

A Double Causal Contrast Theory of Moral Intuitions in Trolley Dilemmas

Michael R. Waldmann (michael.waldmann@bio.uni-goettingen.de)

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

Department of Psychology, University of Göttingen,

Gosslarstr. 14, 37073 Göttingen, Germany

Abstract

In trolley dilemmas a train is about to kill several victims who could be saved if instead a different victim is harmed. A number of theories have been proposed which assume that permissibility judgments in these harm-based moral dilemmas are mediated by an analysis of the underlying causal structure. For example, it has been postulated that it is permissible to harm people as a side effect but not as a means. We have developed a different causal theory which claims that moral judgments are influenced by two contrasts, the global contrast between the number of victims in the presence and absence of the act, and an additional local contrast that compares the fates of the morally relevant target (i.e., threats, victims) of the proposed intervention in the presence versus absence of the act. This double causal contrast theory explains intuitions in various types of trolley dilemmas better than its competitors.

Keywords: moral reasoning; trolley dilemmas; causal reasoning; doctrine of double effect

Introduction

Trolley dilemmas have become the drosophila for testing alternative philosophical and psychological theories of moral judgments in harm-based moral dilemmas (see Kamm, 2007). In the philosopher's Judith Thomson's (1986) version of the trolley dilemma, a situation is described in which a trolley whose brakes fail is about to run over five workers who work on the tracks. However, the trolley could be redirected by a bystander on a side track where only one worker would be killed (bystander problem). Is it morally permissible for the bystander to throw the switch or is it better not to act and let fate run its course? Most people seem to have the intuition that throwing the switch is morally required or at least permissible. However, the intuitions change in another of Thomson's (1986) examples, in which the train could be stopped by throwing a fat person from a footbridge on the tracks, thus stopping the train with his body (footbridge dilemma). Most people find this act outrageous, even though again one person is sacrificed to save five. For philosophical theories these two intuitions present a puzzle. The intuitions in the bystander dilemma seem to be in line with utilitarian or consequentialist theories that focus on the favorable outcome of the act in contrast to not acting (1 vs. 5 dead people). However, the footbridge dilemma yields the same outcomes. The intuitions in this dilemma seem to be more consistent with non-consequentialist reasoning, which focuses on the impermissibility of the act of killing a person.

Not only in philosophy but also in psychology the trolley dilemmas have attracted interest as test cases for psycholog-

ical theories of moral intuitions. Some have derided this research as trolleyology because of the artificiality of the task. It is certainly true that most people never will be in a situation that mimics the trolley problem. However, we would like to defend this paradigm as a valuable tool to study the cognitive basis of moral intuitions. People care about how society should deal with violent death, severe illness, terrorism, or emergency, even though they may never be involved in a dilemma involving these events. Nevertheless, these intuitions influence how our society and law functions. Thus, it is important to understand the mechanisms that underlie people's moral intuitions.

Threat vs. Victim Interventions

From a psychological point of view, the philosophical comparisons between bystander and footbridge trolley versions are flawed because of the various confounds. The footbridge dilemma differs in a number of relevant features from the bystander problem, including the act (re-directing a train vs. pushing a person), the physical distance between agent and victim, the directness, and the saliency of the death, or the degree of intentionality (see also Greene et al., 2009; Waldmann & Dieterich, 2007, for evidence). Unfortunately, in the early research on trolley dilemmas psychologists have often adopted close variants of Thomson's (1986) versions, which makes it hard to interpret the results of these studies (Greene et al., 2001; Mikhail, 2007). In our own research we have therefore tried to create variants of trolley dilemmas, which are better controlled so that some of the already well known factors affecting moral intuitions (e.g., distance, violence of act) are kept constant (Waldmann & Dieterich, 2007). We will first present a new, better controlled experiment which highlights the structural differences between different variants of trolley dilemmas. This experiment will serve as the base example for presenting competing theories, which then will be tested in additional experiments.

General Procedure Unless otherwise noted all experiments were run in groups (including seminars and lectures) with students from the University of Göttingen, Germany. Participants came from various fields, but we excluded philosophy and economics to avoid prior exposure to relevant philosophical positions. Subjects were handed booklets in which they were told that they are going to read about a situation which mentions two options of an agent in the story. All dilemmas used a format in which a fictitious agent in a remote control room of a train company is presented with two alternatives with outcomes, which lie in the future. The outcomes were clearly stated and characterized as cer-

tain. In the instructions it was pointed out that participants should carefully read the stories and attempt to empathize with the situation of the agent. The story was presented in a brief story that described the moral dilemma and the future options. Additionally, images were shown that presented the two options (acting vs. non acting)(see figures for examples). Subsequently, a rating scale was presented. Generally participants were asked to rate whether the agent should act or not in the described situation. The scale ranged from 1 ("not at all") to 6 ("definitely") with separated numbered boxes.

Experiment 1

In Experiment 1 we compared two parallel versions in which we manipulated the locus of intervention, threat versus victim. In the threat intervention condition the threatening train is redirected, in the victim intervention condition the train in which the single alternative victim is sitting is targeted. In both variants of the trolley problems all trains are moving and can only be redirected by employees of the train company who are sitting in a remote control room. The workers on the trains did not have any control over the trains. In the threat intervention condition ($n=15$)(Condition I), which corresponds to the bystander problem, five track workers sit on train A and one on train B. The empty train C, which represents the threat, is, due to a signaling defect, running behind train A and cannot be stopped. Soon it would hit train A with the five workers. However, the control room could throw the switch and redirect the train on the parallel track where it would hit train B. In both cases the victims would be seriously hurt (see Fig 1, I).

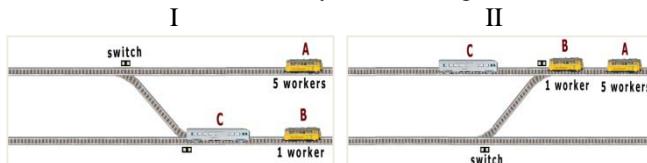


Fig. 1. Illustrations of the consequences of the proposed interventions in the threat (I) versus victim (II) intervention conditions in Experiment 1. In Condition I the threatening train C has been redirected to the side track, in Condition II train B with the victim has been redirected to the main track where it stops train C.

In the victim intervention condition ($n=14$)(Condition II), the first part of the story is identical. However, here the option is to redirect train B by throwing the switch on the parallel track. This way train B would go up to the track where the other two trains are running and would end up in between train C and train A (see Fig. 1, II). Now train C would hit train B which would stop the threatening train C. This would seriously hurt the one worker in train B, but save the five in train A. Consistent with the findings about bystander and footbridge dilemmas, the threat intervention option was rated more acceptable ($M=4.93$, $SD=0.79$) than the victim intervention option ($M=2.57$, $SD=1.02$), $F(1, 27)=48.8$, $p=0.00$.

Causal Theories of Moral Intuitions

How can the different moral assessments of threat and victim interventions be explained? We kept various familiar factors constant so that some simple accounts are ruled out. In both conditions the distance between the intervention and harm is roughly the same, the initial act (re-directing a train via remote control) is identical; in both cases the act only indirectly affects the fate of the victim, and there is no physical closeness or personal force. Moreover, none of the passengers has control over the train so that there are no differences in responsibility. What are the structural differences that may account for the different intuitions? Given that moral judgments are primarily about evaluating the moral quality of acts or interventions which lead to outcomes, causal theories seem to be a prime candidate for an analysis of such differences.

All theories that can be described as causal include the contrast between the outcomes in the presence versus absence of the intervention, and predict that the size of the contrast influences moral permissibility judgments. However, different theories postulate different representations of the acts and focus on different causal features.

Consequentialism Consequentialism is primarily interested in the contrast between the outcomes. Thus, consequentialist approaches choose a fairly abstract level of describing the acts as acting versus not acting, which blurs the differences between threat and victim interventions. This level of representation in both conditions yields the global outcome contrast between one dead person when the agent acts and five dead people when she refrains from acting (i.e., 1:5). Therefore, this theory predicts generally high acceptability ratings for the act. This may be acceptable as a normative principle (see Unger, 1996), but fails as a psychological account. The theory correctly predicts the intuitions in the threat intervention condition but makes wrong predictions for the victim intervention condition.

Doctrine of Double Effect Traditional non-consequentialist or deontological theories focus on moral rules permitting or prohibiting acts. For example, harmful acts, such as killing, are prohibited. However, simply prohibiting such acts also does not explain the intuitions in trolley dilemmas because apparently people find killing in the threat intervention condition acceptable. A more promising variant of a non-consequentialist theory accounting for trolley intuitions is the doctrine of double effect (DDE), an old deontological rule that is based on a causal analysis and also includes contrasts. A number of psychologists have proposed this rule as a moral heuristic (Royzman & Baron, 2002), or part of an innate moral grammar (Hauser, 2006; Mikhail, 2007). According to the dominant reading of the DDE it is permitted to do a neutral or good act as a means to a greater good, although we foresee lesser harm as a side effect, assuming that there are no better alternatives. However, it is impermissible to bring about lesser harm as an end in itself or as a means to a greater good. Thus, the DDE contains two stages: First a global favorable contrast needs to be ascertained ("greater good") (i.e., 1:5 in the trolley dilemmas). We know

already that this global contrast does not explain the effect, although it is certainly the case that the size of this contrast influences judgments (Nichols & Mallon, 2006). The main focus of the DDE is on the causal processes entailed by the proposed act. Here the doctrine distinguishes between two types of causal processes involving the single victim. If the victim is harmed as a side effect, as in the threat intervention condition, the act is permitted. However, if the victim is used as a means to save the five, as in the victim intervention condition, the act is prohibited. Thus, this rule explains the intuitions in the two conditions of Experiment 1. Importantly the DDE explains the different intuitions by analyzing the causal processes in the *presence* of the proposed intervention, whereas a contrast with events in the *absence* of the intervention does not play a role after the initial evaluation stage.

A Double Contrast Theory We are going to propose and test another variant of a causal contrast theory, our double contrast theory, which is an extension of Waldmann and Dieterich's (2007) proposal. Our main assumption is that subjects choose a level of abstraction of the act that brings out the specific causal characteristics of the proposed intervention. Contrasting the two interventions on the abstract level as presence or absence of acting or as killing and saving is too abstract because it does not reveal the differences between the scenarios. Using a very low-level description, such as button pressing on a remote control, also blurs the differences. We believe the most natural basic level description in the scenarios refers to the kind of intervention and the morally relevant target of the intervention. Morally relevant targets in trolley dilemmas are threats or victims, which can be stopped, redirected, derailed and so forth by the interventions. This is also the level of description that is used in the stories describing trolley dilemmas. For example, a natural description of the interventions in Experiment 1 might state that in Condition I the threatening train is redirected, whereas in Condition II the train with the single victim is set into motion towards the threatening train. Thus, in Condition I the threatening train is the target of intervention, whereas in Condition II the train with the single victim is the target of intervention.

Our main claim is that people will focus on the target of intervention and assess the harm directly caused by intervening in this target in contrast to the harm *the target* would cause in the absence of the intervention. This local contrast which focuses on the target of intervention rather than the global outcomes will, according to our theory, heavily influence the acceptability rating.

How does the double contrast theory explain the two standard dilemmas? In general, the morally relevant targets of intervention in our trolley dilemmas are either the trains which pose a threat, or the trains which house a potential victim. In the threat intervention condition (I) the proposed act can be summarized as re-directing the threat. Thus, the morally relevant target is the threatening trolley C. To assess the local contrast we need to focus on the direct harm caused by the target of intervention, train C, which is one

seriously harmed person. This outcome is contrasted with the direct harm caused by the target of intervention (i.e., train C) in the absence of the intervention, which in Condition I are five people who are harmed by train C in the absence of an intervention. Thus, the local and global contrasts are the same in this case (1:5), both favoring the proposed intervention.

In contrast, in the victim intervention condition (II) the proposed act can be described as re-directing train B with its potential victim towards the threatening train C. Thus, train B with its potential victim is the target of intervention, and the local contrast will therefore focus on train B with its single potential victim. Setting this train into motion will directly cause harm to this victim. The fact that five people are saved further in the future is an indirect, more remote consequence of the act and therefore not part of the local contrast. To compute the local contrast the harm caused by the target of intervention in the absence of the act also needs to be considered. Train B with its single passenger, the target of intervention, would safely stay on the side track so that its passenger would not be harmed. Thus, the local contrast focusing on train B would amount to 1:0 (1 harmed vs. 0 harmed). The local contrast implies that the act is harmful, which predicts the lowered acceptability ratings.

As in the other theories we also believe that the global contrast (1:5) additionally plays a role, which explains why the ratings are not at a minimum. However, we assume that these global contrasts are backgrounded. In this regard, the double contrast theory makes similar assumptions as the DDE. But whereas the DDE explains differences of intuitions by focusing on the causal structure entailed by the acts, the double contrast theory focuses on the *contrast* of the fate of the target of intervention. In sum, both the double contrast and the doctrine of double effect explain the patterns in the standard trolley cases (e.g., Experiment 1).

Evidence for the Double Contrast Theory

In order to test our double contrast theory against its competitors we started to look for alternative versions of the trolley problem that better distinguish between the theories. In previous trolley research the target of intervention and the location of the alternative victim were often confounded. Whereas threat interventions typically redirect empty trains, victim interventions more directly intervene in the alternative victim. Other variants of the trolley problem allow us to disentangle these and other confounds, and provide informative tests for the alternative theories.

Experiment 2

In Experiment 2 we ran four conditions with 20 participants in each condition. Condition I is a standard threat intervention condition in which an empty threatening train can be redirected away from five victims towards one. All victims are sitting in trains, as in Experiment 1. As usual, this condition yielded relatively high mean ratings ($M=4.6$, $SD=1.57$), which signals high acceptance for the act. Our theory predicts this pattern as a consequence of the 1:5 contrast (see

above). Condition II is new (see Fig. 2): Here a passenger is sitting in the threatening train C. According to the instructions this passenger has no control over the train. The train is about to kill the five on the main track if nothing is done. However, in our instructions we stated that the passenger on the threatening train will be able to jump off the train before it crashes into the train with the five, and save himself. Thus, in the absence of an intervention five people would die, as in Condition I. Alternatively the threatening train could be redirected. Unfortunately, the train needs to be redirected to a side track which traverses a bridge. This bridge prevents the passenger on the threatening train from jumping off so that he will be killed in the collision between train C and the empty train B, which is parked on the side track behind the bridge. This is a novel condition because the intervention targets a threatening train which also transports a potential victim. Thus, this is a case of both a threat and a victim intervention. Interestingly, this condition descriptively received slightly (although not significantly) higher acceptability ratings ($M=5.0$, $SD=1.3$) than Condition I, which means that most subjects opted for sacrificing the one. Although in this condition a train with a single victim is the direct target of a harmful intervention, this variant of victim intervention is not aversive.

How does our theory explain this finding? According to the double contrast theory subjects will compute a local contrast on the morally relevant target of intervention. In both Conditions I and II the target is the threatening train C, which in one condition is empty and in the other houses a potential victim. In both conditions, train C directly harms one person in the presence of the intervention but harms five people in the absence of the intervention. Thus, both Conditions I and II yield the same 1:5 local (and simultaneously global) contrast, which favors acting.

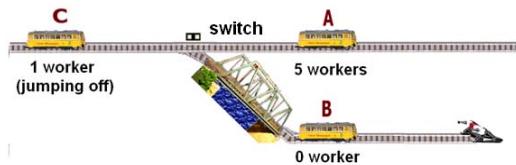


Fig. 2: Illustration of Condition II in Experiment 2 (see text for details).

We ran two more conditions. The most interesting condition of this experiment is Condition III. Here the threatening train C again carries a passenger who has no control over the train, and who is about to jump off (see Fig. 2). In the absence of the intervention, the five in train A at the end of the main track would be killed. On the side track an empty train B is parked, which could be directed upward toward the threatening train C. This empty train would stop the threatening train C on the main track but would kill its single passenger, who, according to the instructions, would not have sufficient time left to jump off. Note that killing the one with the empty train is on the causal path of preventing harm to the five. Thus, the train with its single passenger is

used as a means to prevent harm from the five. Harming people and using them as means against their will should, according to the DDE, be aversive (see Experiment 3 for further discussions of the concept of means). In contrast to the predictions of DDE, however, we got again high acceptability ratings ($M=4.7$, $SD=1.69$), which in fact are statistically equivalent to the ones in the standard threat intervention condition (I).

Condition IV is a standard victim intervention condition, which serves as a control. An empty threatening train C is heading toward a train (A) with five passengers. At the end of a side track, which leads over a bridge, a train (B) with a single passenger is parked. This train B with its passenger can be set in motion in the direction of the main track where it would arrive in time to stop the threatening train C, however with fatal consequences for the single passenger. This condition yielded the expected low ratings ($M=3.15$, $SD=2.01$). In fact, these ratings proved significantly lower than the ratings in the three other conditions, $F(1, 76)=16.2$, $p=0.00$, which were not significantly different from each other.

How does our double contrast theory explain the difference between Condition III and the superficially similar standard victim intervention, Condition IV? Note that in both conditions the train that is parked on the side track is set in motion, and directed towards the threatening train on the main track. Thus, at first sight one might conclude that this empty train is in both scenarios the target of intervention. However, this is wrong according to our theory. In the victim intervention condition (III) the morally relevant target of intervention is indeed the train on the side track with its potential victim, who would either be killed or would stay alive. Thus, the local contrast favors inaction (1:0). However, although the act seems superficially similar in Condition IV, in this condition the train that is being moved is empty. Thus, it neither represents a threat nor is a victim located inside the train. This train is therefore not a morally relevant target of the intervention; it rather plays the causal role of an instrument to stop the threatening train. In this regard the empty train is similar to other morally irrelevant instruments, such as the remote control or button presses. As a consequence, the threatening train C, not the empty train B is the morally relevant target of the intervention in Condition IV. Computing the local contrast over the harmful outcomes train C is causing in the presence versus absence of the intervention yields a 1:5 local (and global) contrast, which favors the intervention. In sum, the results of the experiment favor our double contrast theory over the DDE and related principles (Kamm, 2007).

Experiment 3

In Experiment 3 we ran different variants of some of the conditions in Experiment 2 along with new conditions. This experiment provides further tests of the DDE and our double contrast theory. Again we used the standard trolley instruction about a threatening train on a test site which, due to a brake failure, is about to hit a train with five track workers

at the end of the main track. These five workers would be killed. As in Experiment 1 there is also a parallel side track, which is connected to the main track via a connecting track (see Fig. 3). We ran four conditions. As in the other experiments the passengers inside the trains had no control over the trains and therefore were not responsible for the outcomes in all conditions. In both Conditions I and II we placed the single victim inside the threatening train B in a safe location in the rear of the train. Thus, unlike in the last experiment the passenger does not need to jump off the train to save himself. Doing nothing leads to the death of the five in train A at the end of the main track, but would spare the passenger in the safe location inside the threatening train B. In Condition I ($n=58$), the instructions propose as an alternative that the agents in the remote control station could redirect an empty train C located on the parallel side track up to the main track, thus hitting the threatening train in the rear section and thereby leading to the death of the single passenger (see Fig. 3). However, the threatening train B would be derailed saving the five. This scenario led to fairly high ratings ($M=4.4$, $SD=1.28$).

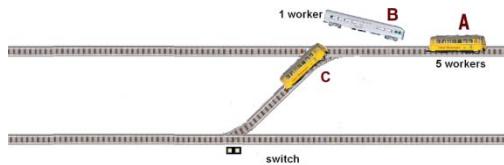


Fig. 3: Illustration of consequence of proposed act in Conditions I and II of Experiment 3 (see text for details).

Condition II was similar. However, to provide a clear cut case against the DDE, we made the role of the victim as a means more salient. Technically one could argue that in all our experiments the means of saving the five were the redirected trains, not the bodies of the passengers, whose deaths could be construed as side effects. However, such an argument would not save the DDE because then the difference between the threat and victim intervention in Experiment 1, for example, would be a puzzle. Moreover, we doubt that people would construe their harm as a side effect if they were sitting in a vehicle that is being used without their consent to save others (see also Kamm, 2007). Anyhow, in Condition II ($n=54$) we stated again that a single passenger in train C, who is unfamiliar with the steering and brake system, is sitting in the rear of the train in a safe location. Now employees in the control station, guided by a camera inside the train, notice that by hitting the train, the passenger would fortuitously be pushed against the brake system, which would lead to a derailment of the train. The passenger would be killed by this act but the five would be saved. In this instruction the body of the victim is clearly specified as a necessary means for the goal to derail the train and save the five. Interestingly, similarly high ratings as in Condition I were obtained ($M=4.81$, $SD=1.04$). In fact, descriptively these were the highest ratings in this experiment. Clearly participants were not sensitive to whether the body

of the victim was causally necessary for saving the five or not.

Both Conditions I and II refute the DDE as a theoretical account. Although in both conditions the single victim was used as a means to save the five, subjects found the intervention highly acceptable. This finding is explained by the double contrast theory. As in Condition III in Experiment 2, in Conditions I and II of Experiment 3 the empty train C plays the role of an instrument, the morally relevant target of intervention is train B, which both constitutes a threat and houses a potential victim. In the presence of the intervention train B, the target of intervention, is involved in the death of one victim while in the absence of the intervention the five passengers in train A die. Thus, this conditions leads to a 1:5 local and global contrast.

To ascertain that the high ratings in Conditions I and II are indeed different from predictably aversive conditions, we also ran Condition III as a control, which is the standard victim intervention condition ($n=49$). In this condition train C on the side track which transports a single passenger is redirected through the connecting track to the main track where the train would hit and derail the empty threatening train B, thus leading to the death of the one in train C, but saving the five in train A. As usual, this intervention was given fairly low ratings ($M=3.76$, $SD=1.64$), which is predicted by our theory as a result of the 1:0 local contrast.

Finally, in Condition IV ($n=51$), a fourth train D in which one worker is sitting was introduced which is parked on the connecting track, thus blocking the way to the main track. The proposed intervention was to send an empty train C located on the parallel side track up the connecting train, thus derailing train D on the connecting track, and thereby killing its passenger. After stating this fact, the instruction mentioned that this event will open up the way to the main track where train C from the side track could derail the empty threatening train B on the main track, thus saving the five in train A. This intervention also yields fairly low ratings ($M=3.88$, $SD=1.37$). How does our theory explain the finding in Condition IV? The initial morally relevant target of intervention in this condition is train D, which is parked with its potential victim on the connecting track. This victim dies in the presence but would be alive in the absence of the intervention, thus creating a 1:0 local contrast.

The general pattern is confirmed by an ANOVA: Conditions I and II, which are statistically equivalent, yielded significantly higher acceptability ratings than Conditions III and IV, $F(1, 208)=11.30$, $p<0.001$.

General Discussion

The goal of our studies was to test theories of moral acceptability in harm-based moral dilemmas. Certainly there are other types of moral problems which might require different theories (Haidt, 2007). Trolley dilemmas represent interesting test cases for cognitive theories because they show that our moral intuitions are influenced by structural factors which go beyond simple comparisons between outcomes (e.g., numbers of victims) or acts (e.g., killing, saving).

Despite identical outcomes and the identical conflict between saving and harming, our moral intuitions differ depending on various factors including the kind of act, distance, intention, contact, legal responsibility, personal force, or the framing of the outcomes (e.g., Greene et al., 2009; Rai & Holyoak, 2010). In our studies we tried to control for these already known factors in order to focus on the remaining structural causal differences between types of scenarios, which pose a puzzle for both psychologists and philosophers.

A number of moral theories focus on causal structures and are therefore candidates for explaining effects of such structural differences. These theories differ in the choice of the level of description and in the postulated relevant causal features. *Consequentialism* focuses on outcomes, and therefore uses abstract descriptions of acts. The moral analysis contrasts global outcomes in the presence and absence of the act. This theory fails as a psychological account.

A second causal account, the non-consequentialist *doctrine of double effect* also tests for a favorable global contrast first, but then focuses on the causal paths entailed by the act under consideration. Here the distinction between harming people as a means versus as a side effect carries most of the weight in explaining differences in intuitions in trolley dilemmas.

A third theory, our *double contrast theory*, also starts by considering the global contrast. But then a local contrast is computed using basic level descriptions of the interventions targeting threats or victims. For example, in the victim intervention conditions people represent the intervention as re-directing the victim, and consider what will happen to this victim in the presence versus absence of the proposed act.

Three experiments have shown that the double contrast theory wins over the doctrine of double effect. People clearly find it acceptable to use people as means without their consent when the local contrast favors the act.

Directions for Future Research

More research is needed on how people choose the level of description in moral dilemmas. It would be interesting to present subjects with still movies, and have them describe the scenarios in moral and non-moral settings.

Another interesting goal would be to further explore the factors influencing local contrasts. In our experiments we have chosen interventions in which the acts were morally innocuous (e.g., throwing a switch). In the contrast between re-directing a victim and not re-directing the victim, the morally relevant contrast is surely about what happens to the victim. However, if the intervention was shooting a victim versus not shooting her, the contrast between shooting and not shooting would certainly impact on the moral evaluation of the contrast. A clear example of this case is, for example, the famous *Jim and the Indians* dilemma, in which Jim is given the choice of watching twenty Indians be shot or shoot one of these twenty Indians himself, thus saving the rest (Williams, 1973). Although the local contrast for the Indian, Jim could shoot, would be 1:1 (he is dead regardless

of the act), the act is certainly aversive because of the shooting component of the contrast shooting the Indian vs. not shooting the Indian.

Finally it would be interesting to get a more quantitative assessment of the relative weight between global and local contrasts. Global contrasts surely affect moral assessments, as can easily be seen if we consider a 1:1.000.000 contrast in a disaster variant of a trolley problem (Nichols & Mallon, 2006). Note that none of the previous theories includes assumptions about how global contrasts quantitatively affect judgments because moral philosophers typically ask about permissibility, not about degree of permissibility. Our experiments clearly suggest that local contrasts dominate judgments but they do not allow us to answer the question how much weight these contrasts have relative to the global contrast.

References

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.

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.

Haidt, J. (2007). The new synthesis in moral psychology. *Science*, 316, 998-1002.

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

Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford: Oxford University Press.

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

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, 100, 530-542.

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

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

Thomson, J. J. (1986). The trolley problem. In J. J. Thomson, *Rights, restitution, and risk. Essays in moral theory* (pp. 94-116). Cambridge, MA: Harvard University Press.

Unger, P. (1996). *Living high and letting die*. New York: Oxford University Press.

Waldmann, M. R., & 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.

Williams, B. (1973). A critique of utilitarianism. In J. J. C. Smart & B. Williams, *Utilitarianism: For and against* (pp. 82-117). Cambridge: University Press.