If You're Going to Do Wrong, at Least Do it Right:

Considering Two Moral Dilemmas at the Same Time Promotes Moral Consistency

We study how people reconcile conflicting moral intuitions by juxtaposing two versions of classic moral problems: the trolley problem and the footbridge problem. When viewed separately, most people favor action in the former and disapprove of action in the latter, despite identical consequences. The difference is often explained in terms of the intention principle – whether the consequences are intended or incidental. Our results suggest that when the two problems are considered together, a different judgment emerges: participants reject the intention principle and embrace either the principle of utilitarianism, which favors action in both problems, or the action principle, which rejects action in both problems. In subsequent studies, we find that when required to choose between two harmful actions, people prefer the action that saves more lives, despite its being more aversive. Our findings shed light on the formation of moral judgment under normative conflict, the conditions for preference reversal, and the potential polarization of moral judgment under joint evaluation. Organizational implications are discussed.

Keywords: *joint evaluation; preference reversal; normative conflict; moral judgment; deontology; utilitarianism; action principle; intention principle*

1. Introduction

The past 15 years of research in moral psychology mind is beset with moral conflict (Graham et al. 2009, Cushman and Greene 2012). Two types of moral judgment are particularly in competition: one that assesses morality based on the actions that a person performs, and another that focuses on the outcomes of those actions (Cushman 2013, Greene et al. 2001, Koenigs et al. 2007). This tension has fueled centuries of philosophical debate between deontological moral theories, concerned with prohibiting and constraining actions, and consequentialist/utilitarian moral theories, concerned with maximizing outcomes and increasing utility. The tension between actions and outcomes also animates classic moral dilemmas: is it permissible to harm some people (a wrongful action) in order to help many others (a superior outcome)?

People who face these dilemmas often provide apparently conflicting responses. A famous example is given by the "trolley problem," which has several versions. In brief, the trolley problem presents a runaway train that is headed towards a group of innocent individuals and is about to kill them. In the *Footbridge* version, people are asked whether they would physically push a person off a footbridge and in front of the train so that the train would hit the fallen person and slow down, allowing the group of individuals to escape. In the *Drop* version, this same outcome can be accomplished by flipping a switch that would drop the person off the footbridge. Most people say that they would not push or drop the person to save five others. Nonetheless, in another version, which we refer to as the *Divert* version, they typically say that they would flip a switch that diverts the train away from the five people and toward one bystander who would be killed as a result. These competing versions of the trolley problem (formulated by Thomson 1985; Foot 1967) are among the most widely known, exhaustively studied, and clearly understood embodiments of an apparent moral conflict.

Although the determinants of action-based and outcome-based moral judgments have been exhaustively studied, surprisingly little is known about how people reconcile moral conflict. When people need to consider cases that invoke competing moral judgments, how do they resolve the conflict?

Prior research indicates that people are motivated to resolve dissonance between cognitive states (Shultz et al. 1999, Festinger 1957, Brehm 1956) and achieve consistency among their moral judgments, such that they often seek to provide the same answer in similar cases (Haidt 2001, Schwitzgebel and Cushman 2012, Lombrozo 2009; Uhlmann, Pizarro, Tannenbaum, & Ditto 2009). Our goal is to understand whether and how people actually achieve moral consistency in a normative conflict. Like many recent studies, we use two competing versions of the trolley problem in order to present a conflict between action-based and outcome-based moral preferences. If people are made to face their inconsistent inclinations, how will they reconcile the conflict?

There are some disadvantages to studying moral behavior through the context of the trolley problem (Bauman et al. 2014); for instance, it is hypothetical, requires suspension of disbelief, and involves atypical circumstances and choices. On the other hand, the problem has notable advantages to studying moral judgment and, in particular, the relationship between intuition and principled reasoning. First, the trolley problem captures the conflict between competing moral values vividly, robustly, and reliably across individuals, and more so than stimuli we have considered from other literatures. Second, the trolley problem has been extensively studied, and our experimental approach and hypotheses draw upon the specific lessons of this existing literature. This includes a precise and detailed accounting of both the principled and intuitive basis of judgments in trolley problems, which is not available for other experimental paradigms targeting moral reasoning. Finally, like other stylized paradigms such as the prisoner's dilemma or the ultimatum game, trolley problems embody a tradeoff between rights and welfare that models fundamental social problems. For instance, the introduction of driverless cars offers a vivid contemporary example of trolley-like challenges presented to engineers and policymakers (Bonnefon, Shariff, & Rahwan, 2016).

2. The Competing Moral Principles

The research on moral judgment points to several principles that could be used to resolve the conflict between the trolley problems. The first principle is *utilitarianism*, the moral principle that evaluates actions based on their net benefit and seeks to maximize the good (Mill 1863, Bentham 1789).

According to *act utilitarianism*, the resolution of the Drop and the Divert dilemmas should be identical; in both cases, one should sacrifice the one to save the many, as this is the better outcome. (Other, more complex versions of utilitarianism exist that we do not address in the present paper.)

The remaining two principles reflect action-based, deontological evaluation. The *action principle* categorically prohibits any harmful action regardless of its consequences, positing that actively causing harm is worse than passively allowing it (Cushman et al. 2006, and in a different context, Baron and Spranca 1997, Ritov and Baron 1999). Applying the action principle yields the conclusion that one should act in neither case, as both diverting and dropping cause harm to innocent individuals, whereas inaction merely allows harm to occur.

Whereas the action principle says "do no harm," the third principle—the *intention principle*—can be summarized as "do no *intended* harm." This principle, which philosophers refer to as the *doctrine of double effect* (Foot 1967, Thomson 1985), prohibits only actions that are intended to bring about harm, and specifically actions that intend to use a person as a means to an end. In contrast, actions that involve unintended but foreseen incidental harms are permitted if they yield better outcomes overall. Applying the intention principle leads to the conclusion that one should reject action in the Drop version, in which harm to the one is intended as a means to save the others, yet take action in the Divert version, in which harm to the one is foreseen but not intended. Notably, intention is the only principle discussed in this paper that distinguishes the Drop and Divert versions, whereas the action principle and utilitarianism each yields consistent (and opposite) conclusions across the Drop and Divert versions.

3. Reconciling Competing Intuitions under Joint Evaluation

We adopt a simple method to examine how people reconcile competing and inconsistent moral intuitions. Specifically, we compare how people judge the rival drop and divert versions of the trolley problem when they engage with each case separately versus when they engage with them simultaneously. Asking people to choose which version is more permissible creates a conflict between a case typically evaluated on the basis of its consequences (trolley/divert) and a case typically evaluated on the basis of its action (footbridge/drop).

This basic method, contrasting *joint evaluation* (JE) versus *separate evaluation* (SE), is widely employed in research on decision making. The thrust of the SE/JE paradigm can be illustrated through the following example: suppose you were offered a job that would pay $75,000 a year to you and $75,000 to candidates who have credentials similar to yours (the fair option). Now suppose you were offered a job paying $85,000 a year to you and $95,000 to similar candidates (the beneficial option). Which offer, if any, would you choose? (Bazerman et al. 1994). People tend to assess the fair option more positively than the beneficial option when they evaluate the options one at a time, yet shift their preferences to the more beneficial option when the options are presented simultaneously. Many studies suggest that the reversal of preferences from fairness to outcome maximization persists across a wide range of contexts (Bazerman et al. 1992, Bazerman et al. 1994, Hsee et al. 1999, Tenbrunsel et al. 2000). More generally, JE was found to reduce the effect of emotions on decision making and to promote reflective and controlled decision processes (Irwin et al. 1993, Bazerman et al. 1998, Kahenman and Ritov 2004, Ritov and Baron 2011; Hsee 1996; Hsee et al. 1999, Bazerman et al. 1994).

The main conclusion of this body of research is that under conditions of JE, people are motivated to define and apply an accessible and convincing principle of decision making—one that they are explicitly aware of and would endorse upon reflection. Often, this principle appears to be outcome maximization. Following this tradition of research, we propose that when people are made to face their conflicting moral intuitions under JE, they would be motivated to find and apply a moral principle that they can readily articulate and endorse upon reflection. Specifically, we expect that under JE, more people would apply utilitarianism, but as we discuss next, JE may also prompt people to become more action averse as well. We parse out these two possibilities in our paper.

4. Hypotheses

Which moral principle(s) is JE expected to promote? One likely option, supported by the JE literature and the dual process model of moral judgment (Greene 2004, 2008, Greene et al. 2001), is for JE to promote utilitarianism. The dual process theory—which was extensively applied to the trolley problem—posits that utilitarian moral judgments are the product of cognitive control, whereas

deontological (i.e., action-based) moral judgments are the product of automatic and often emotional psychological processes (for a review, see Cushman and Greene 2013; for a critical review, see Baron et al. 2012). Given JE's record of promoting controlled decision processes, and ones that involve outcome comparisons in particular, a natural prediction is that people will become more utilitarian under JE. This hypothesis is supported by previous research showing that JE shifted focus to the magnitude of harm (an outcome-based factor) rather than the directness of the action (an action-based factor) (Paharia, Kassam, Bazerman and Greene 2009).

However, there is an alternative prediction: JE may also promote the action principle. First, although the dual process model emphasizes the contribution of cognitive control to utilitarian judgment, there is substantial agreement that some deontological concerns are also formalized and used as explicit rules (Bartels 2008, Cushman et al. 2010, Paxton and Greene 2010, Paxton et al. 2011, Pizarro and Bloom 2003). More specifically, previous studies suggest that people engaged in the controlled, deliberative task of justifying their moral judgments readily articulate the action principle as a basis for judgment (Cushman et al. 2006). If so, under conditions of JE, people may be more likely to endorse the action principle and, as a result, less likely to engage in (any) action.

The third possibility is that under conditions of JE, people may be prompted to reason from the intention principle. If so, under JE, we should expect participants' preference for action in the divert case to increase as compared with the drop case. However, recent research offers some reason to doubt this possibility. Several studies that presented the divert and the push (not drop; more on the importance of this difference later) dilemmas sequentially found that the judged permissibility of the divert choice decreased when it was viewed after the push option (Lombrozo 2009; Schwitzgebel and Cushman 2012; Kelman and Kreps 2014). Were the comparison to have highlighted the intention principle, the result would have been different. More direct evidence against the intention principle was obtained in a study that asked people to provide explicit justifications for their diverging judgments of the Footbridge versus Trolley problems. A large majority of the participants were unable to do so, and only a small minority succeeded in unambiguously appealing to the intention principle (Cushman et al. 2006). Given the

difficulty in reflectively arriving at the intention principle, it seems unlikely that this principle would play a dominant role in moral judgments under JE. In the absence of the intention principle, however, preferring action in divert over drop (as people often do under SE) appears inconsistent. Hence, we expect that the preference for the divert case would decrease under JE, as other principles—that provide an articulable rationale for choosing one of the other options—would move to center stage.

Overall, we expect that JE will prompt people to reason either from utilitarianism or from the action principle, two moral principles that appear to be readily articulable and explicitly used in conditions allowing reflection and deliberation. Conversely, we expect that JE will discourage people from endorsing the "divert" option, which neither maximizes welfare (in violation of utilitarianism) nor avoids direct harm-doing (in violation of the action principle), and is consistent only with the intention principle or the doctrine of double effect—a moral principle that people rarely articulate or endorse. Notably, these hypotheses expect conditions of deliberation to shift moral judgments to conform with articulable moral principles. Whereas some models of moral judgment argue that moral reasoning primarily shapes post-hoc rationalizations of prior intuitions (e.g. Haidt, 2001), we explore whether moral reflection can alter moral judgment in its formation.

II. Overview of Experiments

We adapted the trolley problem to the standard template of JE versus SE by directly juxtaposing the drop and divert cases, while varying the utility of action in each case. In this design, the drop case involves harming one person to save five others (1 vs. 5), whereas the divert case involves harming one person to save three others (1 vs. 3). In the joint condition, we present the two cases as occurring simultaneously with time to take action only in one of the cases (Figure 1; the full text is included below). Under SE, people are expected to prefer divert to drop because diverting the boxcar appears more permissible than dropping the person; in the absence of drop, diverting is supported by both outcome-based and action-based moral evaluations (the intention principle). Under JE, however, an outcome-based evaluation (utilitarianism) would rank action in Drop as superior to action in Divert because it saves more lives, whereas an action-based evaluation (the intention principle) would rank action in Divert as superior

to action in Drop because it does not *intentionally* harm anyone. Alternatively, an evaluation based on the (in)action principle would rank inaction as superior to action in both Drop and Divert (because it does not *actively* harm anyone). Table 1 shows the evaluation of each choice under each principle.

Table 1 about here

Several recent studies asked participants to rate the trolley and the footbridge cases sequentially (Lombrozo 2009, Schwitzgebel and Cushman 2012, Kelman and Kreps 2014) or side by side (Shallow et al. 2011) in order to study a variety of questions. Lombrozo (2009) used these ratings to examine whether moral commitments influence the rating of the trolley dilemmas; Schwitzgebel and Cushman (2012) asked whether trained philosophers show the same order effects that lay people do in moral judgment; and Kelman and Kreps (2014) studied the effect of kinship and identifiability on people's moral intuitions. The present study is the first that contrasts the two cases to study how normative conflict promotes certain moral principles. To do so, we introduce several new and important features into the JE paradigm.

First, the JE condition directly juxtaposes the two cases as one dilemma rather than presenting them side by side as two separate dilemmas. In addition, instead of asking participants to rate each case separately, participants are asked to choose between them and therefore are forced to reconcile the moral conflict (by comparison, see Lombrozo 2009, Schwitzgebel and Cushman 2012, Kelman and Kreps 2014; Shallow et al. 2011). This design thus creates a salient and clear moral conflict.

In addition, we modified the original trolley and footbridge problems to create a one-to-one association between options and moral principles. First, reducing the number of lives saved in the divert case from five to three lives ensured that under JE, only the "drop" option will be supported by act utilitarianism, and that the "divert" option will only be supported by the intention principle. (When the Divert case is evaluated separately, act utilitarianism and the intention principle are aligned in endorsing action.) The second innovation of the design is the use of the drop version instead of the original footbridge version. This change was needed to untangle the bundle of intention and physical contact that

characterizes action in the footbridge problem. To illustrate the problem, note that the contrast between the original trolley and footbridge dilemmas (as studied in Lombrozo 2009, Shallow et al. 2011, Schwitzgebel and Cushman 2012, Kelman and Kreps 2014; Shallow et al. 2011) involves at least one other dimension beyond the intentionality of the act: in the footbridge problem, the agent has to push the one to save the five, whereas in the trolley problem, the agent merely has to flip a switch to accomplish this goal. This physical interaction was found to play a reliable role in shaping moral judgments (Cushman et al. 2006). To study how the intention principle compares to utilitarianism in the absence of physical contact, the present study replaces the act of pushing with an act of flipping a switch that mechanically drops the man. The resulting drop case differs from the divert case on the dimension of the intention principle, but not on the dimension of physical contact, as a switch is used in both cases. This design allowed us to isolate the influence of the three contrasting moral principles—utilitarianism, the intention principle, and the action principle—by associating each choice with only one principle. In our set of choices, drop is consistent with utilitarianism, divert with the intention principle, and inaction with the action principle.

Finally, the study was designed to test any changes in preferences under SE versus JE in the absence of potential similarity effects (Shallow et al. 2011). Shallow and his colleagues asked participants to rate the push, divert, and inaction alternatives, with the consequences of two of the three options being presented as having more similar outcomes to each other in terms of the number of fatalities. Thus, in one of their conditions, divert resulted in four deaths, an outcome numerically similar to the five deaths resulted by inaction. In another condition, divert resulted in two deaths, an outcome numerically similar to the one death resulted by push. In both conditions, similarity lowered the rating of the option that was designed to be numerically similar to divert (Shallow et al. 2011). The present study examines the effect of JE in the absence of similarity effects by setting the numerical consequences of the divert option to be exactly intermediate between those of the drop and inaction options, so that the numerical outcome is not uniquely similar to either case.

Insert Figure 1 about here
(1.5 column fitting)

Figure 1: Illustrations of the Moral Problem

Each experimental condition presented either the Drop (upper) or the Divert (lower) illustration immediately after the text description of the case. The JE condition presented the combined illustration depicted in this figure.

1. Experiment 1: Joint versus separate evaluation

1.1. Participants

Three hundred people recruited through Amazon Mechanical Turk (mTurk) participated in the study online in exchange for $0.15.

1.2. Procedure

Participants were randomly assigned to one of three groups: *SE-Divert*, *SE-Drop*, or *JE*. In each of the SE conditions, participants read the scenario below and were asked: "What would you do?" They then marked their choice of either flipping the switch or not flipping the switch. In the divert version, flipping the switch would divert the train; in the drop version, flipping the switch would drop the man. In the JE condition, participants chose among three options: flip switch A, flip switch B, or do not flip either switch.

In the SE-Divert version, the scenario read:

> You are working by the train tracks when you see an empty boxcar break loose and speed down the tracks. The boxcar is heading toward three workmen who do not have enough time to get off the main track. If you do nothing, these three workmen would be killed.
>
> [Just before the three workers there is a side track branching off of the main track. On this side track there is one other worker. You can run over and flip a

switch that will send the boxcar down the side track. The man on the side track will be killed, but the boxcar will not hit the three workmen on the main track.]

In the SE-Drop condition, the number of workmen at risk was changed to *five* instead of three, and the text in brackets was replaced with the following:

> Above the main track is a platform with another worker, who is very large. This worker is not threatened by the boxcar. But, he is standing over a trap door. You can run to a switch that will open the trap door and drop this large man in front of the boxcar. The man will be killed, but his body will get caught in the wheels of the boxcar, slowing it down enough for the five workmen to escape.

Participants in the JE condition were presented with a "trilemma" in which they stood between two sets of railroad tracks (see Figure 1). On the left, the drop version was unfolding, and the participant could run to a switch that would drop one person in front of a train to prevent it from hitting *five* people (the drop option). On the right, the divert version was unfolding, and the participant could run to a different switch that diverts the train to hit one person on a side track in order to save *three* people (the divert option). Participants were told they only had time to activate one switch or do nothing at all (the third option of inaction). All conditions asked the identical question: "What would you do?"

1.3. Results

Table 2 summarizes the main results. The data was subjected to an exact test for two-way contingency tables with structural zeros, due to the structure of our contingency table that included built-in empty cells (West and Hankin 2008, using the aylmer package in R). This test indicated that the choice rates for drop, divert, and inaction differed significantly between the SE and the JE conditions ($p = .0005$; $\chi^2 = 160.38$, $Phi = .731$). Several proportion tests (z-tests) were then performed to investigate the nature of the difference.

In line with past results, we found that under SE, people were more likely to endorse diverting over doing nothing (75.5%) than to endorse dropping over doing nothing (41%) ($Z = 3.73$, $p < .00001$, d

= .82). In the present study, the preference for diverting in SE was higher despite the fact that dropping saved five people and diverting saved only three people.

Comparing the SE and JE conditions, the total number of options available to participants increases from two to three. It thus follows that some options will attract a reduced proportion of participants; the key research question is, which options will bear the brunt of this reduction, and which will attract more support?

In contrast to SE, where divert was preferred to drop, under JE, drop was the modal choice, and the proportion of participants who chose drop (45%) was the highest among the three options and significantly higher than that of those who chose divert (18.1%; $Z = 4.713$, $p < .00001$, $d = .715$). The proportion of those choosing drop under JE did not differ significantly from the proportion of those choosing drop under SE (41%, $p = .367$, n.s.). Yet given that the total number of competing options increases from one (in SE) to two (in JE), it is notable that there is no reduction and even a slight increase in the proportion of participants selecting drop.

In contrast, the proportion of people who endorsed the divert option decreased dramatically with JE, as compared with SE-Divert, from 75.5% to 18.1% ($Z = -4.519$, $p < .00001$, $d = 1.46$). Finally, we observe a marginally significant increase in preference for inaction when we compare SE-Divert to JE (24.5% vs. 37.2%), ($Z = 1.945$, $p = .051$, $d = .372$).

Insert Table 1 about here

1.4. Discussion

Experiment 1 reveals a shift in preferences under separate versus joint evaluation. Under separate evaluation, divert was preferred to drop by a large margin. Under joint evaluation, divert became the least-preferred option, drop became the highest-endorsed option, and support in inaction increased.

Our results have several implications. First, they suggest that the intention principle has little impact under JE. The intention principle provides an explicit rationale for judging the drop case as

impermissible; yet, under JE, participants were no less likely to endorse drop than under SE. In fact, drop was the most popular choice. The intention principle also provides an explicit rationale for judging the divert case as permissible; yet, under JE, participants were significantly less likely to endorse divert than under SE.

The low endorsement of the divert option under JE is also notable because divert could have been perceived as a *compromise* between drop and inaction, given its intermediate numerical outcome and its action-based rationale; on this basis, it might have been expected to draw a greater share of choices in a setting of moral conflict (Simonson and Tversky 1992, Tversky and Simonson 1993). However, there is no evidence of a compromise effect. In sum, there is no evidence for an increased application of the intention principle under JE; in fact, the opposite is true.

Second, the results suggest that the action principle becomes more influential under JE; compared to divert, more people choose inaction under JE. One possible explanation for this effect is that people choose inaction because the action principle provides a readily articulable rationale for their intuition that it is wrong to drop a person in front of a train in order to save several lives. Having arrived at a principle, they apply it across the choice set, which leads them to refrain from any action, dropping or diverting. This interpretation is in line with past studies that examined order effects on trolley problems. This research suggests that the intuition elicited by the push case may become the basis for decision in the divert case (Petrinovich and O'Neill 1996, Schwitzgebel and Cushman 2012). In other words, people presumably use the action principle to impose consistency on their moral judgment of cases.

Finally, the influence of utilitarian reasoning on moral judgment under JE is harder to evaluate. It depends on whether we consider the drop case or the divert case. For the drop case, 41-45% of participants favored the utilitarian action both under separate and joint evaluation. Thus, we find strong evidence for a utilitarian influence on moral judgment under moral conflict, but this influence does not outweigh action-based influences in total. (It outweighs the influence of the intention principle, but not the intention and action principles together.) For the divert case, we find a decrease in utilitarian judgment

under JE, as the rise in inaction appears to "cannibalize" some would-be utilitarians. On the whole, the application of utilitarianism is substantial but does not appear to be significantly stronger under JE.

2. Experiment 2: Second Choices

Given the difficulty of evaluating the magnitude of utilitarian influence on moral judgment in Experiment 1, Experiment 2 was designed to further probe the effects of JE on the competition between outcome-based and action-based moral evaluations. If we remove the inaction option from JE and ask participants to endorse either the drop or the divert action, which action will they prefer? This method can be seen as a revealed-preferences approach: when forced to choose, would participants adhere to outcome-based or action-based moral evaluation, to utilitarianism or the intention principle?

If choosing inaction indicates an increased overall influence of action-based moral evaluation, one may expect that participants who endorsed inaction will switch to the "next best" deontological principle—the intention principle—and therefore endorse divert. Colloquially, this hypothesis might read: "I don't want to harm anybody at all, but if I must do harm, I will do it in the most unintended way possible." Alternatively, choosing inaction may not necessarily indicate a strong commitment to action-based moral evaluation. Participants could also endorse inaction because they experience difficulty deciding between the drop and divert options, and the action principle provides an easily accessible justification to avoid the decision. If difficulty is the primary factor underlying choices for inaction, we might expect these participants to fail to shift to the intention principle due to the difficulty of articulating and comprehending it (Cushman et al. 2006). In this case, they might opt for the drop option, finding utilitarianism easier to articulate. In colloquial terms: "I don't want to harm anybody at all, but if I must do harm, then at least I should save the most lives possible."

In sum, our forced-response design at the second-choice stage helps us understand the relative influence of outcome-based versus action-based moral evaluation and establish a hierarchy among the various psychological motivations present during JE. Note that it was necessary to remove the inaction option, and not the divert option, to solve this problem. Although both inaction and divert are supported by action-based principles, inaction may also stem from other reasons, such as the difficulty of the choice.

Hence, keeping inaction while removing divert would not have allowed us to identify which type of moral evaluation ultimately shapes people's reconciliation of the moral conflict. An additional reason to remove inaction and not divert was that, absent divert, the JE dilemma would have reverted into the original SE case, essentially identical to SE-Drop.

Experiment 2 expanded on Experiment 1 in another way, namely by examining whether participants would reconcile the conflict differently when asked what they *would* do versus what they *should* do. By adapting the dilemma to ask participants "What should you do?" we aimed to overcome the potential confound between normative and psychological considerations that might be present when participants respond to the "would" question (for example, "I think I should drop the man, but I would not be able to bring myself to do that.")[1] The "should" question aimed to remove this ambiguity and focus participants on normative considerations. We collected two datasets, each asking participants what would/should they do.

2.1. Participants

"Would" version. Two hundred and eight people (104 female, M_{age} = 29.76, *SD* = 9.5) recruited through Amazon Mechanical Turk (mTurk) participated in the study in exchange for $0.30. Prior to data analysis, seven participants were filtered out after failing a simple attention check (being unable to describe what decision they had just made for the previous item).

"Should" version. Three hundred and twenty-five people (179 female, M_{age} = 36.9, *SD* = 12.3) were recruited through mTurk. Prior to data analysis, 30 participants were filtered out after failing the same attention check used in the "would" version.

2.2. Procedure

Participants were randomly assigned to one of three groups: *SE-Divert*, *SE-Drop*, or *JE*. In all conditions, participants read the respective scenarios from Experiment 1 and were either asked what they would do (in the "would" version) or what they should do (in the "should" version). The options were identical to those presented in Experiment 1.

[1] We thank an anonymous reviewer for this insightful suggestion.

In the JE condition, after participants marked their choice and pressed the button to continue to the next page, they were presented with the following instruction: "Please assume that the option you chose is no longer available. What would [should] you do? (please do not mark again your previous choice)." This question aimed to probe participants' second moral preferences. Participants were presented again with the same three options and had to select one. Other than the explicit request noted above, we did not prevent participants from marking the same option again.

2.3. Results: First Choice

The first part of Experiment 2 replicated Experiment 1 (Table 2). An exact test for two-way contingency tables with structural zeros (using the same method in Experiment 1) was conducted on the "would" and "should" datasets. Results in both were highly significant (*would*: $\chi^2 = 110.051$, $p = .000$, *Phi* = .74; *should*: $\chi^2 = 162.6$, $p = .000$, *Phi* = .69), and followed the same pattern of Experiment 1. Several proportion tests (z-tests) were performed to investigate the nature of the differences.

As in Experiment 1, under SE, people were more likely to endorse diverting over doing nothing than to endorse dropping over doing nothing (*would*: 87.5% vs. 41.4%, $Z = 4.16$, $p = .000$; *should*: 73% vs. 38.4%, $Z = 3.77$, $p = .000016$). In contrast, under JE, drop was again the choice with the largest share of endorsement (42.5%), significantly higher than divert (23%; $Z = 3.632$, $p = .0003$, $d = .5$). As in Experiment 1, participants were more likely to endorse drop under JE as compared with under SE, despite having more options to choose from, but this was not a significant difference (*would*: 42.5% vs. 41.4%, $p = .92$; *should*: 45.6% vs. 38.4%, $p = .50$). As for inaction, in the "would" data, the rate of inaction significantly increased from SE-Divert to JE (from 12.5% to 34.5%, $Z = 2.93$, $p = .003$, $d = .72$). The trend was similar but not significant in the "should" data (from 27% to 35.6%, $Z = 1.28$, $p = .20$). To examine the overall significance of the rise in inaction and any effects of wording, we conducted a logistic regression with condition (SE/JE) and wording (would/should) as predictors and choice (inaction/action) as a dependent variable. The overall model was significantly more predictive than the null, $\chi^2(3) = 11.91$, $p = .007$, and condition (JE vs. SE) had a significant impact on the proportion of participants choosing inaction ($B = -.849$, $SD = .28$, $p = .002$). The effect of wording was marginally

significant (B = -.502, SD = .28, p = .07), and there was no significant interaction between condition and wording (p = .102), indicating that the increase in inaction from SE to JE is likely not specific to either wording.

Insert Table 3 about here

2.4. Results: Second Choice

We next investigated whether the participants who chose inaction were inclined towards utilitarianism or deontology as the next best option. In the "would" data, among the 30 participants who chose inaction, 19 (63%) preferred the drop option as their second choice, nine chose the divert option, one participant reiterated a choice of inaction, and one participant did not indicate a second choice. The trend was similar in the "should" data: among the 32 participants who chose inaction, 18 (56%) chose drop, 11 chose divert, and three reiterated their choice of inaction. In sum, when forced to choose between the active options, the majority of inaction participants chose the utilitarian option (drop) as their next preferred choice.

What do the results tell us about people's ultimate moral preferences? We combined these responses with the divert and drop original choices to examine the effect of the second preferences on the overall pattern of response (Table 2, last row in each of the "would" and "should" cells). In other words, we used this analysis to consider the overall level of preference for drop versus divert across the total population of participants in JE when inaction is removed as an option. A Z-test on the "ultimate action choice" JE data was highly significant in both datasets (*would*: Z = 4.13, p < .00001, d = .73; *should*: Z = 4.59, p < .00001, d = .794), showing a significant preference for the utilitarian option, drop (65% in both datasets), over the deontological option, divert (31%, 33.7%).

2.5. Discussion

Participants' first-choice responses replicated our findings from Experiment 1. Endorsement of the divert option decreased under JE, consistent with a decrease in the influence of the intention principle.

Meanwhile, endorsement of the inaction response increased under JE, as compared with the divert SE case, which is consistent with an increase in the use of the action principle. Utilitarianism remained a strong influence across JE and SE. The "would" versus "should" wording did not impact these results.

In Experiment 2, we aimed to understand the effect of JE on moral preferences when people are "forced" to act. We did so by evaluating the second choices of participants who first chose inaction under JE. Would these participants apply a deontological pattern of response consistently and endorse divert, or switch to a utilitarian pattern of response and endorse drop? We found that when inaction was not available, participants were more likely to switch to utilitarianism. Integrating these preferences with the first preferences for drop and divert clarified the pattern of relative application of outcome-based versus action-based moral evaluation in the trolley problem. Under SE, more people find diverting (to save three people) a more acceptable option than dropping (to save five people). Yet, under JE with no inaction option available, nearly twice as many people prefer drop to divert. If we view these results through the lens of the candidate actions (drop, divert, inaction), there appears to be a dramatic reversal of preferences between drop and divert under SE versus JE.

Through the lens of the candidate moral principles, however, we see a more nuanced effect. Utilitarianism is the first preference of many participants (first-choice droppers), and the second preference of most remaining others. Participants whose first choice was inaction are not likely to shift to the next best action-based deontological principle but instead are likely to shift to utilitarianism. This shift suggests that the first inaction choices do not necessarily reflect a stronger endorsement of deontological reasoning under JE, but rather an endorsement of one easily accessible justification to avoid action in a complex moral conflict. Absent such an option, choices shift to utilitarianism, which is again easier to articulate than the intention principle. Overall, then, the shift in preferences reveals that JE can weaken the influence of more difficult forms of action-based moral evaluation and strengthen the influence of outcome-based moral evaluation. This is a nuanced effect because the increased endorsement of utilitarianism among participants who would have preferred inaction suggests that these participants may

favor utility maximization only when forced to choose between two varieties of harmful action: drop and divert.

Some caution is necessary in the interpretation of the choices. Notably, the results do not speak to people's original commitment to the action principle, nor do the results suggest that people necessarily hold this principle "lightly." Instead, the results reveal that an endorsement of one deontological principle—the action principle—does not necessarily entail or produce a strong endorsement of deontological reasoning more broadly. Besides a greater influence of outcome-based evaluation under JE and the relative difficulty of articulating the intention principle, people may also avoid the next-best deontological option simply because they find the intention principle less persuasive.

Overall, by considering both first and second choices together, Experiment 2 helps establish a hierarchy among the moral principles available under JE. Utilitarianism, or outcome maximization, is the highest-rated first choice overall and the highest-rated second choice of the inaction participants. The intention principle is the lowest-rated first and second choice. In between these two appears to be the action principle, with increased influence under JE but lower influence overall as compared with utilitarianism.

3. Experiment 3: Preference Reversal in Moral Judgment

Experiment 2 provides some evidence that JE reverses people's moral preferences from deontological to utilitarian, through their second choices. We designed Experiment 3 to test whether we could obtain the same apparent preference reversal directly, without asking participants for their second choice. To avoid this step, we switched from asking participants what they would/should do, a question that presupposes an option to do nothing, to asking participants to directly compare the moral status of the drop and the divert actions against each other, leaving aside the matter of how each compares to inaction. This design, which we adapted from Paharia and colleagues (2009), makes for a more streamlined test of our hypothesis.

3.1. Participants

Three-hundred and thirty individuals (198 male, M_{age} = 29.63, *SD* = 9.66) recruited through Amazon Mechanical Turk (mTurk) participated in the study online in exchange for $0.30.

3.2. Procedure

Participants were randomly assigned to one of three groups: *SE-Divert*, *SE-Drop*, or *JE*. They read the scenario used in Experiments 1 and 2. In the SE conditions, participants were instructed to "consider flipping the switch." Following the design of Paharia et al. (2009), they were then asked, on a scale of 1 (not at all moral) – 10 (very moral), "How moral do you think your behavior would be in this decision?" In the JE condition, the participants were asked: "In which case would your behavior be more moral?" They responded using an eleven-point scale anchored at -5 as "flipping switch A is much better than flipping switch B," 0 as ''both choices are equally moral/immoral," and 5 as "flipping switch B is much better than flipping switch A" (see Paharia et al., 2009). In the JE scenario, switch A was *Drop* and switch B was *Divert.*

3.3. Results

The SE results replicated the previous findings and showed that divert was rated as significantly more moral (*M* = 6.0357, *SD* = 2.36) than drop (*M* = 5.00, *SD* = 2.78), $t(224) = 3.012$, $p = .003$, $d = .41$.

The JE condition was analyzed using a one-sample t-test with zero (scale midpoint) as the test value. The results revealed a significant effect, with the average response located to the left of the midpoint (*M* = -.45, *SD* = 2.16), indicating that drop was rated as significantly more moral than divert when the two cases were evaluated together ($t(103) = -2.131$, $p = .035$, $d = .21$). We further bootstrapped the mean to examine the robustness of this result (Efron 1979). A bootstrap simulation based on 5,000 samples resulted in a confidence interval of (-.87, -.02), reaffirming the significance of the preference reversal.

3.4. Discussion

Replicating the pattern we observed for second choices in Experiment 2, we find that evaluating the two problems together shifts participants' relative moral attitudes concerning the utilitarian (drop) and deontological (divert) options. Specifically, participants consider drop to be less moral than divert when

these actions are evaluated separately, but shift to considering drop to be more moral than divert when the actions are evaluated together. This finding demonstrates an increased application of outcome-based moral evaluation under JE, at the expense of action-based moral evaluation.

III. General discussion

Our study of trolley problems under JE versus SE yields two general findings. First, under JE, we find that participants show an increased preference for inaction (Experiments 1 and 2). One likely explanation for these results appeals to the finding that under JE, people are more likely to engage in controlled, principled comparisons between competing options (Bazerman et al. 1998, Hsee et al. 1999, Kahneman and Ritov 2004, Ritov and Baron 2011). People may account for their aversion to the drop case by endorsing the action principle, according to which it is impermissible to cause harm. Applying this principle to the divert case, they conclude that inaction is again required.

There is a principle that would allow people to justify their divergent intuitions in the drop and divert cases. This is the intention principle, which received substantial attention and some support in the philosophical literature (Foot 1967, Thomson 1985, Fischer and Ravizza 1992) and appears to underlie many patterns of moral and ethical decision making (Greene et al. 2001, Paharia et al. 2009, Royzman and Baron 2002). However, past research suggests that people are either unable to spontaneously articulate this principle or they reject it as unpersuasive (Cushman et al. 2006). Conversely, people are able to articulate the action principle, and when they reflect back on their moral choices, they also tend to endorse it (Cushman et al. 2006, Hauser et al. 2007). However, the methods employed in these past studies were explicitly described as insufficient to establish whether explicit consideration of the action principle plays a causal role in moral judgment. The present study helps to resolve this matter, illustrating that this non-utilitarian assessment can intervene in the decision-making process and alter moral judgment in its formation, and identifying the specific conditions in which it does so. We find that deliberation over deontological principles is not limited to a post-hoc justification or rationalization processes (Haidt 2001), but can have an effect on moral decision making.

Our second finding is an increased utilitarian influence on moral judgment under JE, specific to the relative moral value of divert versus drop. Across our studies, the rate of the fully utilitarian decision was the highest when we asked participants to choose between all options (Experiments 1 and 2). People were more likely to endorse the fully utilitarian decision of drop than the deontological decision of divert when comparing the two cases (Experiment 3), and this was also true for people who actually preferred inaction overall (Experiment 2). In other words, such individuals reason that if they must choose between two bad actions, they prefer the action that saves more lives, even if the action itself is more aversive. This helps us understand the hierarchical ranking of moral interests under JE: when required to abandon a deontological principle prohibiting all harmful action, many people revert to a utilitarian principle rather than maintaining the next-best deontological stance.

Although this second finding provides some support for the position that utilitarian judgments are enhanced under JE, some of the effect emerges only under the constraint of second choices (Experiment 2) or limited choices (Experiment 3). This might be considered a surprising finding, given that utilitarianism is hypothesized to be a result of controlled processing (Greene et al. 2001, Greene et al. 2004), and JE is hypothesized to enhance such processing (Bazerman et al. 1998, Hsee et al. 1999, Kahenman and Ritov 2004, Ritov and Baron 2011). Why did we not observe a more dramatic shift to utilitarian choice?

One possible explanation is that JE is not necessary to highlight the principle of utilitarianism in the trolley problem. Both the drop case and the divert case involve a tradeoff between lives, with clear numerical comparisons. It is unlikely that participants failed to notice this fact or to consider the possibility of a principled decision to maximize welfare, even under SE. In fact, past research indicates that participants readily consider the principle of utility maximization when responding to moral dilemmas (reviewed in Cushman et al. 2010). Joint evaluation may make the utilitarian comparison more salient, but without substantially enhancing sensitivity to utilitarian considerations because those considerations were already salient under separate evaluation.

A second possible explanation has to do with the differing value of the inaction option across choice settings and domains. The majority of previous SE/JE studies examined decision-making in consumption or employment, in settings where choosing neither option made little sense: why should someone choose to accept no job offer rather than one of the two (Bazerman et al. 1994), or no ice cream rather than one ice cream (Hsee et al. 1999)? There is little value in choosing to be empty-handed, at least when the choice is between a small number of well-defined alternatives. In contrast, when it comes to moral choices, refraining from action is often considered to be the right choice (Ritov and Baron 1999). Whereas leaving empty-handed when facing consumption choices poses few benefits, keeping one's hands clean when facing moral choices has the value of following a moral principle. It is perhaps less surprising, then, that JE increases inaction in the moral dilemma by convincing some participants of the moral value of refraining from causing harm. Among those individuals, however, a large fraction believes that if you *must* harm a person, you should at least save as many others as possible.

A third possible explanation for the relatively modest increase in utilitarian response together with an increase in inaction is decision avoidance. Choosing between two aversive actions may itself be aversive. When people need to make difficult decisions, they often avoid them, resorting to a non-consequentialist choice (Tversky and Shafir 1992, Shafir 1994, Anderson 2003). Conflicts between moral options are difficult to manage, and the number of lives at stake may increase anticipated regret or blame. Both these factors – selection difficulty and anticipated regret – were shown to increase avoidance (Anderson 2003), and inaction may mitigate the negative emotions associated with them (Luce 1998).

Taken together, our findings offer a counterpoint to the prevailing assumption that principled reasoning contributes little to moral judgment (see also Pizarro & Bloom 2003; Paxton & Greene 2010, Bartels 2008). For instance, in his "social intuitionist" model, Jonathan Haidt (2001) argues that people rarely rely on moral principles except as post-hoc rationalizations of prior intuitive judgments. Whereas the model posits that moral principles are mostly ineffective, our evidence indicates that the application of moral principles can alter moral judgment in its formation. We accept Haidt's proposition that intuition drives much of moral judgment and Greene's evidence that much of this intuition is consistent with

deontological ethics. But, Haidt's results are focused on conditions where System 1 processes dominate and we are highlighting a System 2 lever (joint decision-making) that can powerfully shape moral decisions. Prior research consistently found that joint evaluation contexts produce greater reflection and deliberation. We find that the more deliberative processes prompted by joint decision-making can have an effect on moral decision-making and move judgments in the direction of articulable moral principles. This suggests that deliberative processes have a causal effect on moral judgments, not only on people's stated justifications for these judgments. According to the dual process model (Greene 2008), deliberation should principally favor utilitarian outcome maximization over action-based prohibitions. Our findings nuance this prediction. Indeed, conditions promoting reflection yield a big camp that favors utilitarianism, yet also a substantial camp that favors the strong deontological position that prohibits harmful action in the service of welfare maximization. This suggests that a salient normative conflict encourages people to seek principles that would impose moral consistency, in the sense that such principles would provide the same answer to similar cases. This form of reasoning accepts principles that can be either outcome- or action-based, but is less sensitive to principles that offer nuanced distinctions between similar cases. The result, reported in Experiments 1 and 2, is that people may split into two opposite hardline camps: either utilitarian harm is always justified, or else it is never justified. That polarization in moral judgment could be the outcome of principled reasoning may be beneficial or worrisome, depending on the context. When the divide clarifies issues of disagreement, it may be beneficial. But when creating coalitions and reaching agreement is important, moral polarization is a cause for worry. Whether this effect persists across other moral, political, and organizational dilemmas is an issue for further research.

Although our paper draws on a more abstract set of problems, we firmly believe that it taps into some of the core concerns facing management today. Moral judgments are components of daily life and daily decisions in organizations. Analogous to the competing moral principles highlighted in our current work with trolley problems, managers and employees are confronted with and often conflicted in the pursuit of profit- or growth-oriented outcomes, versus the means of achieving those outcomes, which may entail disruptions and even harm to various stakeholders. How do people or managers reconcile the ever-

recurring conflict between increasing welfare and avoiding harm in circumstances that make such trade-off inevitable? In what settings would they be more likely to choose welfare and in what settings would they avoid action? And when people or managers have an impulsive response that something is morally wrong, would that impulse hold up with greater deliberation? Although our discussion of the shifting value of the inaction option suggests that caution is needed when extrapolating beyond choice domains, we believe that our findings shed light on these important problems and offer some potential organizational implications for management scholars and practitioners. A concrete example is provided by Bohnet, van Geen, and Bazerman (2016), who explored a related moral problem, namely whether to weigh gender when making hiring decisions for gender-stereotyped tasks. Bohnet et al. (2016) find massive gender-based discrimination when employers choose employees under separate evaluation. Their participants preferred males for math tasks and females for verbal tasks. But, under JE, employers switched from the intuitively appealing and stereotypical set of choices to the more utilitarian set of choices, hiring employees based on performance, not gender. This result further demonstrates the potential of the JE framework to encourage utilitarian judgment in managerial settings that involve moral and ethical concerns.

The results in the current paper and in Bohnet et al. (2016) are consistent with earlier work on joint versus separate preference reversals, which showed that people focus on social comparison processes under separate decisions, yet largely ignore social comparisons under joint decision making (Bazerman et al., 1994). The more deliberative JE mode may prompt certain advantageous choices, such as recognition of one's own economic well-being (Bazerman et al. 1994) or of the value of public goods (Irwin et al. 1993). The broader managerial pattern is that people are more likely to follow their deontological instincts under separate decisions, and to deliberate more under joint evaluation. As they deliberate, people are likely to seek out decisions with a clear, accessible principled basis. This has clear implications for the mindset that managers want to create within their employees in ethical and moral contexts. Managers should particularly strive to provide decision-makers in their organizations who are

called upon to make routine moral judgments with principles that provide a rationale for the manager's desired outcome.

Yet the nuanced findings that emerge from our work suggest that organizations hoping to harness joint evaluation to nudge people toward more beneficial choices in settings that involve moral conflict may encounter some increased inaction. One way to avoid this effect is adopting inaction-discouraging choice architecture. Our studies provide some evidence that under an architecture of limited options that includes only active choices (study 3) or asks people to set aside their inaction preferences (study 2), the application of principled reasoning may result in more beneficial choices. When managers face moral conflict due to a requirement to lay off one of two good employees, or experience ethical qualms over shutting down failing programs, inaction is likely (Barak-Corren and Bazerman, 2016). In such cases and others, organizations might be wise to acknowledge people's preference for inaction and, where appropriate, construct choice-sets that require employees to set this preference aside and choose amongst available beneficial actions. Our paper highlights the need for more research on whether and when being confronted with moral conflict leads to inaction – and how managers can encourage action in such circumstances, if needed.

References

Barak-Corren, N., Bazerman, M.H. Forthcoming.Moral conflict and inaction. *Organ. Dynamics*.

Baron, J., Gürçay, B., Moore, A., Starcke, K. 2012. Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese* **189**(1) 107-117.

Baron J, Spranca M. 1997. Protected values. *Organ. Behavior Human Decision Processes* **70** 1-16.

Bartels, DM. 2008. Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition* **108**(2) 381-417.

Bauman, CW., McGraw, AP., Bartels, DM., & Warren, C. 2014. Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Soc. Personality Psychol. Compass* **8** 536–554.

Bazerman, M., Tenbrunsel, A., Wade-Benzoni, K. 1998. Negotiating with yourself and losing: Understanding and managing conflicting internal preferences. *Acad.Management Rev.* **23** 225-241.

Bazerman, M., Loewenstein, G., White, S. 1992. Reversals of preference in allocation decisions: Judging an alternative versus choosing among alternatives. *Admin. Sci. Quart.* **37** 220-240.

Bazerman, M., Schroth, H., Shah, P., Diekmann, K., Tenbrunsel, A. 1994. The inconsistent role of comparison others and procedural justice to hypothetical job descriptions: Implications for job acceptance decisions. *Organ. Behavior Human Decision Processes* **60** 326-352.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*.

Bohnet, I., van Geen, A., Bazerman, M.H. In press. When performance trumps gender bias: Joint versus separate evaluation. *Management Science.*

Cushman, FA. 2013. Action, outcome, and value: A dual-system framework for morality. *Personality Soc. Psych. Rev.* **17**(3) 273–292.

Cushman, F., Young, L. 2011. Patterns of moral judgment derive from nonmoral psychological

representations. *Cognitive Sci.* **35**(6) 1052–1075.

Cushman, F., Young, L., Greene, J. 2010. Multi-system moral psychology. In JM Doris & T. M. P. R. Group (Eds.), *The Oxford Handbook of Moral Psychology* (Oxford University Press, New York).

Cushman, F., Young, L., Hauser, M. 2006. The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psych. Sci.* **17**(12) 1082-1089.

Efron, B. 1979. Bootstrap methods: Another look at the jacknife. *Annals of Statistics* **7** 1-26.

Fischer, J., Ravizza, M. 1992. *Ethics: Problems and Principles* (Holt, Rinehart & Winston, New York).

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

Graham, J., Haidt, J., Nosek, B. 2009. Liberals and conservatives rely on different sets of moral foundations. *J. Personality Soc. Psych.* **96**(5) 1029.

Greene, J. 2004. The neural bases of cognitive conflict and control in moral judgment. *Neuron* **44**(2) 389–400.

Greene, J. 2008. The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology* (Vol. 3) (MIT Press, Cambridge, MA).

Greene, J., Sommerville, R., Nystrom, L., Darley, J., Cohen, J. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* **293** 2105–2108.

Haidt, J. 2001. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psych. Rev.* **108** 814-834.

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

Hsee, C., Loewenstein, G., Blount, S., Bazerman, M. 1999. Preference reversals between joint and separate evaluation of options: A review and theoretical analysis. *Psych. Bull.* **125** 576-590.

Kahneman, D., Ritov, I. 1994. Determinants of stated willingness to pay for public goods: A study in the headline method. *J. Risk Uncertainty* **9** 5-38.

Kelman, M., Kreps, T. 2014. Playing with trolleys: Intuitions about the permissibility of aggregation. *J. Empirical Legal Stud.* **11** 197–226.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., Damasio, A. 2007. Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature Materials* **446**(7138) 908–911.

Luce, M 1998. Choosing to avoid: Coping with negatively emotion-laden consumer decisions. *J Consumer Res* **24**(4) 409-433.

Lombrozo, T. 2009. The role of moral commitments in moral judgment. *Cognitive Sci.* **33** 273–86.

Mill, JS. 1863. *Utilitarianism*.

Paharia, N., Kassam, K., Greene, J., Bazerman, M. 2009. Dirty work, clean hands: The moral psychology of indirect agency. *Organ. Behavior Human Decision Processes* **109**(2) 134-141.

Paxton, J., Greene, J. 2010. Moral reasoning: Hints and allegations. *Topics Cognitive Sci.* **2**(3) 511–527.

Paxton, J., Ungar, L., Greene, J. 2011. Reflection and reasoning in moral judgment. *Cognitive Sci.* **36**(1) 163-177.

Petrinovich, L., O'Neill, P. 1996. Influence of wording and framing effects on moral intuitions. *Ethology Sociobiology* **17** 145–71

Pizarro, D., Bloom, P. 2003. The intelligence of the moral intuitions: comment on Haidt (2001). *Psychol Rev* **110**(1) 193–198.

Ritov, I., Baron, J. 2011. Joint presentation reduces the effect of emotion on evaluation of public actions. *Cognition Emotion* **25** 657–675.

Ritov, I., Baron, J. 1999. Protected values and omission bias. *Organ Behavior Human Decision Processes* **97** 79-94.

Schwitzgebel, E., Cushman, F. 2012. Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind Language* **27**(2) 135-153.

Shafir, E. 1994. Uncertainty and the difficulty of thinking through disjunctions. *Cognition* **50** 403-430.

Shafir, E., Tversky, A. 1992. Thinking through uncertainty: Nonconsequential reasoning and choice. *Cognitive Psychology* **24** 449-474.

Shallow, C., Rumen, I., Medin, D. 2011. Trolley problems in context. *Judgment Dec. Making* **6**(7) 593-601.

Shultz, T., Léveillé, E., Lepper, M. 1999. Free choice and cognitive dissonance revisited: Choosing "lesser evils" versus "greater goods". *Personality Soc. Psych. Bull.* **25**(1) 40-48.

Tenbrunsel, A., Wade-Benzoni, K., Messick, D., Bazerman, M. 2000. Understanding the influence of environmental standards on judgments and choices. *Acad.Management J.* **43**(5) 854-866.

Tversky, A., Shafir, E. 1992. Choice under conflict: The dynamics of deferred decision. *Psychological Science,* **3** 358–361

Thomson, J. 1985. The trolley problem. *Yale Law J.* **94**(6) 1395–1415.

Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., Ditto, P. H. 2009. The motivated use of moral principles. *Judgment and Decision Making*, **4**(6) 476-491.