Cognition 131 (2014) 28–43

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Transfer effects between moral dilemmas: A causal model theory

Alex Wiegmann *, Michael R. Waldmann

Department of Psychology, University of Göttingen, Germany

ARTICLE INFO

Article history:
Received 26 April 2013
Revised 15 December 2013
Accepted 17 December 2013

Keywords:
Moral judgment
Moral dilemmas
Causal models
Analogical mapping
Transfer effects

ABSTRACT

Evaluations of analogous situations are an important source for our moral intuitions. A puzzling recent set of findings in experiments exploring transfer effects between intuitions about moral dilemmas has demonstrated a striking asymmetry. Transfer often occurred with a specific ordering of moral dilemmas, but not when the sequence was reversed. In this article we present a new theory of transfer between moral intuitions that focuses on two components of moral dilemmas, namely their causal structure and their default evaluations. According to this theory, transfer effects are expected when the causal models underlying the considered dilemmas allow for a mapping of the highlighted aspect of the first scenario onto the causal structure of the second dilemma, and when the default evaluations of the two dilemmas substantially differ. The theory's key predictions for the occurrence and the direction of transfer effects between two moral dilemmas are tested in five experiments with various variants of moral dilemmas from different domains. A sixth experiment tests the predictions of the theory for how the target action in the moral dilemmas is represented.

1. Introduction

Some readers might recognize the following anti-piracy advertisement that has recently been shown in movie theaters. A person is sitting in front of a computer and about to start illegally downloading a film. In the following scene big letters appear on the screen, saying "You wouldn't steal a car", and you see a person trying to steal a car. A second later, you are told "You wouldn't steal a television", and you see the respective scene. Eventually, you read "You wouldn't steal a movie", and you can see a person running away after he has grabbed a DVD in a movie store. Back to the initial scene the text appears "Downloading pirated films is stealing".

The strategy behind this anti-piracy advertisement is clear. Showing several instances of stealing highlights one aspect of downloading a pirated film, namely the aspect of taking away someone's property. Highlighting a specific aspect of an action can be achieved in several ways: The easiest way is to simply point out the respective aspect. A more implicit, but possibly more successful strategy is to present analogous cases in which people's intuitions are more clear-cut than in the target dilemma. For instance, in the discussions about the moral status of abortions, anti-abortionists may present ultrasounds showing the heartbeat of embryos inside the womb, whereas pro-abortionists may point to the case of a pregnant woman that has been brutally raped. This argumentation strategy is chosen with the aim that people presented with the analogies will automatically transfer their moral intuitions onto the target case. If in the example above people agree that stealing is wrong in one of the clear cases, and they cannot point out why downloading pirated films significantly differs from stealing, they might feel committed to judge that downloading pirated films is wrong as well.

Research on moral judgment strongly suggests that there are indeed *transfer effects* (e.g., Horne, Powell, & Spino, 2013; Lanteri, Chelini, & Rizzello, 2008; Liao, Wiegmann, Alexander, & Vong, 2012; Lombrozo, 2009;

* Corresponding author. Address: Department of Psychology, Gosslerstr. 14, 37073 Göttingen, Germany. Tel.: +49 5513991250.
E-mail address: awiegma@uni-goettingen.de (A. Wiegmann).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.12.004

Petrinovich & O'Neill, 1996, Wiegmann, Okan, & Nagel, 2012). Such transfer effects have been found to be robust and strong, sometimes even counteracting other factors known to influence moral judgments (Wiegmann et al., 2012). The aim of the present article is to propose a new theory of transfer effects between moral dilemmas that focuses on highlighting of components of causal models.

To test theories of transfer, cases that demonstrate that transfer can be asymmetric are particularly informative. In the following section we will report recent studies showing that sometimes transfer only occurs when moral dilemmas are ordered in a specific sequence, but not when the sequence is reversed. This is an interesting finding because transfer effects are often justified as attempts of reasoners to be consistent across different situations. However, while attributing the goal to be consistent may be a plausible hypothesis when transfer is observed, this does not explain why transfer occasionally is absent in the opposite order. Different theories have been proposed but so far the effect of asymmetric transfer resists a convincing explanation.

Asymmetries of transfer are not only interesting as test cases for cognitive theories, they are also important for predicting how intuitions influence each other outside the laboratory. A politician, for example, may think about how she can influence the intuitions of her constituents about privacy issues or about a military invasion. A poll about several political propositions may yield different results depending on the ordering of the votes. Parents may think about how they can influence intuitions of their children about theft on the Internet, or advertising agencies may work out strategies on how to convince buyers to buy more expensive organic food items. The research on asymmetric transfer indicates that it will be of crucial importance to pick the right cases in the right order to make the analogies work. Our research will focus on specific cases of asymmetrical transfer that are designed to test between competing theories.

1.1. Asymmetrical transfer effects between intuitions about trolley dilemmas

Most of the studies in which transfer effects were found investigated trolley dilemmas. Trolley dilemmas have been extensively discussed in moral philosophy, which has stimulated various empirical studies in moral psychology (see Gräfenhain & Wiegmann, 2012, chap. 81; Waldmann, Nagel, & Wiegmann, 2012, for overviews). In trolley dilemmas, an out-of-control train is threatening a group of people who are about to die if nothing is done to stop the train. To save this group, a bystander could intervene at the cost of the death of another person who would otherwise not be in danger.

Push and *Switch* are the two best-known trolley variants discussed in both the philosophical and psychological literature (e.g., Foot, 1967; Thomson, 1985).[1] In both dilemmas three people are threatened by an out-of-control train. In Push, the only possibility to save the three persons is to throw a heavy person from a bridge in front of the train, resulting in the death of the heavy person but saving the three (Thomson, 1985). In Switch, the threatening train can be redirected away from the three onto another track where one different person would die in the collision with the train (Foot, 1967). Research in moral psychology has shown that the majority of people disapprove of intervening in Push, whereas they favor an intervention in Switch (e.g., Bartels, 2008; Cushman, Young, & Hauser, 2006; Greene et al., 2009; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Sloman, Fernbach, & Ewing, 2009; Waldmann & Dieterich, 2007; Waldmann & Wiegmann, 2010; see Waldmann et al., 2012, for an overview).

Interestingly, transfer effects for these two moral dilemmas have been shown to be asymmetrical (e.g., Lanteri et al., 2008; Lombrozo, 2009; Petrinovich & O'Neill, 1996; Wiegmann et al., 2012). Presenting Push before Switch affects people's judgment for the proposed action in Switch: In this condition, subjects are less likely to approve of the proposed action in Switch than when being confronted with Switch alone. However, presenting Switch before Push does not change people's judgment for the proposed action in Push. Explaining this asymmetry is a key challenge for every theory of transfer effects between evaluations of moral dilemmas.

2. Selective highlighting within causal models: a theory of transfer effects

We will introduce our theory in four stages: First, we will specify the scope of our theory. Second, we will present its general spirit by using an ambiguous image as an analogy. Third, we will describe the core components of our theory. Fourth, we will outline the predictions of our theory, and show how it handles the asymmetrical transfer effect between moral dilemmas, such as Push and Switch.

2.1. The scope of our theory

The target domain of our theory is moral dilemmas in which potential victims are threatened by physical (often deadly) harm. The principal goal is to explain transfer effects between harm-based moral dilemmas that are consecutively presented and individually evaluated. Apart from these constraints, the scope of our theory is not limited to specific kinds of dilemmas, for example trolley dilemmas. In Section 8.2 we will discuss possible extensions of our theory.

2.2. The ambiguous image analogy

To illustrate the intuition motivating our theory, we would like to use an ambiguous image as an analogous case. Fig. 1 illustrates an asymmetrical transfer effect (adapted from Medin, Goldstone, & Gentner, 1993). Most people perceive four prongs if they look at the image on the left side first, but if they see the left image after having been shown the right image first, they instead see three

[1] Conventionally, these two dilemmas are labeled "Footbridge dilemma" and "Trolley dilemma", respectively. Since both dilemmas involve trolleys, the conventional labels are somewhat misleading so that we will instead use the terms "Push" and "Switch". In most of our experiments, three individuals (instead of the usual five in the philosophical literature) can be saved.

prongs. Thus, the perception of the number of prongs in the left image is *ambiguous*, whereas the object on the right side is *unambiguous*.

We are using this analogy to get a better understanding of the asymmetric transfer effects in moral dilemmas. Previous research on moral dilemmas has demonstrated that the action proposed in Push is generally perceived as clearly wrong, regardless of what other dilemmas have been evaluated before. Thus, Push seems to have the characteristics of an *unambiguous moral dilemma*. In contrast, the proposed action in Switch, if considered in isolation (i.e., without prior consideration of another dilemma), is generally judged to be morally right (see Waldmann et al., 2012, for an overview). Moreover, previous research has shown that moral intuitions for Switch are highly context sensitive (Di Nucci, 2013; Wiegmann et al., 2012). Thus, Switch seems to play the role of an *ambiguous moral dilemma*.

2.3. Causal structures of moral dilemmas

How can the claim be theoretically motivated that Push seems unambiguous and Switch ambiguous? Our theory focuses on the underlying causal structures of these two dilemmas (cf. Kamm, 2007; Mikhail, 2011; see Fig. 2).

We will first consider the causal structure of Switch, which serves as an example of an ambiguous dilemma. Performing the proposed action in Switch leads to two outcomes – saving three persons and causing the death of one person. The negative outcome does not lie on the same causal path as the good outcome (see also Kamm, 2007; Mikhail, 2011; Timmons, 2002). To illustrate why causing the death of the one person does not lie on the causal path saving the three, imagine that the threatening train would stop for some reason shortly after being redirected but before reaching the one person. In this case, the one person would not get killed but the three persons would still be rescued.

The feature that the different outcomes lie on different causal paths allows for *selective highlighting* of the causal relationship between the intervention and the good outcome. If this path is highlighted, the aspect of saving becomes salient in this dilemma. Analogously, the good outcome does not lie on the causal path leading from the intervention to the bad outcome. If this path is selectively highlighted, the aspect of killing dominates. In sum, the causal structure of Switch allows for selective highlighting of the causal relationship between the intervention and either the good or the bad outcome; as a consequence, the other causal relation is backgrounded. The causal model underlying this dilemma is the key reason for the postulated ambiguity of Switch.

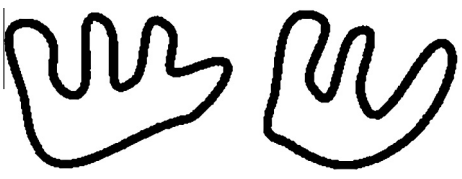

Fig. 1. Ambiguous (left) and unambiguous (right) images. Demonstration of asymmetrical mapping effects (material adapted from Medin et al., 1993).

In contrast to Switch, intervening in Push first leads to the death of the one person, whose death lies on the causal path leading to the saving of the three potential victims (see also Foot, 1967; Kamm, 2007; Knobe, 2010; Mikhail, 2011; Sloman et al., 2009; Waldmann & Dieterich, 2007; Waldmann & Wiegmann, 2010). Without the involvement of the one person, there is no way to stop the threatening train. In Push, the one person is harmed as a means to save the three. Due to this causal structure, and in contrast to Switch, there is, apart from ignoring the causal model, no way to represent the causal path leading from the intervention to the good outcome without passing the bad outcome. Hence, the causal structure of Push does not permit selective highlighting of the causal relationship between the intervention and the good outcome. It is, however, possible to describe the causal path leading from the intervention to the death of the one person without attending to the final good outcome. If the initial path is selectively highlighted, the aspect of killing dominates. In summary, the causal structure of Push allows for selective highlighting of the causal relationship between the intervention and the bad but not the good outcome. This feature makes Push, as well as structurally similar dilemmas, unambiguous according to our terminology.

2.4. Default evaluations and highlighting

While most people disapprove of the proposed action in Push, they tend to approve of the proposed action in Switch if this dilemma is judged in isolation, or as the first of several dilemmas (see Waldmann et al., 2012). This finding indicates that per default the causal path leading from the intervention to the good outcome is highlighted in Switch, which leads to its positive evaluation. According to our terminology, dilemmas like Switch are *positive dilemmas*. Analogously, if most people disagree with an action in a specific dilemma, as in Push, this empirical finding indicates that the causal path leading from the intervention to the bad outcome is highlighted. Such dilemmas are *negative dilemmas*, according to our terminology. Numerous theories compete for explaining these different default evaluations (see Waldmann et al., 2012). For our present purposes it suffices to state that the default evaluations differ in these two types of dilemmas.

2.5. Selective highlighting and mapping

If two moral dilemmas are presented consecutively, the highlighted causal path in the first dilemma can in some cases be *analogically mapped* onto the causal structure of the second dilemma so that the same causal path becomes selectively highlighted in the second scenario. For such a transfer to occur, the causal structure of second dilemma must allow selective highlighting of the same causal path that is highlighted by default in the first dilemma.

For instance, if the causal path leading from the intervention to the bad outcome is highlighted in the first dilemma, analogical mapping is only possible if this path

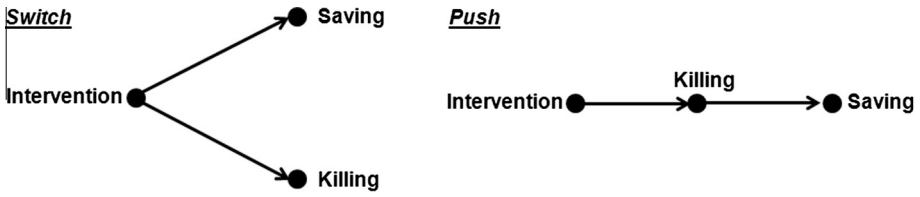

Fig. 2. The causal structures of Switch (left) and Push (right).

can selectively be highlighted in the second dilemma. This is only possible if there is a causal path in the second dilemma that leads from the intervention to the bad outcome without containing the good outcome.

If the second dilemma is an ambiguous dilemma, analogical mapping from the first dilemma is possible, regardless of whether the first dilemma highlights the good or the bad outcome by default, because an ambiguous dilemma allows for selective highlighting of the causal relationship between the intervention and either the good or the bad outcome.

Analogical mapping is also possible if the second dilemma is an unambiguous dilemma, but only in cases in which the path highlighted by default is the same in the two dilemmas (see the following section for why in these cases judgments for the second dilemma are often not expected to be altered). In contrast, if the first dilemma is an unambiguous positive one and the second dilemma an unambiguous negative one, for example, analogical mapping is not possible. In this case, the second dilemma does not permit selective highlighting of the path leading from the intervention to the good outcome as this path necessarily contains the negative outcome.

2.6. Changing people's intuitions

An important consequence of mapping and selective highlighting is the possibility of a change of the moral intuitions of people. Two conditions have to hold in order to make such a transfer effect between two dilemmas likely. First, the aspect of the first dilemma that is highlighted by default can be analogically mapped onto the causal structure of the subsequent dilemma. Second, the default evaluations of the two dilemmas differ significantly. If these two conditions are met, we predict that the first dilemma will affect the evaluations of the second dilemma by shifting the evaluation of the second dilemma towards the evaluation of the first dilemma. The strength of the predicted transfer effect depends on how strongly the default evaluations of the two dilemmas differ. If the difference is small, there is little room for change.

2.7. Specific predictions

In this section we will first show how our theory handles the asymmetrical transfer effect between moral dilemmas, such as Push and Switch. Typically transfer is observed when Push precedes Switch, but not when Switch precedes Push. This pattern is predicted by our theory. Whereas Push is an unambiguous dilemma that highlights the bad outcome per default, Switch is an ambiguous dilemma and hence potentially subject to selective highlighting. Thus, the first condition for mapping holds in this sequence. Moreover, there is also room for change. Whereas Push is per default evaluated negatively, Switch is per default evaluated positively. Thus, the second condition also holds. Hence, we expect that people will change their evaluation of Switch after having read about Push.

No such change is predicted when Switch precedes Push, however. Push is unambiguous, and does not allow selective highlighting of its good outcome, which is per default highlighted in Switch. Thus, Switch cannot be mapped onto Push. Under this condition, no change in the moral evaluation of Push, the second dilemma, is expected (see Fig. 3).

So far we have concentrated on cases in which the order of presentation of an ambiguous positively evaluated and unambiguous negatively evaluated dilemma is being varied. These cases are particularly interesting because they demonstrate asymmetry of transfer effects between moral dilemmas. However, other combinations which provide interesting test cases for our theory are possible.

Push constitutes an unambiguous negative dilemma. To examine an unambiguous positive dilemma, we will use a positive version of Push, hereafter called PosPush. In PosPush, three persons are standing on the tracks and threatened by an out-of-control train. Further down the tracks a single person is standing. If no action is taken, the train will kill the three persons who would stop the train with their bodies. In this case, the single person further down on the track would live. The only possibility to avoid the death of the three persons is to push them off the tracks, which would save them but kill the one. In PosPush the saving of the three persons is necessary for the death of the one person and precedes it; in Push it is the other way around. Hence, in PosPush only the path leading from the intervention to the good outcome can be selectively highlighted, which makes PosPush an unambiguous dilemma. Since (based on previous studies) we expect that most people agree with the proposed action in PosPush, it is a positive dilemma, according to our terminology (see also Royzman & Baron, 2002).

What does our theory predict if PosPush is presented either in the first or second position? If PosPush is preceded by a significantly more negative dilemma, a transfer of this negative surplus is not possible because the causal structure of PosPush does not allow for selective highlighting of its negative outcome. Hence, the first condition for change is not met. If PosPush is preceded by a dilemma with a similar evaluation, transfer may be possible if the same causal components can be selectively highlighted but the second condition is not met. Finally, if the

evaluation of the first dilemma is significantly more positive than the evaluation of PosPush, the two conditions for transfer could be met. However, since the evaluation for PosPush is expected to be positive already, such a constellation may be hard to establish due to a possible ceiling effect. Therefore, PosPush is not likely to be influenced by a preceding dilemma. By contrast, when it is placed in the first position, it should be capable of influencing the second dilemma if the two conditions for transfer are met.

A further possible type is an ambiguous dilemma that is per default negatively evaluated. We are not aware of the existence of such a dilemma so that this possibility only constitutes a theoretical one. However, if there was such a dilemma, predictions from our theory could be derived.

2.8. Overview of experiments

In the first four experiments we tested the predictions of our theory for cases in which two moral dilemmas are presented one after another. We explored under what conditions transfer effects between moral judgments are to be expected. More specifically, Experiment 1 set the stage by replicating the basic effect of asymmetry of transfer between Switch and Push. We predicted that Push will influence Switch but not vice versa. Experiment 2 generalized this finding to the case in which Switch is preceded by a dilemma that is unambiguous and more positive than Switch (which already is positive). In Experiment 3 an unambiguous positive (PosPush) and an unambiguous negative dilemma (Push) were paired. According to our theory, no transfer effects are predicted in this combination of cases. Finally, in Experiment 4 we aimed to show that asymmetrical transfer effects do not exclusively occur between trolley dilemmas but also between dilemmas from other domains. Experiment 5 served to rule out an alternative hypothesis for asymmetrical transfer that focuses on the fact that the levels of confidence for agreeing with the proposed action typically differ between Switch and Push. According to this alternative account of asymmetrical transfer, moral intuitions about proposed actions should be more prone to transfer effects when subjects' confidence hovers around an intermediate level (which is typically the case in Switch) than when confidence is high (as in Push). Intermediate ratings leave more room for change than ratings at one of the extremes (see also Schwitzgebel & Cushman, 2012).

Whereas the first five experiments tested our theory by exploring different variants of transfer effects on moral judgments, Experiment 6 investigated a more fine-grained implication of our theory by testing our assumption that Switch is represented as ambiguous and can be both represented as a case of killing or saving depending on the prior dilemma with which it is paired. In this experiment, we also paired unambiguous dilemmas with different default evaluations with each other to show that in these situations people's representation as killing or saving will not be changed.

3. Experiment 1

The first experiment set the stage by attempting to replicate the key finding predicted by our theory, namely asymmetric transfer, in an online population. The moral dilemmas used were the standard Switch and Push paradigms. We predicted that presenting people first with Push would affect their judgment for Switch while presenting people first with Switch would not affect their judgment for Push.

3.1. Method

3.1.1. Participants

197 subjects, each receiving £0.50, were recruited via an online database located in the U.K. They were invited via an email. The email contained a link that directed subjects to the experiment.

3.1.2. Design, materials, and procedure

The experiment was conducted on the Internet. Upon clicking on a link which subjects received via e-mail, they were redirected to a website containing the experiment. Subjects first read general instructions. These familiarized them with the rating scale, asked them to read the following dilemmas carefully, and appealed to them to take the task seriously. This part of the procedure was the same for all experiments. Subsequently, subjects were randomly assigned to one of two conditions: In *Push_Switch* participants were first presented with Push and then with Switch, in *Switch_Push* it was the other way around. Both scenarios were accompanied by a graphical illustration of the initial situation.

After reading about each dilemma, participants were asked whether the proposed action (i.e., redirecting the runaway train in Switch; pushing the person from the bridge in Push) should be carried out. To express their judgment, participants were asked to mark one point on a 6-point Likert scale ranging from 1 (''certainly no'') to 6 (''certainly yes''). After having judged the second dilemma, participants were asked some demographic questions, and were given a very simple logical question to identify participants who did not pay sufficient attention to the task.

3.2. Results and discussion

25 participants were excluded from the analyses because they did not finish the experiment, finished it in less than 40 s, or failed to answer the simple logical question correctly. As can be seen in Fig. 4, we managed to replicate the asymmetrical transfer effect. While the ratings for Push did not differ significantly depending on whether it was presented first ($M = 2.4$; $SD = 1.45$) or second ($M = 2.45$, $SD = 1.47$), $t(170) = .2$, the ratings for Switch were significantly decreased when Switch was presented second ($M = 3.51$, $SD = 1.53$), as compared to ratings for Switch when presented first ($M = 4.40$, $SD = 1.27$), $t(170) = 4.18$, $p < .0001$. The relevant two-way interaction (type of dilemma (Switch vs. Push) $\times$ order of presentation (Switch first vs. Push first)) was also significant, $F(1,170) = 11.27$, $MSE = 15.24$, $p < 0.001$.

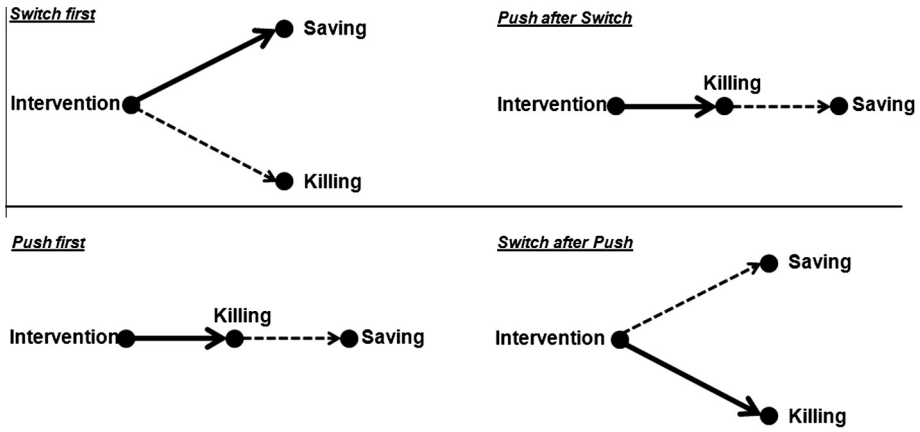

Fig. 3. The highlighted path of the causal structure of Switch and Push in the default evaluation (left side) vs. when preceded by the other dilemma (right side). Bold solid lines represent the highlighted part of the causal structure.

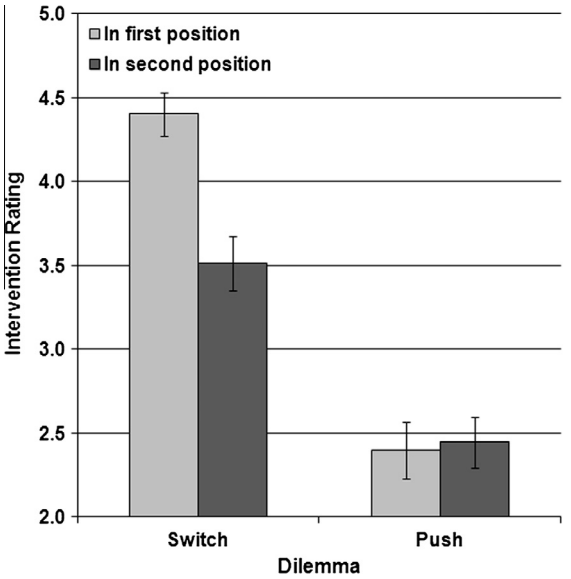

Fig. 4. Ratings for Switch and Push as a function of the order of presentation. The scale ranged from 1 ("one should certainly not do the proposed action") to 6 ("one should certainly do the proposed action"). Error bars represent standard errors of means.

4. Experiment 2

While we showed in Experiment 1 that participants' level of approval of the proposed action in Switch can be decreased if Switch is preceded by Push, our aim in Experiment 2 was to demonstrate that participants' level of approval for Switch is not only diminished by a previous more negatively evaluated dilemma, but can also be increased by a previous dilemma that is rated more positively than Switch. Since the default evaluation for Switch is already fairly high, finding such a dilemma is more difficult than finding a dilemma for which people's level of approval of the proposed action is lower (as it is, for example, the case in Push). We have used a dilemma labeled PosPush (see also Section 2.7) in which saving three persons from being run over by a train can be achieved by pushing them off the tracks. The consequence of this intervention is that another person further down the tracks who had not been in danger before will then lose his life. In this dilemma, the primary action is to save three people by pushing them out of harm's way. Unlike in Switch, in which killing is a direct consequence of saving, PosPush harms the one person at the end of the tracks in a more indirect way. Therefore we expected slightly more positive evaluations for PosPush than for Switch (see Royzman & Baron, 2002). The transfer effect should be relatively small, though, because both dilemmas are typically rated positively on the rating scale. In sum, our theory predicts that participants' evaluation for Switch will probably become more positive, but not as positive as the default ratings for PosPush.

4.1. Method

4.1.1. Participants

157 subjects, each receiving £0.50, were recruited via an online database located in the U.K.

4.1.2. Design, materials, and procedure

Participants were randomly assigned to one of two conditions. In *PosPush_Switch* participants were first presented with PosPush and then with Switch, in *Switch_PosPush* it was the other way around. Both descriptions were accompanied by an illustration of the initial situation. Otherwise the same procedure and test question were used as in the first study.

4.2. Results and discussion

22 participants were excluded from the analyses by the same criteria as in Experiment 1. The results for the remaining subjects are summarized in Fig. 5. Analogous to Push in the previous experiment, the evaluation for PosPush did not differ significantly depending on whether it was presented first ($M = 4.87$; $SD = 1.39$) or second ($M = 4.75$, $SD = 1.11$), $t(133) = .53$, $p = .59$. By contrast, ratings for Switch were lower when presented first

($M = 4.28$, $SD = 1.21$), than when this scenario was preceded by PosPush ($M = 4.69$, $SD = 1.34$). This difference was significant if a one-tailed test is used, $t(133) = 1.86$, $p = .07$ (two-tailed). Given that we predicted a small effect in a specific direction, this statistical test seems appropriate. The relevant interaction (dilemma × order of presentation) was also marginally significant, $F(1, 133) = 2.8718$, $MSE = 1.43$, $p = .09$.

In sum, we found evidence for two predictions of our theory. The unambiguous PosPush is not affected by the previous dilemma, similar to Push in the previous experiment. This finding supports the view that not the valence but rather the degree of ambiguity is the basis of transfer effects. Moreover, we found a small effect showing that not only a negative dilemma can diminish ratings for an ambiguous dilemma, but also that a positive dilemma can have an effect in the opposite direction. The small effect size is predicted by the second criterion of our theory because both paired dilemmas generated evaluations that were positive, and only weakly different.

5. Experiment 3

An important prediction of our theory, setting it apart from other theories, is that moral intuitions about an unambiguous dilemma should only be affected by a previous dilemma if the same causal path is highlighted and if the default evaluations differ in both dilemmas (see also General Discussion). Otherwise, no transfer effect is predicted to occur, even if the standard evaluations of the involved dilemmas differ strongly. To test the prediction that transfer is blocked when different causal paths are highlighted despite different default evaluations, we paired two unambiguous dilemmas with these characteristics in Experiment 3, Push (negative) and PosPush (positive). No transfer effect is predicted regardless of order in these pairings.

5.1. Method

5.1.1. Participants

182 subjects, each receiving £0.50, were recruited via an online database located in the U.K.

5.1.2. Design, materials, and procedure

Participants were randomly assigned to one of two conditions. In *Push_PosPush* participants were first presented with Push and then with PosPush, in *PosPush_Push* it was the other way around. Both conditions were accompanied by an illustration of the initial situation. Otherwise the same procedure and test question were used as in the previous experiments.

5.2. Results and discussion

44 participants were excluded from the analyses by the same criteria as in the previous studies. The results for the remaining subjects are summarized in Fig. 6. As predicted by our theory, no transfer effects were observed regardless of order. The difference for Push when presented first ($M = 2.08$, $SD = 1.28$) vs. second ($M = 2.20$, $SD = 1.53$) was not significant, $t(136) = .48$, $p = .63$. Similarly, there was no effect for PushPos when presented first ($M = 4.27$, $SD = 1.35$) vs. second ($M = 4.34$, $SD = 1.39$), $t(136) = .30$, $p = .77$.

It is worth noting how similar Push and PosPush are on the surface. If one considers the two graphical illustrations accompanying Push and PosPush, the only differences between the two scenarios seem to be the missing bridge in PosPush and that the one person has switched places with the three persons. Moreover, in both dilemmas people

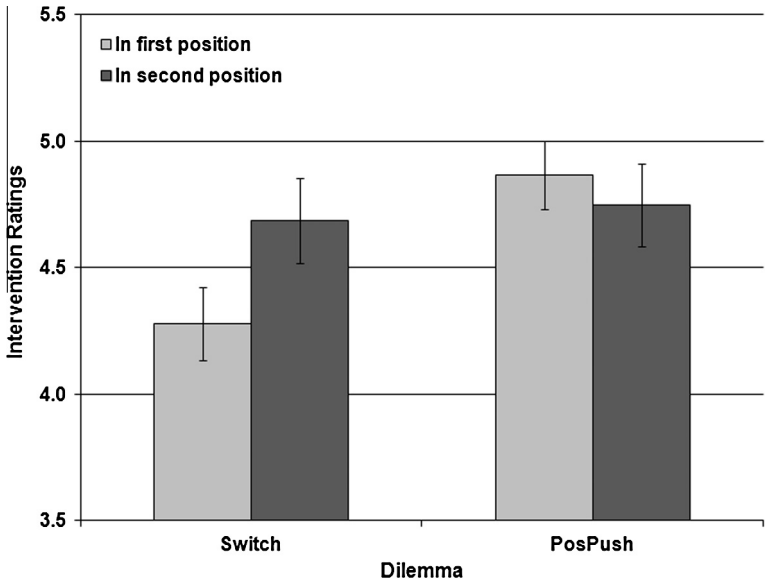

Fig. 5. Ratings for Switch and PosPush as a function of the order of presentation. The scale ranged from 1 (“one should certainly not do the proposed action”) to 6 (“one should certainly do the proposed action”). Error bars represent standard errors of means.

are physically pushed. The key difference leading to the different default evaluations is the direct outcome of the push action (saving vs. harming). Yet, despite the high discrepancy in their default evaluations, the two dilemmas did not affect each other.

6. Experiment 4

A critic of our experiments might counter that so far we have only tested the predictions of our theory with trolley dilemmas, which means that we cannot rule out that some specific features of trolley dilemmas have caused the effects. To test the generality of our theory, in Experiment 4 we therefore used unambiguous and ambiguous dilemmas from three different non-trolley domains. The distinction between ambiguous and unambiguous dilemmas is again based on the underlying causal model that either allows highlighting of separable causal paths (ambiguous) or combines the positive and negative outcomes on a single path. As in the first two experiments, we decided to use unambiguous dilemmas that per default are evaluated negatively, whereas the ambiguous ones are evaluated positively. This way there is enough room for a possible transfer effect. The key prediction again is that unambiguous dilemmas should influence ambiguous ones, but not vice versa.

The experiment also allowed us to test whether there are transfer effects across different domains. So far we have demonstrated asymmetrical transfer within a single domain, namely trolley dilemmas. In Experiment 4 we paired both ambiguous and unambiguous dilemmas from three different domains within each domain and between the different domains. Finding asymmetrical transfer across different domains would strengthen the theory that it is the mapping of the underlying causal models rather than superficial domain features that drive the predicted transfer effects.

6.1. Method

6.1.1. Participants

643 subjects, each receiving £0.50, were recruited via an online database located in the U.K.

6.1.2. Design, materials, and procedure

Three stories – *Burning House*, *Sharks* and *Fumes* – were used. There were two versions of each story, an unambiguous negative and an ambiguous positive one. In all stories, the proposed action resulted in the saving of three people and the death of one other person.

In *Burning House*, three people are trapped in a burning building. There is only one emergency exit through which all of them could escape to safety, but burning debris blocks the exit. Martin, a bystander in safe distance, notices a fourth person in the hallway leading to the exit; this person has been injured but is about to crawl to safety through a small hole at the bottom of the exit door. The three people do not have time to climb through the small hole. In the unambiguous negative version, Martin realizes that he could grab the injured man and use his body as a battering ram to break through the burning blockage in the hallway that is preventing the escape of the three people. The three people would be saved but the injured person would die. In the ambiguous positive version, the hallway's emergency system is capable of putting out fire by eliminating oxygen from the hall, and Carl could activate the system by pressing a nearby button. Again the three people would be saved, but the injured person would suffocate and die.

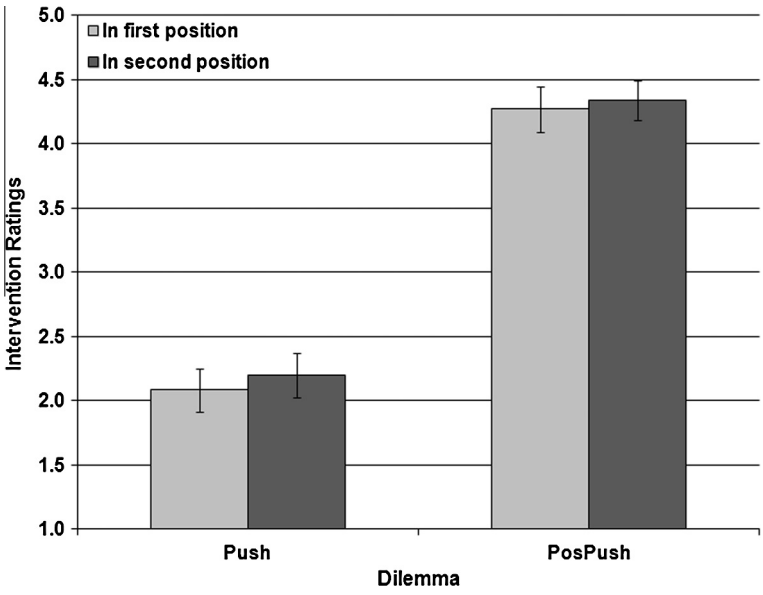

Fig. 6. Ratings for Push and PosPush as a function of the order of presentation. The scale ranged from 1 (''one should certainly not do the proposed action'') to 6 (''one should certainly do the proposed action''). Error bars represent standard errors of means.

In *Sharks*, a shark threatens three people. In the unambiguous negative version, a bystander could shoot a fourth person standing on the pier which would cause him to fall into the water where he would be eaten by the shark. In the ambiguous positive version, the three people can be rescued by a bystander making loud noises, which would divert the shark into another direction where a fourth person is swimming whom the shark will then kill.

In *Fumes*, poisonous gas is emitted from a room in a hospital, which soon would kill three patients in the next room. In the unambiguous negative version, the three patients could be rescued by pulling a fourth patient with a rare disease into the room with the gas because this patient has a rare blood type that allows absorbing the gas through his skin. In the ambiguous positive version, a switch controlling the ventilation system could be flipped so that the gas can be directed away from the room with three patients into a room with a different patient.

Each ambiguous positive version was paired with each unambiguous negative version with the order of presentation being counterbalanced. Hence, there were eighteen conditions in total to which participants were randomly assigned. Each subject was only exposed to one of these eighteen conditions. Otherwise, the same procedure and test question were used as in the previous experiments.

6.2. Results and discussion

111 participants were excluded from the analyses using the same criteria as in the previous experiments. Fig. 7 shows that the pattern of judgments for the pairings of the ambiguous and unambiguous dilemmas was qualitatively similar in the different domains with the ratings for ambiguous dilemmas being more influenced by the preceding dilemma than the ratings for the unambiguous dilemmas. The figure shows 12 conditions because we collapsed over the domains with which the dilemmas shown on the *X*-axis were paired (see Appendix for the full information about all pairings). In order to see the transfer patterns when pairs came either from the same or different domains, Fig. 8 shows the ratings for the two kinds of dilemmas as a function of their position in the sequence and the equality of the domain (same vs. different). The figure shows a qualitatively similar pattern regardless of whether scenarios were paired that came from the same or from different domains. This impression is supported by an analysis of variance which yielded no significant three-way interaction (ambiguity × sameness of domain × order of presentation), $F(1,528) = 2.83$, $MSE = 3.66$, $p = .09$.

For the follow-up analyses we collapsed the data from the three domains. For unambiguous negative dilemmas, it did again not make a significant difference whether they were shown first ($M = 2.60$, $SD = 1.62$) or after an ambiguous positive dilemma ($M = 2.69$, $SD = 1.62$), $t(514) = 0.63$; $p = .53$. However, when an ambiguous positive dilemma was presented first ($M = 3.87$, $SD = 1.37$), its rating was significantly higher than when it was presented after an unambiguous negative dilemma ($M = 3.33$, $SD = 1.48$), $t(514) = 4.35$; $p < 0.0001$. The interaction (ambiguity × order of presentation) was significant, $F(1,530) = 10.45$, $MSE = 16.88$, $p < .01$.

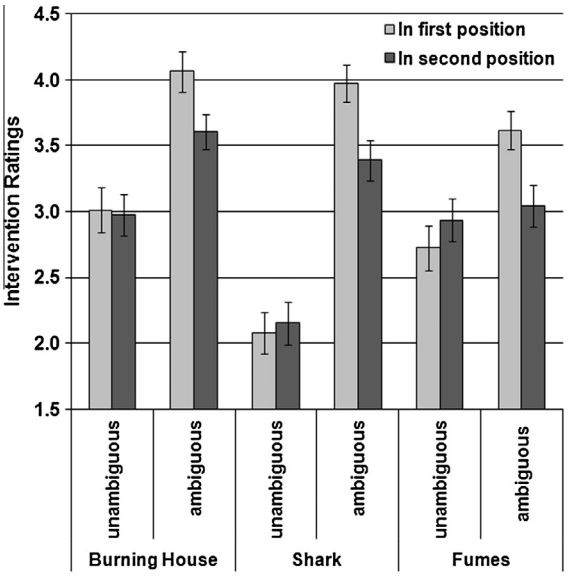

Fig. 7. Ratings for the unambiguous negative and ambiguous positive dilemmas as a function of the order of presentation. The scale ranged from 1 ("one should certainly not do the proposed action") to 6 ("one should certainly do the proposed action"). Error bars represent standard errors of means. For example, the second bar from the left represents the mean rating for the unambiguous Burning House when preceded by an ambiguous dilemma from any of the three domains.

These results add two important new findings to the previous studies. First, they demonstrate asymmetric transfer in domains different from trolley dilemmas. Furthermore they provide further strong support for our theory by showing that transfer can be obtained both within and between domains. Apparently the key driving force predicting the presence or absence of transfer in our scenarios is the underlying causal models rather than surface features.

7. Experiment 5

A critic[2] might point out that transfer is only observed in our experiments when the ratings are in the midsection of the rating scale, but not when they are more extreme (see also Schwitzgebel & Cushman, 2012; General Discussion). Thus, a possible alternative account might be that transfer depends on the confidence in the judgments. Judgments for Push, for example, elicit high confidence in the wrongness of the proposed action, which might make this scenario immune to change, whereas judgments for Switch express medium confidence, which might offer more room for change. One way to test this hypothesis against our theory is to modify the scenarios in a way that shifts Switch towards the extreme, and Push towards the midsection. Whereas our theory would still predict transfer from the

[2] We thank one anonymous reviewer for bringing this possible alternative explanation to our attention, and for suggesting the experiment.

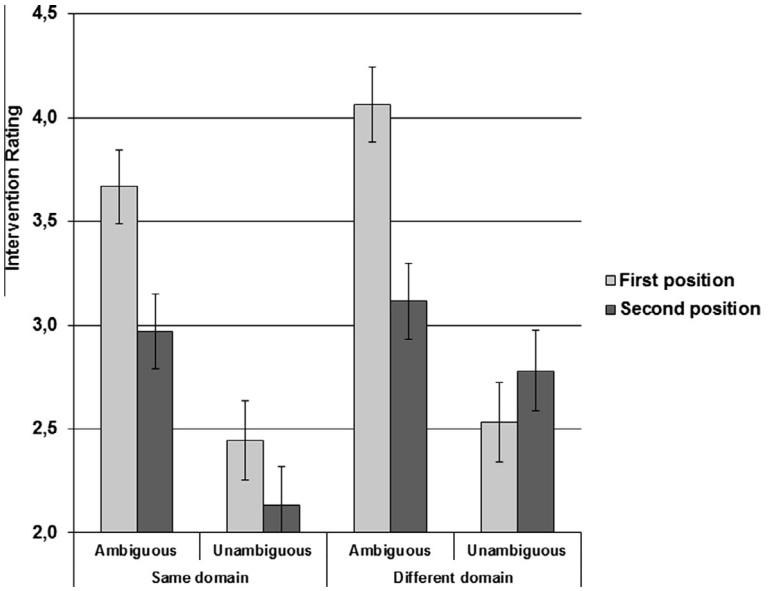

Fig. 8. Ratings for the unambiguous negative and ambiguous positive dilemmas as a function of the order of presentation. The scale ranged from 1 (“one should certainly not do the proposed action”) to 6 (“one should certainly do the proposed action”). Error bars represent standard errors of means. The data from different domains were collapsed.

unambiguous to the ambiguous scenario but not versa, the confidence account would predict a reversal. According to this theory, Switch should now influence Push but not the converse. To manipulate confidence, we manipulated the number of people that are saved by the action (see Bartels, 2008; Nichols & Mallon, 2006, for related manipulations). Saving 1000 people by killing one should increase confidence in the permissibility of the action in both the Push and the Switch scenarios. To test whether increasing the number of victims to 1000 indeed shifts confidence in the direction of more positive ratings we also included the standard conditions with three victims.

7.1. Method

7.1.1. Participants

314 subjects, each receiving £0.50, were recruited via an online database located in the U.K.

7.1.2. Design, materials, and procedure

Participants were randomly assigned to one of four conditions, namely Push_Switch, Switch_Push, Push1000_Switch1000, or Switch_1000_Push_1000, with the ordering within the labels signaling the position in the sequence. In the first two conditions three victims can be saved, whereas in the two other conditions 1000 people are threatened.

7.2. Results and discussion

70 participants were excluded from the analyses for the same reasons as in the previous experiments. As can be seen in Fig. 9, our manipulation of the number of victims was successful. The main effect of number of saved victims (3 vs. 1000) was significant, $F(1,310) = 10.489$, $MSE = 39.83$, $p < 0.01$. Follow-up tests showed that the difference between the default ratings (i.e., ratings for the dilemmas in the first position) of the two Push versions (3 vs. 1000) was significant $t(310) = 2.44$, $p < .05$. Moreover, the default ratings for the two Switch versions were also significantly different when a one-tailed test was applied, $t(310) = 1.78$, $p = .08$. Thus, increasing the number of victims led to an upward shift, resulting in ratings for Push being close to the midpoint and ratings for Switch being closer to the positive end of the scale. However, contrary to the prediction of the confidence account, no reversal of the transfer effect was observed.

Fig. 9 shows that the typical asymmetrical transfer effect was obtained for the original as well as for the modified versions of Push and Switch. The interaction (ambiguity $\times$ order of presentation) was significant when three people could be rescued, $F(1,139) = 8.4891$, $MSE = 12.60$, $p < .01$, as well as in the modified version in which 1000 persons could be saved, $F(1,171) = 20.331$, $MSE = 21.06$, $p < .001$. Importantly, the three-way interaction (number of victims $\times$ ambiguity $\times$ order of presentation) was not significant, $F(1,310) = .16$, $p = .69$.

Separate analyses of the new conditions with 1000 victims and of the control conditions reveal similar patterns as in previous experiments. The ratings for Push with three victims did not differ significantly depending on whether the story was presented first ($M = 2.41$; $SD = 1.54$) or second ($M = 2.70$, $SD = 1.77$), $t(139) = 1.04$. By contrast, the ratings for Switch with three victims were again significantly decreased when Switch was presented second ($M = 3.39$, $SD = 1.55$), as compared to ratings for Switch when presented first ($M = 4.52$, $SD = 1.31$), $t(139) = 4.70$, $p < .001$.

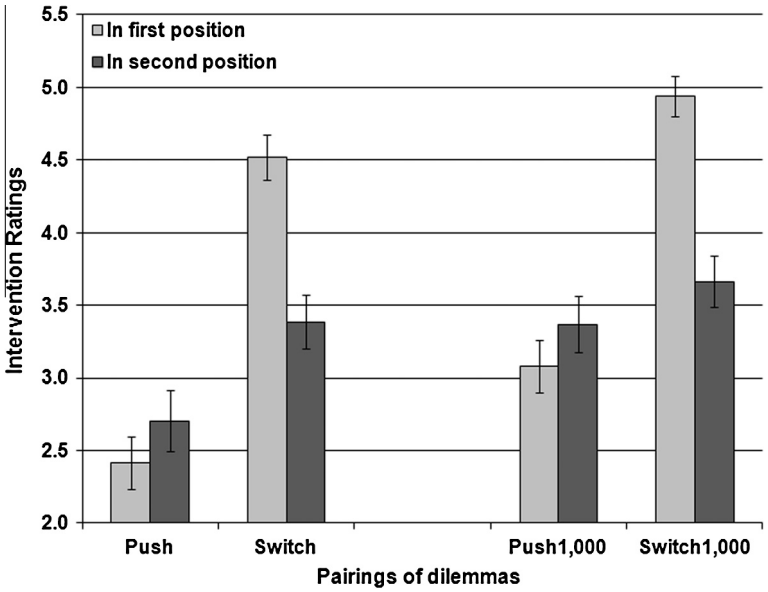

Fig. 9. Ratings for Switch, Push, Push1000 and Switch1000 as a function of the order of presentation and of dilemma pairing. The scale ranged from 1 (“one should certainly not do the proposed action”) to 6 (“one should certainly do the proposed action”). Error bars represent standard errors of means.

A similar pattern was found for the modified versions of Push and Switch with 1000 victims. While the ratings for Push1000 did not differ significantly depending on whether the scenario was presented first ($M = 3.08$; $SD = 1.69$) or second ($M = 3.37$, $SD = 1.79$), $t(171) = 1.10$, $p = .27$, the ratings for Switch1000 were significantly decreased when Switch1000 was presented second ($M = 3.66$, $SD = 1.65$), as compared to ratings for Switch1000 when presented first ($M = 4.94$, $SD = 1.26$), $t(171) = 5.67$, $p < .001$. These results are consistent with our causal model account, but are inconsistent with theories focusing on confidence.

8. Experiment 6

One of the key claims motivating the asymmetric transfer predictions is that in the ambiguous scenarios the action can be easily mentally separated in two different components, saving and killing, which can be differentially highlighted. In Experiment 6 we aimed to provide more direct support for the claim that it is easier to shift the focus between killing and saving in an ambiguous than an unambiguous dilemma. More specifically, we predicted that Switch should be perceived as a case of saving per default. However, after having been presented after a negative unambiguous dilemma, such as Push, the focus in Switch should be shifted towards the killing aspect. We expected no such changes for the unambiguous dilemmas Push and PosPush in the second position.

8.1. Method

8.1.1. Participants

280 subjects, each receiving £0.50, were recruited via an online database located in the U.K.

8.1.2. Design, materials, and procedure

Participants were randomly assigned to one of four conditions, namely Push_Switch, Switch_Push, Push_PosPush, or PosPush_Push.[3] The labeling of the conditions expresses in which order the respective dilemmas were presented. After each dilemma participants were requested to rate the proposed action on a 6-point Likert scale ranging from 1 (“performing the action is a clear case of killing”) to 6 (“performing the proposed action is a clear case of saving”). Subjects were told that they should use the points in between to express their intuitions about the relative weight of these two extremes.

8.2. Results and discussion

47 participants were excluded from the analyses for the same reasons as in the previous studies. The results for the remaining participants are summarized in Fig. 10. When paired with Switch, it did not make a significant difference for Push whether this dilemma was presented before Switch ($M = 2.56$, $SD = 1.63$) or afterwards ($M = 2.52$, $SD = 1.62$), $t(116) = .13$, $p = .89$. In both cases, killing was the dominant description for Push. In contrast, judgments for Switch differed significantly depending on whether this dilemma was presented before ($M = 3.95$, $SD = 1.41$) or after Push ($M = 3.13$, $SD = 1.55$), $t(116) = 3.06$, $p < .01$. The two-way interaction (ambiguity $\times$ order of presentation) was also significant, $F(1,116) = 9.69$, $MSE = 10.98$, $p < 0.01$. When presented first, Switch was more strongly judged as a case of saving than when it was preceded by Push. This shift confirms the prediction of our theory that ambiguous

[3] In this experiment we did not pair Switch with PosPush because their default evaluations are too similar.

dilemmas are more amenable to shifts of focus than unambiguous dilemmas.

As a control, we also tested conditions in which the unambiguous scenarios Push and PosPush were placed in the second position. Here our theory predicts no shifts of the interpretation of the action. When paired with PosPush, it did not make a significant difference whether Push was presented before PosPush ($M = 2.60$, $SD = 1.69$) or afterwards ($M = 2.79$, $SD = 1.56$), $t(113) = .63$, $p = .53$. In both positions, the interpretation as a case of killing prevailed. Ratings for PosPush also did not differ depending on whether it was presented before ($M = 4.56$, $SD = 1.53$) or after Push ($M = 4.64$, $SD = 1.38$), $t(113) = 0.29$, $p < .77$. Here the interpretation of saving was predominant.

9. General discussion

We have proposed a new theory of transfer effects between moral dilemmas. In our theory we focus on two components of moral dilemmas, namely their causal structure and their default evaluations. On the basis of the causal structure we distinguish between ambiguous and unambiguous dilemmas. In unambiguous dilemmas the good and bad outcomes are ordered on a single causal path. Thus, it is only possible to highlight one single (either bad or good) aspect in this type of dilemma, which often makes this dilemma immune to transfer effects. Intuitions about unambiguous dilemmas can only be changed when the default evaluations differ, and when the same causal path is highlighted in both dilemmas. In contrast, ambiguous dilemmas present good and bad outcomes on different separable causal paths, and therefore allow for selective highlighting of either the positive or negative aspect of the dilemma. The evaluation of such a dilemma can be shifted by a preceding dilemma if the default evaluations between the consecutively presented dilemmas differ. Our theory accounts for asymmetrical transfer effects that have been found in previous studies (e.g., Lanteri et al., 2008; Liao et al., 2012; Lombrozo, 2009; Petrinovich & O'Neill, 1996; Wiegmann et al., 2012).

The theory also makes a number of novel predictions, which we have confirmed in six experiments. We showed that ratings of Switch cannot only be lowered by a preceding negative (Experiment 1), but also be raised by a preceding positive unambiguous dilemma (Experiment 2). We also showed that evaluations for an unambiguous dilemma were unaffected by a previous unambiguous dilemma despite strongly differing default evaluations (Experiment 3). We explained this lack of a transfer effect as a consequence of the highlighting of different causal paths in the two dilemmas.

To rule out the possibility that asymmetric transfer effects only occur between trolley dilemmas, unambiguous negative and ambiguous positive dilemmas from different domains were paired. Our general prediction of asymmetric transfer was confirmed with these other domains. Importantly, this experiment also demonstrated that transfer effects are also obtained across domains with different surface features, which supports our view that transfer effects were driven by the structure of the underlying causal models (Experiment 4). In Experiment 5 we have successfully defended our theory against the theory that asymmetries of transfer are due to differential position on the rating scale expressing different degrees of confidence. Finally, we have shown that the interpretation of the proposed action in an ambiguous dilemma can be more easily shifted between the dimensions killing and saving than the interpretation of the action in unambiguous dilemmas (Experiment 6).

9.1. Alternative explanations

Our theory provides a comprehensive account of the moral transfer effects we have found in our studies. It also

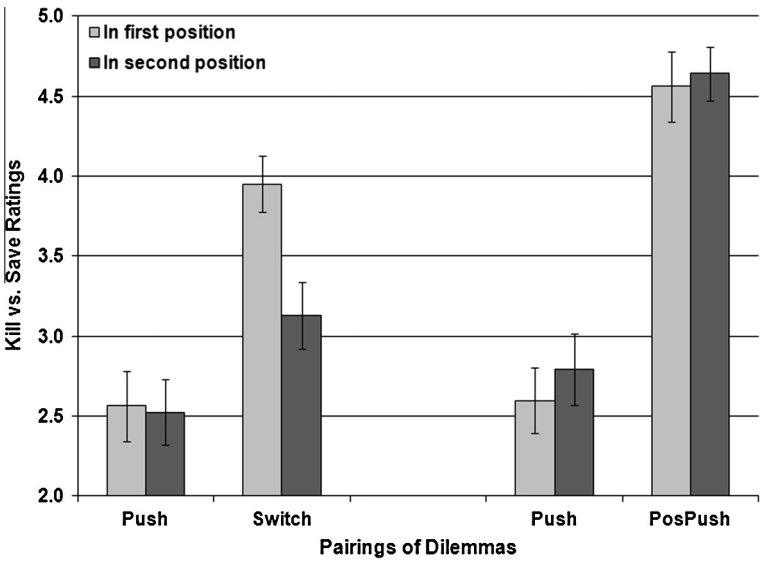

Fig. 10. Ratings for Switch, Push, and PosPush as a function of the order of presentation and of the paired dilemma. The scale ranged from 1 ("clear case of killing") to 6 ("clear case of saving lives"). Error bars represent standard errors of means.

explains findings in previous studies. However, alternative theories have been proposed, which we would like to briefly discuss here.

Greene, Sommerville, Nystrom, Darley, and Cohen (2001) have proposed a dual-process theory of moral judgment that claims that two distinct brain systems are involved in moral judgment, one mainly devoted to emotional and the other one to rational processes. According to this theory, dilemmas like Push should activate negative emotions, whereas dilemmas like Switch should activate rational processes. Thus, a straightforward account of asymmetrical transfer may postulate that emotions but not rational reasoning is transferred to the subsequent dilemma.

However, although such an explanation would be able to predict the effects found for the pairings of Push and Switch, they do not predict the lack of a transfer effects between unambiguous dilemmas, such as Push and PosPush, which is predicted by our theory as being caused by their underlying causal structures. If transfer effects were triggered by emotions, transfer effects should have been observed between Push and PosPush.

One possible post hoc explanation of the lack of transfer between Push and PosPush could add the assumption that negative emotions can only be transferred to neutral scenarios, not to those that elicit positive emotions (as possibly PosPush). However, there is a more general problem with emotion-based explanations. Recent studies that analyzed the signatures of emotional processing for Switch and Push concluded that people's different reactions to Switch and Push could not be attributed to differences in emotional involvement. Nakamura (2013) analyzed the 62 moral dilemma tasks used in Greene et al. (2001) by means of a factor analysis and structural equation modeling, and found no differences between Switch and Push with regards to emotional involvement. Horne and Powell (2013) arrived at the same conclusion using self-report measures of emotions.

Lanteri et al. (2008) offer a further possible mechanism of transfer effects. According to this explanation, Push might highlight the right of a single victim against harmful interventions, thereby triggering a negative evaluation of the proposed action in Push. Sensitivity to this right may then be transferred to Switch. However, this explanation does not explain why the intuitions triggered by Switch or PosPush in which the rights of the three potential victims to be saved seem to dominate the right of the single victim do not lead to a transfer effect in Push.

Schwitzgebel and Cushman (2012; see also Lanteri et al., 2008) proposed a different explanation for the asymmetrical transfer effect that integrates emotion-based and principle-based accounts. According to these authors, asymmetrical transfer effects might arise because of an interaction of the strength of the intuitions elicited by the two dilemmas and the general desire of people to maintain consistency between judgments. Maintaining consistency in this context means that the same moral status should be assigned to actions that cause identical outcomes. According to Schwitzgebel and Cushman (2012), the intuitions triggered by Push are stronger than the ones triggered by Switch because the negative evaluation of Push is backed up by a (negative) emotional reaction. Hence, if people are first presented with Push, they acquire a strong intuition that it is morally wrong to perform an action that results in the killing of one person. This strong intuition then overrides the default evaluation in Switch, leading to similar evaluation for the actions proposed in Push and Switch. In contrast, the intuition that it is permissible to perform the proposed action in Switch only results in a relatively weak intuition that it is morally right to perform an act that leads to the saving of three people and the killing of one person. This relatively weak intuition in combination with the desire to maintain consistency is not sufficient to override the strong emotionally backed up intuition that it is wrong to intervene in Push.

Experiment 5 can be viewed as a refutation of the alternative theory that it is a difference in confidence regarding the judgments for Push and Switch that cause the asymmetrical transfer effect between these two dilemmas. In this experiment we have shown that the underlying causal structure rather than patterns of confidence predict transfer. Moreover, the recent failures to find support for the role of emotions in intuitions about trolley dilemmas also create problems for this theory (Horne & Powell, 2013; Nakamura, 2013).

A different direction to look for explanations for asymmetric transfer is the extensive literature on analogical transfer (see Holyoak, 2012, for an overview). Ortony (1979) claimed, for example, that asymmetric transfer between analogical cases might be driven by "salience imbalance".[4] According to this theory, transfer is more likely from the more salient to the less salient scenario than in the opposite direction. Features that are of high salience in the base domain increase the salience of features in the target domain. This theory can account for asymmetric transfer between Push and Switch, for example, if Push is considered to be more salient than Switch due to its highlighting of killing.

In our view, salience accounts suffer from the unconstrained notion of the term salience. Once we know the outcome of the transfer experiments, it is always possible to select features that justify the desired salience imbalance. One can certainly affect the salience of scenarios by modifying irrelevant features, such as the color of the tracks, which in all likelihood would not affect transfer of moral intuitions. We actually believe that our causal model theory provides a principled account that could explain intuitions about salience imbalances. The theory explains, for example, why killing is highlighted in Push, but both outcomes are highlighted in Switch. Thus, it provides the necessary constraints for making predictions for various pairings.

9.2. Limitations and directions for future research

Our experiments only represent a first step in the direction of testing theories of transfer between moral intuitions. The large literature on analogical transfer points to many possible routes for future research. Interestingly, there has been an increased interest in combining causal

[4] This explanation was suggested by an anonymous reviewer.

model accounts with theories of analogical transfer (see Holyoak, Lee, & Lu, 2010).

In Experiment 4, we have already demonstrated that transfer may not only be observed between highly similar scenarios but also between scenarios coming from different domains. We have shown that the underlying causal model provides the bridge between different scenarios. However, the literature on analogy has taught us that both structural and superficial features contribute to transfer (see Holyoak, 2012). Thus, there are many more variations that need to be tested. For example, the dilemmas used in our experiments are restricted to cases in which decisions about life and death need to be made. It would be interesting to look at dilemmas in which the stakes for the victims are lower (e.g., wounds, monetary losses). Our theory would make similar predictions there, although some of the effects may decrease due to the greater similarity of default evaluations.

A second restriction in the design of our studies is that we did not vary the number and kind of victims in the consecutively presented dilemmas. Although we would expect that the effects of the causal structures are unaffected by such manipulations, the numbers and kinds of victims would certainly affect the default evaluations, and hence the possible sizes of effects. In one previous study, Wiegmann and Okan (2012) found that presenting Push prior to a version of Switch in which one hundred persons could be saved led to a decrease of the approval for the proposed action in Switch despite the larger number of saved victims. This finding confirms the importance of the underlying causal structure. Another interesting direction to explore is to study alternative, more complex causal structures. We have concentrated on the standard cases but dilemmas can obviously present themselves in various more complex causal setups (see, for example, Unger, 1996).

Further variations that we would like to explore in future research are situations in which the temporal lag of the presentation of the consecutive dilemmas is manipulated. One could also increase the number of presented dilemmas, or manipulate when or whether each dilemma is being evaluated (see, e.g., Lombrozo, 2009). Research on analogical transfer has found that temporal lags, awareness of potential relevance of some scenario for a target problem, and the induction of abstract schemas due to the presentation of multiple analogical cases may moderate the degree of transfer effects (see Holyoak, 2012, for an overview).

One important restriction of the present research is that we have focused on dilemmas in which victims were physically harmed by some instrument. We chose these scenarios because they make it relatively easy to manipulate causal models. To date, little is known about the potential role of causal intuitions in other types of moral norm violations (e.g., violations of hierarchical authority regulations, of purity rules, or cases of lying and cheating). Although we believe that causal considerations will play a role in these domains as well, there might be other knowledge structures (e.g. belief-desire reasoning, norm-based reasoning) that will additionally affect both moral intuitions and transfer. Our theory only represents a first step in the direction of developing a comprehensive theory of transfer between moral intuitions.

Acknowledgements

This research was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG WA 621/21-1), and the Courant Research Centre 'Evolution of Social Behaviour', University of Göttingen (funded by the German Initiative of Excellence). We thank Gunda Johannes for proofreading, and Jonas Nagel for valuable comments on an earlier draft.

Appendix A

Descriptive statistics for all dilemmas used in Experiment 4.

Domain	Causal structure	Position	Followed or preceded by	Mean	*SD*	*N*
Fire	Ambiguous	First	Fire	3.92	1.52	26
Fire	Ambiguous	First	Fumes	4.32	1.25	28
Fire	Ambiguous	First	Shark	3.93	1.41	27
Fire	Ambiguous	Second	Fire	3.67	1.39	27
Fire	Ambiguous	Second	Fumes	3.66	1.45	32
Fire	Ambiguous	Second	Shark	3.43	1.53	28
Fire	Unambiguous	First	Fire	2.96	1.45	27
Fire	Unambiguous	First	Fumes	2.63	1.67	30
Fire	Unambiguous	First	Shark	3.36	1.69	36
Fire	Unambiguous	Second	Fire	2.88	1.58	26
Fire	Unambiguous	Second	Fumes	2.50	1.68	30
Fire	Unambiguous	Second	Shark	3.72	1.57	25

(*continued on next page*)

(*continued*)

Domain	Causal structure	Position	Followed or preceded by	Mean	*SD*	*N*
Fumes	Ambiguous	First	Fire	3.37	1.73	30
Fumes	Ambiguous	First	Fumes	3.66	1.47	35
Fumes	Ambiguous	First	Shark	3.81	0.95	31
Fumes	Ambiguous	Second	Fire	3.10	1.54	30
Fumes	Ambiguous	Second	Fumes	2.91	1.59	32
Fumes	Ambiguous	Second	Shark	3.11	1.47	35
Fumes	Unambiguous	First	Fire	2.53	1.68	32
Fumes	Unambiguous	First	Fumes	2.75	1.74	32
Fumes	Unambiguous	First	Shark	2.92	1.35	26
Fumes	Unambiguous	Second	Fire	3.07	1.56	28
Fumes	Unambiguous	Second	Fumes	2.66	1.41	35
Fumes	Unambiguous	Second	Shark	3.14	1.63	28
Shark	Ambiguous	First	Fire	3.92	1.32	25
Shark	Ambiguous	First	Fumes	4.32	1.28	28
Shark	Ambiguous	First	Shark	3.68	1.19	28
Shark	Ambiguous	Second	Fire	3.86	1.20	36
Shark	Ambiguous	Second	Fumes	3.12	1.45	26
Shark	Ambiguous	Second	Shark	3.04	1.60	28
Shark	Unambiguous	First	Fire	1.93	1.74	28
Shark	Unambiguous	First	Fumes	2.14	1.33	35
Shark	Unambiguous	First	Shark	2.14	1.48	28
Shark	Unambiguous	Second	Fire	2.41	1.72	27
Shark	Unambiguous	Second	Fumes	2.42	1.63	31
Shark	Unambiguous	Second	Shark	1.61	1.07	28

References

Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition, 108*, 381–417.

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*, 1082–1089.

Di Nucci, E. (2013). Self-sacrifice and the trolley problem. *Philosophical Psychology*. Advanced online publication.

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

Gräfenhain, M., & Wiegmann, A. (2012). The scientific study of morals. In C. Lütge (Ed.), *Handbook of the philosophical foundations of business ethics*. Springer Press.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI study 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.

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

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 234–259). New York: Oxford University Press.

Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General, 139*, 702–727.

Horne, Z., & Powell, D. (2013). Emotional arousal and moral judgments. In M. Knauff, M. Pauen, N. Sabanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.

Horne, Z., Powell, D., & Spino, J. (2013). Belief updating in moral dilemmas. *Review of Philosophy and Psychology, 4*, 705–714.

Kamm, F. M. (2007). *Intricate ethics*. Oxford: Oxford University Press.

Knobe, J. (2010). Action trees and moral judgment. *Topics in Cognitive Science, 2*, 555–578.

Lanteri, A., Chelini, C., & Rizzello, S. (2008). An experimental investigation of emotions and reasoning in the trolley problem. *Journal of Business Ethics, 83*, 789–804.

Liao, S. M., Wiegmann, A., Alexander, J., & Vong, G. (2012). Putting the trolley in order: Experimental philosophy and the loop case. *Philosophical Psychology, 25*, 661–671.

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science, 33*, 273–286.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York: Cambridge University Press.

Nakamura, K. (2013). A closer look at moral dilemmas: Latent dimensions of morality and the difference between trolley and footbridge dilemmas. *Thinking and Reasoning, 19*, 178–204.

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

Ortony, A. (1979). The role of similarity in similes and metaphors. In A. Ortony (Ed.), *Metaphor and thought* (pp. 186–201). Cambridge: Cambridge University Press.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology, 17*, 145–171.

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

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind and Language, 27*, 135–153.

Sloman, S. A., Fernbach, P. M., & Ewing, S. (2009). Causal models: The representational infrastructure for moral judgment. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral judgment and decision making. The psychology of learning and motivation: Advances in research and theory* (pp. 1–26). San Diego, CA: Elsevier.

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

Timmons, M. (2002). *Moral theory: An introduction*. Lanham: Rowman & Littlefield.

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

Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The oxford handbook of thinking and reasoning* (pp. 364–389). New York: Oxford University Press.

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

Waldmann, M. R., & Wiegmann, A. (2010). A double causal contrast theory of moral intuitions in trolley dilemmas. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 2589–2594). Austin, TX: Cognitive Science Society.

Wiegmann, A., & Okan, J. (2012). Order effects in moral judgments. Searching for an explanation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 1143–1148). Austin, TX: Cognitive Science Society.

Wiegmann, A., Okan, Y., & Nagel, J. (2012). Order effects in moral judgment. *Philosophical Psychology, 25*, 813–836.