation*, Vol. 50: Moral Judgment and Decision Making. San Diego, CA: Academic Press.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York, NY: John Wiley & Sons, Inc.

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

Mikhail, J. (2009). Moral Grammar and intuitive jurisprudence: A formal model of unconscious moral and legal knowledge. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Psychology of Learning and Motivation*, Vol. 50: *Moral Judgment and Decision Making*. San Diego, CA: Academic Press.

Paharia, N., Kassam, K., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes, 109*, 134–141.

Petrinovich, L., O'Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolu-

tionary ethics. *Journal of Personality and Social Psychology*, *64*, 467–478.

Pettibone, J. C., & Wedell, D. H. (2000). Examining models of non-dominated decoy effects across judgment and choice. *Organizational Behavior and Human Decision Processes*, *81*, 300–328.

Rai, T. S., & Holyoak, K. J. (2010). Moral principles or consumer preferences? Alternative framings of the trolley problem. *Cognitive Science*, *34*, 311–321.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*, 263–277.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multi-alternative decision field theory: A dynamic connectionist model of decision-making. *Psychological Review,* 108, 370–392.

Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, *15*, 165–184.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158–174.

Simonson, I., & Tversky, A. (1992). Choice in context: tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*, 281–295.

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

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*, 76–105.

Swann, W. B., Gómez, A., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates response to intergroup versions of the trolley problem. *Psychological Science, 21,* 1176–1183.

Thomson, J. (1985). The Trolley Problem, *Yale Law Journal*, *94,* 1395–1415.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, *76*, 31–48.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review, 79*, 281–299.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science, 39*, 1179–1189.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multi-alternative choice. *Psychological Review*, *111*, 757–769.

Waldmann, M. & Dieterich, J. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*, 247–253.

Wolfgang, M., Figlio, R., Tracy, P., & Singer, S. (1985). *The National Survey of Crime Severity**. Washington, D.C.: Bureau of Justice Statistics.