Research Article

Throwing a Bomb on a Person
Versus Throwing a Person on a Bomb

Intervention Myopia in Moral Intuitions

Michael R. Waldmann and Jörn H. Dieterich

University of Göttingen, Göttingen, Germany

ABSTRACT—Most people consider it morally acceptable to redirect a trolley that is about to kill five people to a track where the trolley would kill only one person. In this situation, people seem to follow the guidelines of utilitarianism by preferring to minimize the number of victims. However, most people would not consider it moral to have a visitor in a hospital killed to save the lives of five patients who were otherwise going to die. We conducted two experiments in which we pinpointed a novel factor behind these conflicting intuitions. We show that moral intuitions are influenced by the locus of the intervention in the underlying causal model. In moral dilemmas, judgments conforming to the prescriptions of utilitarianism are more likely when the intervention influences the path of the agent of harm (e.g., the trolley) than when the intervention influences the path of the potential patient (i.e., victim).

Reflecting on moral dilemmas has always fascinated human-kind. Although most people are rarely in a situation in which they personally make life-and-death decisions, many people feel that moral issues are important and that they should care about how society handles these problems.

In philosophy, a number of competing ethical positions have been proposed. Utilitarianism is a moral doctrine stating that the right act is the one that produces the best overall outcome, as judged from an impersonal standpoint that takes into account the interests of everyone (see Darwall, 2003a). Consequences are evaluated independently of the acts that lead to them. The value or disvalue of the consequences, not moral evaluation of the acts, determines the appropriateness of actions. Although assessing the value of consequences may often be difficult, there are many situations in which most people would agree in their assessment. For example, most people would agree that saving more people’s lives is, ceteris paribus, preferable to saving fewer people’s lives.

Deontology represents a different view (see Darwall, 2003b; Thomson, 1990). According to this ethical position, there is a set of moral rules, rights, duties, or obligations that people must honor regardless of the consequences. For example, many people would agree that one should not kill other people even when doing so prevents more deaths.

Often these two views will lead to the same conclusions. However, there are situations in which the recommendations of the two positions diverge. Philosophers have constructed a number of moral dilemmas that concisely demonstrate the conflicts between these two ethical positions (see Foot, 1967; Thomson, 1976, 1985, 1990). The famous trolley problem describes a situation in which a trolley with brakes that have failed is about to run over five workers on the tracks. The trolley could be redirected to a side track where only one worker would be killed. Is it morally permissible for a bystander to throw the switch that will redirect the trolley, or is it better for the bystander not to act and let fate run its course? Although redirecting the trolley is clearly an act of directly killing one person (imagine the same situation without the five victims on the original track), most people find it permissible to throw the switch, a choice that seems to violate the deontological rule not to actively kill people but is in line with utilitarianism.

Now consider the organ-transplant dilemma that often has been contrasted with the trolley problem. In this scenario, each of five patients in a hospital is dying because of failure of a
different organ. The patients could be saved if the resident surgeon kills a visitor and transplants her organs. Most people would find it outrageous to even consider this option, despite the fact that the choice is also between five dead people or one dead person. Thus, in this situation, deontology seems to prevail. Indeed, in a functional magnetic resonance imaging study, Greene, Sommerville, Nystrom, Darley, and Cohen (2001) found that people react more emotionally to dilemmas like the organ-transplant problem than to dilemmas like the trolley problem.

Whereas philosophers are interested in the development of a normative theory, the goals of psychology are primarily descriptive. Often the cases discussed in philosophy (e.g., trolley vs. organ-transplant problems) vary in too many features to provide a basis for localizing the psychological reason for the differences in people's judgments. The goal of psychological studies is to experimentally pinpoint specific relevant factors without pursuing philosophers' more ambitious goal of giving a complete normative account.

CAUSAL MODELS AND POINTS OF INTERVENTION

Moral dilemmas, such as the trolley problem, describe causal scenarios in which actions or interventions generate physical processes that lead to more or less morally desirable outcomes. Given that causal models include information about actions as well as their consequences (i.e., causal effects), causal models provide a conceptual basis that integrates the utilitarian focus on consequences with the deontological focus on acts. The trolley and the organ-transplant dilemmas differ in a large number of respects, including the type of action, physical distance (direct killing by the physician vs. more remote intervention in the trolley case), and kind of victims (track worker vs. visitor). Continuing our work on the causal-model theory (see Waldmann & Hagmayer, 2005; Waldmann, Hagmayer, & Blaisdell, 2006), which has provided a framework for studying the role of abstract, domain-general features such as the structure of causal models in reasoning and learning, we are interested in whether such abstract features of causal models affect moral judgment.

One such feature is the point within the causal model that is targeted by the intervention. Adopting the terminology of force-dynamics theory, we distinguish between agents and patients (Jackendoff, 2002, p. 363; see also Wolff & Song, 2003). Agents (e.g., the trolley) play an active role in bringing about a harmful effect (e.g., death) to patients (e.g., the track workers). In the standard trolley problem, the trolley, which is the potential agent of harm, is redirected by throwing a switch. We call this type of action agent intervention (Fig. 1, top panel). The intervention changes the causal path of the trolley so that it is redirected to a path where it will harm only one person, rather than five people. We compare this causal setup with a similar situation in which the intervention targets the patient. Imagine a situation in which one person is sitting in a bus at an intersection near the workers on the main track. An intervention moves this bus toward the main track, so that the bus pushes the workers off the track and they are in a safe location. Thus, the one bus passenger becomes a victim, in place of the five workers (i.e., patient intervention; Fig. 1, bottom panel). In these two scenarios, the actions are similar (changing the path of the trolley or of the bus). If one does not take into account the consequences, neither of these actions by itself (e.g., redirecting a trolley) is immoral. Moreover, the outcomes are identical (five people vs. one person harmed), which should lead to an identical utilitarian assessment. Also, both hypothetical interventions (throwing a switch or moving the bus) involve a causal interaction between two causes, the trolley and the intervention. Neither intervention is sufficient to cause death to the one victim in the absence of the other cause, the trolley. Thus, the interventions do not directly, by themselves, cause harm, as in the organ-transplant dilemma. Despite these similarities between the two scenarios, we hypothesized that subjects judge the agent intervention as more permissible than the patient intervention.

What could be the cognitive basis of this hypothesized effect? An abstract causal analysis suggests that the two situations are similar. What differs is the causal mechanism, or causal path, that generates the causal outcomes. How could this make a difference in moral judgment?
To assess the causal effect of an intervention, a causal analysis must compare the path of the object or person targeted by the intervention in the hypothetical presence versus absence of the intervention. Comparing effects in the presence versus absence of causes is a standard procedure in most theories of causal learning (see Shanks, Holyoak, & Medin, 1996). However, focusing on the presence or absence of causes and effects is not sufficient in the present case because the global outcomes are identical in the two versions of the dilemma.

The difference between the scenarios becomes salient if we compare causal paths. In the agent-intervention condition, the trolley heading toward potential victims (i.e., the agent) is the direct object of intervention. In the absence of an intervention, the trolley will continue on its original path and will kill five people. In case of an intervention, the trolley will be turned away from the five to a different path that will lead to the death of one person. This comparison between the moral consequences of the causal paths of the agent of harm in the presence versus absence of the intervention highlights the utilitarian contrast between the deaths of five people and the death of one person. According to the ethical intuitions of most people, five dead people is a worse outcome than one dead person, so people lean toward finding the redirection morally acceptable.

In contrast, in the patient-intervention condition, the bus with one person is the target of intervention. The comparison now focuses on the moral consequences of the intervention for this one person (i.e., the patient). This person will live in the absence of an intervention, but the intervention will lead to the certain death of this person. Moreover, in this case, an autonomous human being is moved and sacrificed without being asked to consent, which contradicts the moral values of most societies.

In sum, the general hypothesis is that people tend to focus on the causal paths of agents or patients targeted by an intervention, and neglect other causal processes occurring outside this focus, in the background. Whereas in the agent-intervention condition, the intervention highlights the two alternative paths with their moral consequences (one vs. five dead people), thus leaving no victims in the background, in the patient-intervention condition, the focus on the causal fate of the one victim leads to a neglect of other changes in the world, such as the death of the five people. These deaths lie outside the causal paths that are being compared when assessing the intervention and are therefore part of the background in the representation of this version of the dilemma. Thus, the focus on the action may sometimes lead to what we call intervention myopia, an attentional focus on the victims highlighted by the intervention, and a neglect of other victims located in the background. We are not saying that in cases of intervention myopia, people are completely blind to the victims in the background (i.e., the death of the five people); rather, we are saying that because of an attentional focus on the effects of interventions, people who are evaluating the morality of options may give victims in the background less weight than victims in the attentional spotlight.

**EXPERIMENT 1**

In this experiment, our aim was to test whether similar scenarios with identical outcomes but different points of intervention (agent vs. patient intervention) lead to different moral evaluations. More specifically, we hypothesized that in our scenarios, agent interventions would be considered more morally acceptable than patient interventions. We tested this hypothesis in four completely different content domains, which included dilemmas in terrorist, medical, and military settings.

**Method**

**Subjects and Design**

Sixty-four students from the University of Göttingen, Germany, participated in this experiment. The experiment was run at the beginning of lecture classes in different academic areas. Eight of the 64 subjects did not fill out the questionnaires and were therefore excluded from the analyses. Of the remaining 56 subjects, 2 responded only to two stories. These 2 subjects were included in the analyses. In a between-subjects design, four moral dilemmas were presented to each subject in a counterbalanced order (Latin square). Each subject received either the agent-intervention or the patient-intervention variants of the scenarios (summarized in Table 1). We did not manipulate conditions within subjects because pilot research showed that this would make the similarity of the dilemmas salient and lead to transfer of the moral evaluations between the agent- and patient-intervention versions, thus eliminating the predicted effect.

**Materials and Procedure**

Subjects were told that the study was about moral dilemmas and then received a booklet. The instructions on the first page said that the task was to read descriptions of several situations and to indicate in each case whether a person in the scene should choose to take a proposed action or should refrain from acting. Responses were made on a rating scale ranging from 1 (definitely not) to 6 (definitely yes). The following four pages described the scenarios and included a rating scale on each page. At the end of each story, the description pointed out that either the larger number or the smaller number of people would be seriously harmed or killed, depending on the bystander’s choice (i.e., fewer harmed if the bystander took action).

**Results**

Figure 2 displays the results based on 25 (bomb and virus scenarios) or 26 (trolley and torpedo scenarios) subjects in the agent-intervention condition and 29 (bomb and torpedo scenarios) subjects in the patient-intervention condition. For every scenario, a reversal was observed. That is, whereas subjects generally tended to recom-
The results support our hypothesis. Subjects were sensitive to the way the alternative outcomes were causally generated. Of course, for individual stories, there may have been additional factors that influenced subjects’ judgments. For example, in the bomb story, physical distance between the action and the harmful event varied (direct pushing of a person vs. hitting a person outside). Such variation can hardly be avoided when testing different scenarios, but the contrasted conditions in other scenarios did not vary in physical proximity and yielded virtually identical results.

**EXPERIMENT 2**

Dilemmas in which it is unavoidable that people are harmed present a general problem to deontology. An old solution to this problem that can be traced in Roman Catholic theology back to Aquinas is the *doctrine of the double effect*. The correct interpretation of this doctrine is under dispute, but according to one popular reading, harming people is permissible if actions with good intentions lead to unavoidable death as a foreseen, but not intended, side effect. In contrast, it is not right to intend bad things as a means to achieve good outcomes (Foot, 1967). Similarly, Kant (1785/1998) argued that human beings must never be used as means only.

In the stories we used in Experiment 1, people who had not been asked to consent were used as means in the patient-intervention versions. For example, in the case of the trolley dilemma, if the employee pushed the cars so that the bus with fewer passengers would move onto the track and replace the bus with a larger number of passengers, he was using the people in the pushed bus as a means to save the passengers in the other bus. In contrast, if the employee redirected the train, the resulting harm to the people on the side track would be a side effect of saving the people on the main track. (In the actual instructions, we mentioned trains instead of trolleys, which sounds more natural in German.)

Our second experiment, which was inspired by discussions in philosophy (Costa, 1987; Rakowski, 2001; Thomson, 1985), used the trolley scenario from the previous study to test the doctrine of double effect against our hypothesis that the point of intervention is the primary factor influencing people’s moral judgments. To test whether reluctance to use people as means was the driving force behind the effect in Experiment 1, we added to the standard variants of this scenario a third version in which the intervention targeted the agent but, unlike in the standard condition, the victims on the side track served as a
means to save other people. In this condition, the side track made a loop back to the main track (see Fig. 3). The train was again heading toward a bus with 10 passengers that, because of an accident, had stopped on the tracks. Doing nothing would hurt these 10 passengers. However, the train could be redirected to the side track, where a bus with 2 passengers was parked. This bus would stop the train, but the 2 passengers would be hurt. Note that without this bus, the train would go back to the main track and hit the bus with 10 passengers. Thus, the bus with 2 people served as a means to prevent harm to the bus with 10 people.

Whereas the intervention in the original agent-intervention story would lead to the death of fewer people as a side effect, the interventions in the original patient-intervention story and the new agent-intervention story with the looped track would use the fewer people in the second bus as a means to save the larger number of potential victims. As in Experiment 1, the locus of intervention was manipulated. In the patient-intervention story, the intervention would move people in the direction where they would be harmed, whereas in both agent-intervention stories, it was the train that would be moved to the side track. It is important to note that physical distance was kept constant. As in the previous experiment, the patient intervention did not move the bus with the two people directly, but moved it indirectly by pushing other cars. Thus, the distance between the intervention and harm was similar in the three conditions.

Method

Subjects and Design
One hundred twenty-three students from the University of Göttingen participated in this experiment and were randomly assigned to one of three conditions: agent intervention with harm to two people as a side effect (AI/S, n = 41), agent intervention with two people serving as a means to save more people (AI/M, n = 42), and patient intervention (PI, n = 40).

Materials and Procedure
The procedure was similar to the one in the previous study except that every subject read only one variant of the trolley dilemma. To ensure that the causal setup was salient in all three conditions, we added a figure illustrating the scenario and the two options (as in Fig. 3). The choice was always between harming 10 versus 2 people. The AI/S and PI versions of the trolley dilemma were taken from Experiment 1. In the AI/S scenario, the train could be redirected to a parallel side track, whereas in the PI scenario, a truck could push cars that would move a bus with 2 people in the way of the train. The new AI/M version differed from the AI/S version in that the side track looped back to the main track. Thus, hitting the bus with 2 people on the side track was a necessary condition for saving the bus with 10 passengers (see Fig. 3). So that the role of the 2 passengers as a means to an end would be salient, the instructions in the AI/M version explicitly mentioned that the redirected train would reach and harm the 10 passengers on the main track if the bus on the side track were not there. Each version of the story pointed out that in each option, the affected passengers would be seriously harmed.

Results
The results were clear. Although, unlike in Experiment 1, there was a general tendency to recommend action, the basic finding was replicated. Mean ratings were higher in the AI/S condition (M = 4.85, SD = 1.01) than in the PI condition (M = 3.73, SD = 1.62). Planned comparisons confirmed that this effect was significant, t(120) = 4.08, p_{rep} > .99, d = 0.83. Most important, ratings in the AI/S and AI/M (M = 4.74, SD = 1.01) conditions did not differ significantly, t < 1. As a consequence, the AI/M and PI ratings differed significantly, t(120) = -3.69, p_{rep} > .99, d = 0.75. Thus, the point of intervention, rather than whether or not people were used as means, seems to have been responsible for the effect.

The absence of a difference between the AI/S and AI/M conditions appears to contradict the findings of a similar study by Hauser, Young, and Cushman (in press), who also tested variants of the trolley dilemma with looping tracks. Responses in
their condition corresponding to the AI/M condition were intermediate between responses in their conditions corresponding to the AI/S and PI conditions. However, their findings may have differed from ours because of differences in the instructions that were confounded with the side-effect/means variable. The instructions in their variant of the AI/M condition but not in their variant of the AI/S condition were elaborated, as follows: “There is a heavy object on the side track. If the trolley hits the object, the object will slow the trolley down, thereby giving the men time to escape. Unfortunately, the heavy object is a man, standing on the side track with his back turned.” It seems likely that comparing a person to a heavy inanimate object and thus highlighting the person’s function as a stopping device makes his cruel death more imaginable. This may have contributed to the reluctance to recommend action in this condition (see Prinz, in press, for a related critique).

GENERAL DISCUSSION

The two studies reported here show that moral evaluations of situations with identical outcomes differ depending on the way these consequences are causally generated. The locus of intervention within the causal model proved to be the key factor. Subjects were more willing to accept harming a smaller number of people in order to save more people if the intervention was targeted at the agent rather than a patient. This effect cannot be explained by the heuristic not to use people as means; in Experiment 2, agent interventions were strongly recommended even when the victims served as means. This result does not rule out the possibility that most people find it immoral to use people as means, as patient interventions are a special case of using people as means. However, the finding shows that it is important to look at the way means-end relations are causally instantiated. The present experiments demonstrate that in their moral evaluations, people tend to focus on the causal path of the agent or patient targeted by their intervention and contrast it with the path of that agent or patient in the absence of this intervention. This narrow focus may lead to intervention myopia in some conditions (e.g., in our patient-intervention scenarios).

How does this finding relate to previous psychological theories of intuitions in moral dilemmas? Royzman and Baron (2002) adopted Quinn’s (1989) revised version of the doctrine of double effect and claimed that people have a preference for indirect over direct harm. It might be instructive to ask whether the difference between agent and patient interventions can be explained by varying degrees of directness. In a typical experiment of Royzman and Baron, subjects assessed a situation in which a missile had mistakenly been fired and was about to hit a civilian airliner. One option was to alter the path of a nearby private jet so that it would intercept the missile (direct harm). This situation is indeed similar to our patient-intervention scenarios. However, Royzman and Baron’s contrasting scenario presented the option of moving the civilian airliner behind the jet so that the jet, rather than the civilian airliner, would be hit (indirect harm). Not surprisingly, subjects preferred the second option. It is easy to see that the second scenario is crucially different from our agent-intervention scenarios. For example, in this version of our trolley dilemma, the individuals on the side track were directly harmed in the case of an intervention. In contrast, Royzman and Baron compared situations in which the target of intervention was either harmed (jet) or saved (airliner). The psychological difference between their situations can be easily explained within our framework. In the first situation, people focus on the jet that is harmed but that would have been spared in the absence of an intervention, whereas in the second situation, people focus on the airliner that is saved but would have been harmed in the absence of an intervention.

In summary, the locus of intervention is one key factor contributing to moral intuitions. Proximity, the causal structure underlying the moral dilemma, and the intentions of the agents (e.g., intended vs. unintended harm) are surely other important factors that eventually need to be integrated into a more complete theory of moral reasoning.

Acknowledgments—We thank T. Lombrozo for suggesting the expression intervention myopia and J. Baron and J. Prinz for comments on a previous draft.

REFERENCES


