

## **UC Merced**

# **Proceedings of the Annual Meeting of the Cognitive Science Society**

### **Title**

Order Effects in Moral Judgment

### **Permalink**

<https://escholarship.org/uc/item/3sf3f8nn>

### **Journal**

Proceedings of the Annual Meeting of the Cognitive Science Society, 32(32)

### **ISSN**

1069-7977

### **Authors**

Wiegmann, Alex  
Okan, Yasmina  
Nagel, Jonas  
et al.

### **Publication Date**

2010

Peer reviewed

# Order Effects in Moral Judgment

Alex Wiegmann (Alex.Wiegmann@psych.uni-goettingen.de)

Jonas Nagel (jnagel1@uni-goettingen.com)

Stefan Mangold (smangol@gwdg.de)

Department of Psychology,  
University of Göttingen, Germany

Yasmina Okan(yokan@ugr.es)

Department of Experimental Psychology, University of Granada  
Campus Universitario de la Cartuja s/n ,18071, Granada, Spain

## Abstract

Explaining moral intuitions is one of the hot topics of recent cognitive sciences. In the present article we focus on a factor that attracted surprisingly little attention so far, namely the temporal order in which moral scenarios are presented. We argue that previous research points to a systematic pattern of order effects that has been overlooked until now: Only judgments of actions that are normally regarded as morally acceptable are affected by the order of presentation. Additionally, this is only the case for dilemmas immediately preceded by a scenario where the proposed action was judged as morally unacceptable. We conducted an experiment that confirmed this pattern and allowed us to analyze the individual level responses it was generated by. We argue that investigating order effects is necessary for approaching a complete descriptive moral theory.

**Keywords:** moral intuitions; trolley dilemmas; order effects

## Introduction

In the past decades, trolley dilemmas have been used extensively for testing philosophical and psychological theories of moral judgments. In the standard description of the trolley dilemma introduced by Philippa Foot (1967), an out-of-control train threatens to kill five people standing on its track. The only way to prevent this is to pull a switch that redirects the train onto a different track where it will kill only one person. In a modification of this scenario (Thomson, 1976), the only possibility to prevent the five people from being killed is to push a heavy person from a footbridge onto the track. This would stop the train but kill the heavy person. Numerous studies (e.g., Hauser, Cushman, Young, Jin, & Mikhail, 2007) have shown that given the same number of people being killed vs. saved, participants approve of acting in the first but not in the second scenario. Several competing descriptive theories explicate psychological principles supposed to underlie this pattern of moral intuitions (e.g. Greene et al., 2001; Hauser, 2006). However, surprisingly little is known about potential effects of the order in which several consecutive scenarios are presented. It is plausible to assume that consecutive scenarios will not be judged independently of each other: A principle or mechanism that is activated when a particular scenario is represented or evaluated might later be applied to a series of subsequent scenarios. However, only few studies

have dealt with this issue so far (Petrinovich & O'Neill, 1996; Lanteri, Chelini, & Rizzello, 2008). Their results suggest that under certain circumstances moral judgments can indeed be transferred from one situation to another.

If such order effects could be replicated systematically, this would have important implications for psychological theories aiming to explain patterns of moral reasoning at a descriptive level. Furthermore, relevant practical implications would arise both for methodological considerations inside the research laboratory (in terms of controlling for order effects when designing experiments) and for everyday judgments outside the lab.

The present work has three main goals. First, we will provide the first comprehensive review of previous empirical research on order effects in moral judgments, and we will demonstrate that a systematic pattern of results has been overlooked so far. Second, we will empirically test the existence, extent and direction of order effects in reasoning about moral dilemmas. Finally, we will discuss the theoretical and practical implications of our findings, focusing on psychological theories of moral reasoning.

## Order Effects in Previous Research

Speaking of order effects in moral judgment, there are at least two possible interpretations that could be labelled “within-scenario order effect” and “between-scenarios order effect”, respectively. The first type of effect results if the order in which information concerning one particular situation is presented affects judgment. If, for example, the task is to judge the permissibility of an action, and the results solely differ as a function of the particular sequence in which positive and negative consequences are presented, a “within-scenario order effect” occurs. Second, a judgment regarding an action in a particular scenario might be influenced by a judgment that had previously been made about a different scenario. To illustrate, consider two conditions in which a given scenario C is preceded by one of two different scenarios (A vs. B). Differences in judgments of the action scenario C between the two conditions would – all other things being equal – instantiate a “between-scenarios order effect”. The present research will focus on this second category of order effects.

One of the few studies addressing “between-scenarios order effects” in moral reasoning was conducted by Petrinovich and O’Neill (1996). Their aim was to analyze whether the presentation order of a set of moral dilemmas would affect participants’ level of agreement or disagreement with the action proposed in each case. In one condition (standard order), the dilemmas were arranged according to decreasing predicted agreement with the potential action, whereas in the second condition (reversed order) the presentation order was reversed. While Petrinovich and O’Neill (1996) did not report any order effects in an experiment comparing three dilemmas that differed with regards to content (Study 2, Forms 1 and 1R), a reanalysis of their data revealed an order effect for the dilemma with the highest predicted agreeability. In particular, the average agreement rating in this scenario was significantly higher if the scenario had been presented first than if it had been preceded by the other two dilemmas ( $t_{57}=2.11$ ;  $p<.05$ , two-tailed). In contrast, the other two dilemmas received almost equally low ratings in both order conditions. A reanalysis of a similar experiment using a different set of dilemmas (Study 2, Forms 3 and 3R) also revealed that the average rating for one of the positively rated dilemmas varied between the two order conditions. The average rating was lower if the scenario had been directly preceded by a dilemma that received lower (as opposed to higher) ratings ( $t_{68}=2.88$ ;  $p<.01$ , two-tailed).

Another experiment reported by Petrinovich and O’Neill (1996; Study 2, Forms 2 and 2R) compared three different versions of the trolley dilemma. As in the previously reported experiments, a reanalysis revealed order effects for the two scenarios with the highest predicted agreeability ( $t_{57}=2.93$ ;  $p<.01$ , two-tailed, and  $t_{57}=2.58$ ;  $p<.05$ , two-tailed, respectively). However, the third scenario that involved pushing a person from a footbridge in order to stop the train (cf. Table 1) was not affected by the order of presentation.

Similarly, Lanteri, Chelini, and Rizzello (2008) reported order effects for the standard trolley dilemma, but not for the footbridge scenario. In addition, similar order effects were found incidentally in some studies. For example, Nichols and Mallon (2006) found that acting in a case equivalent to standard trolley was marginally more likely to be judged as breaking a rule if the scenario was preceded by a footbridge-equivalent case than if presented in the first position. No analogous effects of a preceding standard trolley-equivalent case on judgments in the footbridge-like case were reported. Recently, Lombrozo (2009) incidentally found results analogous to those obtained by Lanteri et al. (2008). Finally, Alistair Norcross (2008) described an interesting order effect outside of an experimental setting that is nevertheless relevant for the present research. He points out that when he asked his students to evaluate the standard switch-trolley dilemma in the first position, the majority judged that diverting the trolley is permissible. However when this dilemma was preceded by a scenario in which saving the lives of five patients requires to kill a healthy person in order to transplant his organs, the

proportion of students judging that diverting the trolley is permissible was considerably lowered.

### **A Systematic Pattern**

We claim that a closer look at the findings reported reveals a systematic pattern: First, all dilemmas that are affected by an order effect were rated positively (in the sense that the proposed action is on average rated as morally right/acceptable). Dilemmas that received a negative rating seem to be unaffected. Second, the dilemmas that were rated positively are only affected if they are directly preceded by a dilemma that was rated negatively. In this case, the ratings were lower or, in those cases in which the response format is dichotomous, the proportion of people that judge the action to be acceptable decreased.

Previous attempts to account for between-scenario order effects failed to fully capture the pattern we are suggesting here. For instance, Petrinovich and O’Neill (1996) argue that the initial strength of the response (agreement vs. disagreement) influences subsequent responses. If this were true, dilemmas that are normally rated negatively should also be affected by the order of presentation. However, this does not seem to be the case, since these dilemmas seem to be rated equally negative in all cases. Lanteri et al. (2008) take this asymmetry into account when explaining their results. However, they focus on properties of specific scenario contents instead of formulating a general pattern.

It is important to note that so far there is no evidence for a major change of people’s judgments at a qualitative level. In Petrinovich and O’Neill’s study (1996), the ratings for the proposed action do not seem to change enough to be regarded as acceptable in one order condition but as unacceptable in the other. In Lanteri et al. (2008), the percentage of people judging the proposed action as acceptable is indeed lowered, but it still remains above 50%.

Taking into account all the previous points, the main goal of our work will be to empirically test the existence of the pattern described above. If an order effect is present we will aim to determine its strength and, in particular, whether it can be strong enough to lead people to disagree with a proposed action that they would normally (i.e., when evaluated independently) agree with. We will use several variations of the trolley dilemma due to the existence of a large body of previous research establishing how the modification of different factors in these dilemmas affects how they are judged.

## Experiment

### Subjects

Fifty participants (35 women) were recruited using the lab in the psychology department at the University of Göttingen. They were randomly distributed to the different experimental conditions. The average age was 23 ( $SD=2.83$ ).

### Materials

We presented participants a series of five moral dilemma scenarios (the standard switch trolley and four modifications; see Table 1). Each scenario included a brief description of a situation and an action that could potentially be carried out in each case, accompanied by a diagram depicting the situation schematically. The initial description of the situational set-up was identical for all scenarios: An out-of-control trolley rapidly approaches three railroad workers who will die if Karl, the only bystander in the scenario, does not intervene.

Table 1: Summaries of the actions proposed in the five dilemmas

Scenario	Proposed action
Push	Push the large person from the bridge in order to stop the train
Trap	Push a button that will open a trap door in order to let the person on top of the bridge fall onto the track and stop the train
Redirect	Redirect a train containing one person that is on a safe parallel track onto the main track in order to stop the train
Run Over	Redirect an empty train that is on a safe parallel track onto the main track in order to stop the train thereby running over a person that is on the connecting track
Standard	Press a switch that will redirect the out-of-control train onto a parallel track where it will run over one person

This introduction was followed by a description of a specific action that Karl could conduct in order to save the three workers. This action was different for each of the five scenarios, but in all cases it resulted in the death of one innocent person (see Table 1). Instructions were included to ensure that participants assumed that the proposed action was the only available option in each case that, if carried out, would always lead to the described outcome. The number of potential victims (3 vs. 1) was kept constant across scenarios.

In order to establish a baseline of agreement with the action proposed in each of the five different scenarios we conducted a pilot study using a different sample consisting of 100 University of Göttingen students. Participants were

individually approached on campus and asked to indicate for one of the scenarios ( $n=20$ ) whether Karl should act in the proposed way or not on a scale from 1 to 6, where 1 was “not at all” and 6 was “absolutely”. Table 2 shows the average ratings for the different scenarios.

Table 2: Mean ratings (standard deviations) of agreement and percentage of subjects disagreeing with the proposed action in the five scenarios when evaluated independently.

Measure	Scenario (each $n=20$ )				
	Push	Trap	Redirect	Run Over	Standard
Mean Rating	1.95	3.4	4.15	4.4	4.45
(SD)	(1.76)	(1.76)	(1.42)	(1.14)	(1.15)
% Disagreement	80	40	30	10	15

*Note.* % Disagreement is the percentage of subjects who gave a rating  $<3.5$  on a scale ranging from 1 to 6.

Based on these results we ordered the five scenarios according to level of agreement with the proposed action (i.e.,  $Push < Trap < Redirect < Run\ Over < Standard$ ). From here onwards we will refer to this ordering as the *level of agreeability* of the scenario, as defined by the extent to which participants agree with the action when the dilemmas are judged independently.

### Procedure

The experiments were run individually on computers. Initially, the instructions were presented on the screen, followed by the five different scenarios. After each scenario, participants were requested to rate, on a scale from 1 to 6, whether Karl should act in the proposed way or not, where 1 was “not at all” and 6 was “absolutely”. Half of the participants saw the sequence of dilemmas in increasing order of agreeability (Least Agreeable First [LAF] condition, beginning with Push), whereas the other half saw the sequence of dilemmas in the reverse order (Most Agreeable First [MAF] condition, beginning with Standard). The computerized format of the task guaranteed that each dilemma was judged before the following one was presented. Furthermore, there was no possibility for participants to withhold their judgment until the end of the sequence or to switch back in order to change a previously given rating.

### Results

To test whether the pattern of ratings of the dilemmas differed in the two orders of presentation, a  $2 \times 5$  mixed analysis of variance (ANOVA) was conducted, where the first factor was the order of presentation (LAF vs. MAF, between-subjects) and the second factor was the scenario judged (within-subjects). The results are shown in Table 3 and Figure 1. They revealed a main effect for order of presentation. Specifically, average ratings were significantly lower in LAF compared to MAF ( $F_{[1,48]}=8.03$ ;  $p<0.01$ ). Furthermore, we found a main effect for scenario

( $F_{4,192}=23.44$ ;  $p<0.001$ ), confirming our expectation of different average agreeability ratings for the scenarios. Crucially, the interaction between order of presentation and scenario was significant ( $F_{4,192}=8.2$ ;  $p<0.001$ ), suggesting the presence of a strong asymmetric order effect, in line with our predictions.

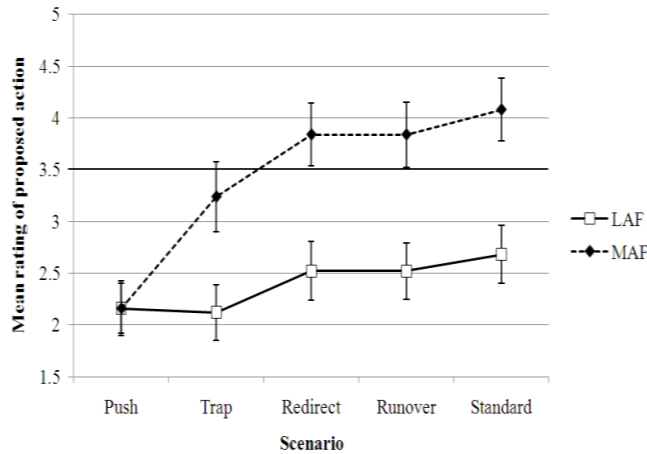


Figure 1: Mean ratings of agreement with the proposed action in the five scenarios when evaluated sequentially, as a function of the order of presentation. Error bars indicate SEM. The bold line at  $y=3.5$  indicates the division between average agreement and disagreement. MAF = Most Agreeable First; LAF = Least Agreeable First.

Table 3: Mean ratings (standard deviations) of agreement and percentage of subjects disagreeing with the proposed action in the five scenarios evaluated sequentially, as a function of the order of presentation.

Order Condition	Scenario				
	Push	Trap	Redirect	Run Over	Standard
Mean ratings (SD)					
MAF ( $n=25$ )	2.16 (1.21)	3.24 (1.69)	3.84 (1.52)	3.84 (1.57)	4.08 (1.53)
LAF ( $n=25$ )	2.16 (1.31)	2.12 (1.33)	2.52 (1.42)	2.52 (1.36)	2.68 (1.41)
% Disagreement					
MAF ( $n=25$ )	76	52	40	32	32
LAF ( $n=25$ )	80	80	72	72	68

Note. % Disagreement is the percentage of subjects who gave a rating  $<3.5$  on a scale ranging from 1 to 6. MAF = Most Agreeable First. LAF = Least Agreeable First.

In order to test our prediction more specifically, we conducted planned comparisons involving Standard and Push as examples of scenarios typically eliciting high and low agreeability ratings, respectively. The average rating for Standard varied considerably depending on the position in which it appeared. When it had been evaluated first, the

average rating was 4.08, while the average was only 2.68 when it appeared at the end of the sequence. This difference was significant ( $F_{1,48}=11.39$ ,  $p<0.01$ ). In contrast, the average rating for the Push scenario was the same in both orders (2.16). Moreover, after computing the within-subjects differences between the ratings for the Standard and the Push scenarios, it can be shown that the average difference is significantly larger in MAF than in LAF ( $F_{1,48}=14.69$ ;  $p<0.001$ ), a result that further supports our prediction of an asymmetrical order effect.

It is worth noting that the difference between the ratings for the Standard scenario in the two order conditions is relevant not only in quantitative but also in qualitative terms: Treating ratings below 3.5 as disagreement and above 3.5 as agreement with the action proposed in a particular scenario, the majority of participants' ratings in LAF would fall into the first category (18 out of 25; 72%) whereas the majority of participants' ratings in MAF would fall into the second (18 out of 25; 72%). This difference is significant ( $\chi^2_1=9.68$ ;  $p<0.01$ ). The same is true for Run Over ( $\chi^2_1=8.01$ ;  $p<0.01$ ), Redirect ( $\chi^2_1=5.20$ ;  $p<0.05$ ), and Trap ( $\chi^2_1=4.37$ ;  $p<0.05$ ), but not for Push ( $\chi^2_1=0.12$ ;  $p=0.73$ ).

## Discussion

In sum, the data were largely in line with the pattern we discovered in previous studies: The judgments of actions that received a positive rating when inquired independently (Standard, Run Over, Redirect) differed significantly in the two order conditions. In contrast, ratings for the action in Push, which was rated negatively when judged independently, did not differ in the two conditions. Furthermore, in the MAF condition, the pattern of the average ratings was very similar to the one obtained when the scenarios were judged independently. In contrast, the average ratings in the LAF condition differed widely from those independent ratings.

It should be noted, however, that the results obtained for one of the scenarios cannot be directly derived from the aforementioned pattern. In particular, Trap was also affected by the order of presentation (both in quantitative and in qualitative terms), even though the proposed action in this scenario was rated slightly negative when judged independently. This finding motivated us to have a closer look at the results at the level of individual participants. In particular, we explored the data treating the ratings as a set of binary choices made by each participant (i.e., treating ratings  $<3.5$  as indication of disagreement and ratings  $>3.5$  as indication of agreement with the proposed action) and observed the tendency that a disagreement with an action was "transferred" to the judgment of the action in the next scenario. That is, an action receiving a positive rating when judged independently received lower ratings when presented as part of a sequence if the preceding scenario was rated negatively by the same participant. In contrast, positive ratings did not affect the ratings of the next action (i.e. changing them into positive ones) if this action was rated negatively in independent ratings. For instance, in the LAF

condition, only three out of 20 participants who disagreed with the proposed action in the initial Push-scenario changed their rating towards agreement during the whole sequence, resulting in 17 votes against the proposed action in the final Standard scenario. In contrast, when participants started with Standard, eleven out of the 17 participants who voted in favor of the proposed action changed their ratings towards disagreement on the way to Push, resulting in only six positive ratings for the proposed action in this final scenario of the sequence. Reformulating the pattern this way allows order effects to occur not only for actions rated positively when judged independently but also for actions rated negatively on average provided that the number of participants who would disagree with the proposed action in a particular scenario is sufficiently higher than the number of participants who would disagree with the action in the subsequent scenario. Within a sequence of scenarios this excess of “disagreements” can be transferred to the next scenario and cause an order effect. On the flipside, an order effect might also occur when the action to be judged in a particular dilemma is preceded by a dilemma where the proposed action is judged positively. Again, it just has to be the case that the number of disagreements in the preceding scenario would be sufficiently higher than in the following scenario if both scenarios were rated independently.

A similar distribution of nominal data could well underlie the results obtained by Petrinovich & O’Neill (1996). Unfortunately, we cannot conclusively confirm this claim because only aggregated results are reported.

It is not possible to determine from our data why the reported asymmetry occurs. However, a possible explanation is the existence of a difference in the urge to justify prohibitions and permissions. When we, e.g., prohibit a child to play with knives we automatically think of – or already have in mind – a justification for this prohibition. Prohibitions seem to call for a justification. In contrast, we do not think about a justification regarding most things we permit. We do not feel an urge to explain or justify to someone why, e.g., he or she is allowed to walk around. Normally, we only justify or explain permissions when a prohibition is the default case. For instance, we might explain to a child that in the case of an emergency an ambulance is permitted to drive over red lights although it is usually prohibited. Applying this line of reasoning to the asymmetric pattern found in our data it might be the case that because participants prohibited the proposed action in Push they were - consciously or unconsciously - thinking about a justification for their prohibition. If they reach a rough justification like “You must not kill an innocent person”, they might keep this principle in mind and apply it to the remaining scenarios. Since an innocent person has to be killed in all scenarios in order to rescue three persons, participants might judge all proposed actions as prohibited. In contrast, when they start with a scenario where they judge the proposed action as permissible it might be the case that no effort is invested in justifying this judgment and,

therefore, no such justification is applied to the remaining scenarios.

### **Implications for Descriptive Moral Theories**

An important goal of descriptive moral theories is to provide an explanation of an average person’s moral judgments that is as comprehensive as possible. A potential source of variance in moral judgments which has received comparably much attention is the structural set-up of the situations in question (e.g., whether the victim serves as means or side-effect in saving the three workers; see Cushman, Young, & Hauser, 2006). However, the effects generated by the manipulation of these factors are usually fairly small, i.e. they account only for a very limited amount of the total variance in moral judgments and thus leave a good portion of between-subject differences unexplained. Thus, considering only factors concerning the objective situational set-up is by no means sufficient to generate a comprehensive descriptive moral theory. Rather, it seems to be necessary to take into account additional psychological mechanisms that influence how a given situational set-up is apprehended, represented, and evaluated. In our experiment, for example, previously judged scenarios seemed to serve as a reference which influenced the judgment of subsequent scenarios. This reference is exogenous to the subsequent scenarios, but indispensable to predicting and explaining the reactions regarding them. Note that, under a certain order condition (LAF), the effects of objective situational parameters that can usually be found have almost entirely vanished. The strength of this effect demonstrates the large predictive potential of such exogenous factors and underpins the importance of spending more efforts on investigating them in the future.

According to our results, differential experiences prior to a moral judgment can have a profound influence on this judgment. Such effects can be expected to be especially large under conditions that strongly suggest the adequacy of transferring a certain judgment from one scenario to the next. This is the case if one person is required to give several subsequent judgments on various cases similar in structure or content in a within-subjects design. As our results suggest, extreme caution is required if responses generated under such conditions are to be attributed to properties of the scenarios themselves.

Finally, we believe that between-scenario order effects might also play a role under conditions outside the laboratory. The viewpoints taken by people discussing moral issues in everyday life might be highly affected by the issues that have been discussed immediately before. This influence might not only be quantitative, but even qualitative. Possible areas of application might be the design of public opinion polls or surveys that consecutively gauge responses to several (moral) issues. Previous research in other contexts showing that such instruments can be highly sensitive to effects of question positioning (e.g., Benton & Daly, 1991;) in combination with our results from the moral domain support this claim. On the other hand, we

acknowledge that the similarity between the dilemmas used in our study might particularly encourage the transfer of judgments between scenarios. It might be that in cases where the issues in question present a larger variability in structure or content, order effects would diminish, and ratings would be more similar to those made independently.

### Summary and conclusions

In this article, we have argued that order effects can have a profound influence on judgments of actions in moral dilemma situations. Amongst order effects, we subsume cases in which a given action is judged differently when rated independently, as compared to when it has been preceded by one or several other scenarios. We began by reviewing the (scarce) literature on order effects in moral psychological research. We then reported the results of an experiment conducted in order to find out whether the pattern of results extracted from the literature reviewed could be replicated. For four out of our five scenarios this was the case: Three scenarios that received positive ratings when evaluated independently received negative ratings when directly preceded by a scenario that had been judged negatively. The ratings for Push were also in line with this pattern, since the proposed action in this case was rated negatively when judged independently and was not affected by the order of presentation.

However, one scenario where the proposed action received slightly negative ratings when judged independently (Trap) was also affected by the order of presentation. This finding motivated us to have a closer look at the results by performing an analysis treating the individual ratings as binary choices. Following this analysis, we reformulated the pattern as follows: In those cases in which a participant disagrees with the action proposed, this judgment is likely to be “transferred” to the judgment of the action in the next scenario, even if this action is rated positively when judged independently. However, positive ratings are not able to change the ratings of the next action into positive ones if people normally disagree with the action proposed in this case.

We went on by speculating what could explain the asymmetry between negative and positive ratings in terms of the potential to be transferred to the next case. One candidate feature discussed was the greater urge to justify prohibitions (negative ratings) compared to permissions (positive ratings). Of course, more research is needed in order to evaluate explanatory mechanisms underlying the observed asymmetry.

In the last section of the paper, we discussed the implications of our findings for descriptive theories of morality. We argued that descriptive moral researchers should be extremely cautious when interpreting results of experiments using within-subjects designs. Furthermore, we contended that they should devote more attention to general psychological mechanisms contributing to moral judgment in addition to focusing on features of particular scenarios.

Overall, the present study should draw the attention of descriptive theories of moral judgment to previously overlooked important sources of variance such as order effects. Due to the crucial implications of these findings, much more empirical and theoretical research needs to be done in the future in order to address determinants, mechanisms, and boundary conditions of the issues discussed here.

### Acknowledgements

We thank Johanna Kirchhoff for data collection.

### References

- Benton, J. E. & Daly, J. L. (1991). A question order effect in a local government survey. *Public Opinion Quarterly*, 55, 640-642.
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral Judgment: Testing three principles of harm. *Psychological Science*, 17(12), 1082-1089.
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, 5, 5-15.
- 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(5537), 2105-2108.
- Hauser, M. (2006). *Moral minds: How nature designed our universal sense of right and wrong*. Ecco.
- Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & Language*, 22(1), 1-21.
- Lanteri, Chelini, & Rizzello (2008). Lanteri, A., Chelini, C., & Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem. *Journal of Business Ethics*, 83(4), 789-804.
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100(3), 530-542.
- Norcross, A. (2008). Off her trolley? Frances Kamm and the metaphysics of morality. *Utilitas*, 20(01), 65-80.
- Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, 17(3), 145-171.
- Thomson, J. J. (1976). Killing, letting die, and the trolley problem. *The Monist*, 59, 204-217.