Available online at www.sciencedirect.com
ScienceDirect
Cognition 108 (2008) 381–417
COGNITION
www.elsevier.com/locate/COGNIT

Principled moral sentiment and the flexibility of moral judgment and decision making

Daniel M. Bartels *

Center for Decision Research, University of Chicago, 5807 S Woodlawn Avenue, Chicago, IL 60637, USA

Received 28 October 2007; revised 1 March 2008; accepted 4 March 2008

Abstract

Three studies test eight hypotheses about (1) how judgment differs between people who ascribe greater vs. less moral relevance to choices, (2) how moral judgment is subject to task constraints that shift evaluative focus (to moral rules vs. to consequences), and (3) how differences in the propensity to rely on intuitive reactions affect judgment. In Study 1, judgments were affected by rated agreement with moral rules proscribing harm, whether the dilemma under consideration made moral rules versus consequences of choice salient, and by thinking styles (intuitive vs. deliberative). In Studies 2 and 3, participants evaluated policy decisions to *knowingly do harm* to a resource to mitigate greater harm or to *merely allow* the greater harm to happen. When evaluated in isolation, approval for decisions to harm was affected by endorsement of moral rules and by thinking style. When both choices were evaluated simultaneously, total harm – but not the do/allow distinction – influenced rated approval. These studies suggest that moral rules play an important, but context-sensitive role in moral cognition, and offer an account of when emotional reactions to perceived moral violations receive less weight than consideration of costs and benefits in moral judgment and decision making.

Keywords: Morality; Judgment; Decision making; Values; Ethics; Intuition; Emotions; Reasoning; Moral rules; Moral dilemmas

* Tel.: +1 773 834 9857.
E-mail address: bartels@uchicago.edu.

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.03.001

1. Introduction

Understanding the processes underlying moral cognition is an active endeavor in psychology; many frameworks have been developed recently with this aim in mind (Baron & Spranca, 1997; Cushman, Young, & Hauser, 2006; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Nichols & Mallon, 2006; Tetlock, 2003). The current studies cross-pollinate between them, testing their predictions concerning the use of two processes: deontology (adhering to moral rules) and consequentialism (balancing costs and benefits). Although many frameworks implicate these processes, little theoretical synthesis has been achieved, perhaps because few studies examine generalization across contexts (i.e., different types of judgment and choice situations).

The current studies aim for synthesis and find that moral cognition is predicated on moral rules, emotional reactions, and assessments of costs and benefits. Specifically, they suggest (1) that contexts that direct attention to violations of moral rules generate deontology-consistent emotional reactions, (2) that deontological response is diminished in contexts that direct attention to utilitarian considerations, and (3) that contextual factors interact with situation-specific values and individual differences to shape moral judgment and choice.

1.1. Consequentialism and deontology as psychological processes

Deontology and consequentialism are two positions in normative ethics that use different bases for judging the moral status of acts and imply different cognitive processes. The goodness of an act's consequences is a relevant feature in both, but it is the *only* relevant feature for consequentialism, which mandates producing the best consequences *by any means* (Darwall, 2003a; Pettit, 1993). Because its only inputs are consequences, the right action is the one that produces the best outcome. In contrast, deontology checks certain qualities of actions against rules that must be honored, thereby setting up constraints on action. The output is that some acts – like knowingly harming an innocent person – are forbidden (even as means to morally-obligatory ends; see Darwall, 2003b; Davis, 1993). In many contexts, deontology and consequentialism yield the same judgments regarding harmful acts, since doing harm often leads to worse consequences overall. But consequentialism treats deontological constraints as rules of thumb which must be broken in cases where doing so would produce better consequences.

The current studies focus on factors that *elicit* or *suppress* judgments consistent with *utilitarianism*, a version of consequentialism that mandates promoting "the greatest good for the greatest number", where goodness is assessed from an impersonal point of view, and each person's welfare counts equally. Because utilitarianism requires summing the degree of satisfaction of *welfare interests* (i.e., basic needs) across individuals, it mandates a simple process for the choices participants face in the current studies: Count the number of lives saved by each alternative and choose accordingly. Choices that maximize lives saved are consistent with utilitarianism. Otherwise, they are not.

1.2. Ethical dilemmas

Philosophers often develop normative arguments in the context of ethical dilemmas intended to distill real-world problems into their "essential" features. Recently, researchers have used responses elicited by dilemmas to develop descriptive accounts of moral cognition. For example, reactions to the (bystander) "trolley problem" (Foot, 1967) – where flipping a switch diverts a runaway train car threatening to kill five railway workers onto a track where it will kill one railway worker – are compared to reactions to the "footbridge problem" (Thomson, 1985) – where the only way to save the five railway workers is to stop the train by pushing a fat man off a footbridge onto the tracks below.

People tend to judge that flipping the switch – causing one death to save five – is acceptable, but that pushing a man to his death is wrong (e.g., Cushman et al., 2006; Mikhail, 2007; Waldmann & Dieterich, 2007). If someone were following a broad deontological rule, such as "It is forbidden to intentionally kill someone (regardless of the consequences)," she might judge both actions wrong. (Deontological constraints are usually more narrowly-directed than this rule, and so on some views, flipping the switch does not count as intentional killing.) In contrast, an exclusive focus on utilitarian considerations (one life versus five) would make both actions acceptable. The intuition that these cases should be treated differently cannot be characterized by strict versions of either normative strategy.

The current studies build on previous research suggesting moral judgment is influenced by (a) emotional reactions to action, (b) magnitude of consequences favoring the action (i.e., lives to be saved), and by (c) individual differences in propensity to incorporate emotional reaction in judgment (Greene et al., 2001; Nichols & Mallon, 2006). The current studies find that moral cognition is highly context-sensitive and that including each of these influences is necessary to account for this context-sensitivity.

Studying *only* responses to trolley problems might limit generalization. Accounts developed around these responses intend to explain *de facto* morally-motivated processes, describing sensitivities to features of dilemmas that generalize across people – some aim to identify universal principles of moral cognition (e.g., Hauser, 2006; Mikhail, 2007). However, in trying to develop parsimonious theories about the fundamental (i.e., not context-dependent) laws of human thought that generalize across a wide range of content domains, these accounts may reflect what Tetlock, Peterson, and Lerner (1996) refer to as the *anti-context* and *anti-content* biases (see also Goldstein & Weber, 1995; Rettinger & Hastie, 2001; Rettinger & Hastie, 2003; Shweder, Mahapatra, & Miller, 1987). The current studies examine, among other factors, the influence of the content of the scenarios under consideration. Whereas many studies use only problems involving threats to human life, studies in the judgment and decision making literature confront participants with threats to resources that are not treated as having moral significance by all individuals or cultures. The literature on "protected values" (Baron & Spranca, 1997) focuses on the restrictive tradeoff rules participants have for certain kinds of (moral) goods and suggests that situation-specific values engender nonconsequentialist decision strategies.

1.3. Protected values

Decision making researchers often assume that people choose so as to produce the best consequences (broadly construed) from their point of view. It is difficult to overstate the similarity between models of decision making and consequentialist philosophy. For example, Amir and Ariely (2007, p. 150) write: "The common view that both laypeople and decision scientists alike hold is consequentialist. That is, people make decisions according to their set of preferences by searching for an optimum, a local optimum, or a close enough estimate when exact algorithms are too costly."

Recent research reports a link between moral values and a *lack of concern* for consequences. Some decisions appear to be driven less by the consequences associated with an action than by moral rules concerning the ways these resources should be treated (Baron & Spranca, 1997; Fiske & Tetlock, 1997; Medin, Schwartz, Blok, & Birnbaum, 1999; Tenbrunsel & Messick, 1999). Moralized goods are often bestowed a "protected" status, and people react strongly to proposed tradeoffs on moral grounds. One formulation that is likely to be recognizable is "You can't put a price on human life!" (e.g., Tetlock, Kristel, Elson, Green, & Lerner, 2000).

The literature on protected values (PVs) focuses on nonconsequentialist moral choices. In some studies, participants read scenarios where harmful actions promoted the best outcomes for a resource (Ritov & Baron, 1999). For example, participants read that the only way to save 20 species of fish upstream was to open a dam. However, opening the dam would kill two species of fish downstream. Participants were asked whether they would open the dam, for the maximum number killed downstream at which they would do so, and later, about the permissibility of killing fish. Those who judged that killing fish is impermissible, *no matter how great the benefits*, were scored as having a PV. People with PVs were less likely to open the dam, and some said they would not want to cause the loss of a single species, *even though it means losing all 20 species*.

These results might tempt some to charge moral cognition with a kind of rigidity. That is, when moral values are implicated in a decision, highly informationally-constrained decision procedures are adopted that may produce errors (from a consequentialist perspective). For example, after espousing the competence of people's commonsense moral intuitions, Hauser (2006, p. 11) writes "We should not conclude from the discussion thus far that our intuitions always provide luminary guidance for what is morally right or wrong. As the psychologist Jonathan Baron explains, intuitions can lead to unfortunate or even detrimental outcomes. Omission bias causes us to favor the termination of life support over the active termination of a life, and to favor the omission of a vaccination trial even when it will save the lives of thousands of children although a few will die of a side effect. As Baron shows, these errors stem from intuitions that appear to blind us to the consequences of our actions."

However, a *complete* lack of concern for consequences seems implausible. If people hold PVs for things they cherish (e.g., family), they likely care immensely about consequences for these resources. For example, consider a parent who is reluctant to vaccinate her son because she might harm him – he might suffer from side effects. If her focus is shifted to the greater risk posed by not vaccinating, she may now feel a (moral) imper-

ative to vaccinate him. Research on PVs suggests that constraints like "do no harm" (Baron, 1996) prevent people from giving the "appropriate" weight to consequences favoring harmful action – (e.g., the risk mitigated by vaccination; Ritov & Baron, 1999).

Recall that consequentialism treats deontological constraints as rules of thumb that must be broken in cases like these. In the current studies, people sometimes behave like utilitarians who (perhaps reluctantly) forego their moral constraints to bring about the best consequences. In other contexts, they give deontology-consistent judgments, suggesting a context-sensitive role for constraints: when people's attention is directed to an action (as in the footbridge case), affectively-charged moral rules, like those associated with PVs, exert a larger influence on judgment than in contexts where people's attention is directed to the consequences of the action.

1.4. Overview

The current studies examine moral judgment in two contexts. Study 1 examines responses elicited by ethical dilemmas, finding that these judgments are predicated on moral rules, emotional reactions, and assessments of costs and benefits. Study 1 suggests that contexts that direct attention to violations of moral rules generate deontology-consistent emotional reactions, but that deontological response can be outweighed by contextual factors that direct attention to utilitarian considerations. Studies 2 and 3 use Study 1's findings as a basis for new predictions about the context-sensitive role of moral rules (PVs, in this case) in the evaluation of life-and-death public policy decisions.

2. Study 1 – Trolley problems, intuitive deontology, and catastrophe cases

Philosophers – and in particular, deontologists – developed variants of the "trolley problem" to elicit intuitions that support their normative arguments. Deontologists use this method to suggest that when harmful actions are judged impermissible, as in the footbridge case, consequentialism is an inadequate moral theory. More generally, Donagan (1977, p. 183) notes: "Common morality is outraged by the consequentialist position that, as long as human beings remain alive, the lesser of two evils is always to be chosen." Recall that deontologists treat actions, and not their outcomes, as the proper unit of analysis. They further argue that because constraints can be narrowly-directed, actions with identical outcomes can represent different kinds of actions and elicit different judgments (Davis, 1993; Quinn, 1989). For example, *killing* can be impermissible while *letting die* can be permissible.

Utilitarians treat many of these distinctions as irrelevant and suggest that people, on a sober second thought, would agree. They sometimes argue that intuitions elicited by trolley dilemmas are untrustworthy. For example, Hare (1981, p. 139) writes: "Undoubtedly, critics of utilitarianism will go on trying to produce examples which are both fleshed out and reasonably likely to occur, and also support their argument. I am prepared to bet, however, that the nearer they get to realism and specificity, and the further from playing trains – a sport which has had such a fascination for them – the

more likely the audience is, on reflection, to accept the utilitarian solution." Other utilitarians argue that because people's intuitions are often elicited by "morally irrelevant" features, they should be discarded, and rational analysis should form the basis of ethical theory (for a book's worth of examples, see Unger, 1996).

Deontologists are not the only philosophers who construct cases to support their arguments, however. Consequentialists have used "catastrophe cases" to elicit intuitions that are difficult to square with deontology, as illustrated in one deontologist's reply to such cases:

> "We can imagine extreme cases where killing an innocent person may save a whole nation. In such cases it seems fanatical to maintain the absoluteness of the judgment, to do right even if the heavens will in fact fall. And so the catastrophic may cause the absoluteness of right and wrong to yield, but even then it would be a non sequitur to argue (as consequentialists are fond of doing) that this proves that judgments of right and wrong are always a matter of degree, depending on the relative goods to be attained and harms to be avoided. I believe, on the contrary, that the concept of the catastrophic is a distinct concept just because it identifies the extreme situations in which the usual categories of judgment (including the category of right and wrong) no longer apply." (Fried, 1978, p. 10).

Normative ethical theories intend to rationalize intuitive judgment and guide deliberative moral reasoning. And, in cases where intuition and deliberation conflict, they aim for resolution: either discard the intuition (Unger, 1996) or add a "catastrophe clause" to accommodate it (Fried, 1978). Study 1 examines preferences elicited by standard and catastrophe case dilemmas. Specifically, Study 1 is motivated by dual-process models that link utilitarian judgment to deliberation and deontology-consistent judgment to intuition.

In Study 1, participants were presented with 14 dilemmas. Each dilemma had a standard version – where six people will die if one is not harmed – and two modified versions: a "vivid" and a "catastrophe" variant. The "vivid" variants add a redescription of the harmful act that is intended to make the scenario more (negatively) affect-laden, triggering sentiment like the outrage described by Donagan (above) and studied by Tetlock et al. (2000). Moral outrage is the "principled moral sentiment" that motivates a number of hypotheses in the current studies. In Study 1, this outrage is expected to elicit deontology-consistent response. The "catastrophe" variant describes the group at risk as 20, rather than six people, making the cost of adhering to a deontological constraint more grave: Participants faced the decision of imparting harm for a net savings of 19 (rather than 5). I expect preferences more consistent with utilitarianism for these scenarios.

2.1. Attributing utilitarianism to cold deliberation and deontology-consistent responding to an emotional reaction

Judgments elicited by dilemmas have proved useful exploring the contribution of automatic and controlled process to moral judgment. (e.g., Cushman, Young,

& Hauser, 2006; Greene, Nystrom, Engell, Darley, & Cohen, 2004; Greene et al., 2001). To revisit the examples earlier, researchers have compared reactions to the bystander and footbridge versions of the trolley problem. Greene et al. (2001) argue that moral judgment is largely a function of the excitation and inhibition of emotional process. People's aversion to pushing the fat man in the footbridge case, they argue, is attributable to an emotional reaction elicited by the up-close and "personal" nature of the act that differs from the "impersonal" nature of flipping the switch in the bystander case. In the most widely-publicized of the trolley problem studies (cited more than X times as of 4/28/08), Greene et al. present as evidence for their claim greater activation in brain areas associated with emotional functioning for "personal" dilemmas and greater activation in areas associated with working memory for "impersonal" dilemmas.

Greene et al. also offer an "inhibition hypothesis" whereby deontology-consistent response is over-ridden by deliberation. They argue that (rarely-observed) utilitarian judgments for "personal" dilemmas are produced by actively suppressing the affectively pre-potent, deontology-consistent response to judge "personal" harm impermissible. Other studies manipulate ancillary emotions and find effects consistent with the idea that negative emotional reactions fuel deontology-consistent responses. For example, Valdesolo and DeSteno (2006) found that inducing positive affect before presenting the footbridge case (which they argue diminished the negative affect "signal" associated with the moral violation) led to more utilitarian responses.

More recent studies conducted by Greene et al. (2004) and Greene, Morelli, Lowenberg, Nystrom, and Cohen (2008) have further explored the inhibition hypothesis, characterizing the controlled processes implicated in over-riding the pre-potent response. These studies have used observed judgments to partition classes of responses that are then analyzed. Specifically, they have contrasted reaction times and brain activation exhibited for trials in which participants respond in a deontological manner to "personal" dilemmas, to reaction times and brain activation exhibited for trials in which participants respond in a utilitarian manner to personal dilemmas.

Areas of the brain associated with higher-order cognition are more active for trials in which participants give utilitarian judgment, and participants respond slower (Greene et al., 2004), especially under cognitive load (Greene et al., 2008). This approach has gone a long way in describing the processes at work when people have judged one way or another. However, these follow-up studies have not manipulated contextual factors (e.g., magnitude of consequences favoring sacrifice, judgment elicitation procedure) to make predictions about levels of deontological and utilitarian judgment, which is the focus of the current studies.

It is clear that moral judgment involves at least *some* emotional processing. However, the recent emphasis placed on emotional functioning in moral cognition may tempt some to conclude that deontological judgment is emotional response *and nothing more*. There is good reason to be skeptical of this claim.

2.2. The importance of deontological constraints for deontological intuition: Nichols' affect-backed normative theory

A more moderate claim is made by Nichols (2002), Nichols & Mallon, 2006), who argues that moral cognition depends on an "affect-backed normative theory." The normative theory consists of a set of proscriptive rules that codify moral and immoral behavior. These constraints are "affect-backed" because they are often accompanied by affect. This account attributes an important influence to affect, but argues that other accounts that emphasize emotional reactions (like Greene et al's, 2001) neglect the role of rules in moral judgment.

On this view, rules establish preconditions for actions being viewed as morally wrong. While this may seem tautological, Nichols and Mallon (2006) note that choosing to circumcise one's son qualifies as "personal" by Greene et al.'s standards, and if this were the only determinant of judgment (cf. Greene et al., 2004), this action would be viewed as morally wrong. Nichols and Mallon argue that because our culture does not have a rule proscribing circumcision, this practice is viewed as permissible.

This account implicates three processes: cost-benefit analysis, checking to see whether an action violates a rule, and an emotional reaction. Two kinds of evidence are presented that suggest moral judgment is mediated by affect. First, Nichols (2002) found that conventional violations that elicited affective reactions (e.g. spitting at the table) were judged as less permissible than violations that did not (e.g. playing with your food). Importantly for the hypotheses tested in Study 1, this effect was moderated by individual differences, being more pronounced for participants high in disgust sensitivity.

Second, Nichols and Mallon (2006) developed trolley-like dilemmas of minimized emotional force and found a distinction between judgments of whether the protagonist broke a rule – what they call "weak impermissibility", and judgments that an action was wrong, all things considered – what they call "all-in impermissibility." They found that violations of affect-backed rules were more likely to generate judgments of all-in impermissibility than violations of non-affect-backed rules.

However, Nichols and Mallon (2006) also found that even affect-backed moral rules could be overwhelmed in catastrophe cases. For example, when told billions of people would die from a virus released into the atmosphere unless the fat man is pushed, 68% of participants judged that the act violates a moral rule. However, only 24% judged that the action was morally wrong, all things considered.

Nichols and Mallon's (2006) results suggest that moral judgment is influenced both by whether violations of rules evoke affective reactions and by whether attention is directed to consequences favoring violating a rule. It appears that non-affect backed rules are treated as consequentialists treat all constraints: In cases where the consequences favor a harmful action, infringing a constraint is morally justifiable. In these cases, one might judge that an action violated a moral rule, and that the action is morally right, *all things considered*. In contrast, the operation of *affect-backed rules* is more consistent with a rigid deontology: violating these rules is forbidden except in the most extreme circumstances.

The research reviewed above suggests that judgments elicited by ethical dilemmas are influenced by (a) whether the harm-producing action elicits a strong emotional reaction (Greene et al., 2001), (b) whether consequences favoring the sacrifice are great enough (i.e., many lives to be saved; Nichols & Mallon, 2006) and (c) individual differences in emotional processing (Nichols, 2002). Study 1 examines each of these factors by having participants respond to ethical dilemmas, some of which are modified to elicit a stronger emotional reaction to action, while others are modified to be more like catastrophe cases.

2.3. Does deliberation lead to utilitarianism? Intuitive and deliberative thinking styles

The studies reviewed above suggest that intuition *and* deliberation shape moral judgment. Both frameworks treat deontology-consistent judgment as intuitive: Greene et al.'s (2001, 2004) research, in particular, suggests that sensitivity to rule violations is often more reflexive than reflective. Nichols and Mallon's (2006) implicates affect (triggered by violations) for judging an action impermissible, *all things considered*.

Also, both frameworks motivate hypotheses about when deontological intuitions will be overridden by consequentialist responses. Nichols and Mallon demonstrate that constraints can be overridden in catastrophe cases, and Greene et al. (2001) argue that some utilitarian judgment is produced by deliberatively overriding the affectively pre-potent, deontology-consistent response. Note that the processes implied by Greene et al.'s account are strikingly similar to the line of argumentation developed by some utilitarians. That is, if people were to engage in rational, logical analysis, their moral sentiment would be drawn away from their initial impressions and brought in line with utilitarianism (Hare, 1981; Unger, 1996).

Study 1 tests predicts about automatic and controlled processes by asking whether deliberative thinkers will exhibit preferences more consistent with utilitarianism than intuitive thinkers. Study 1 uses a modified version of Epstein, Pacini, DenesRaj, and Heier's (1996) Rational versus Experiential Inventory (REI) to measure differences in thinking styles. The REI consists of two subscales: the Need for Cognition scale (Cacioppo, Petty, & Kao, 1984), which measures enjoyment of and reliance on deliberation, and the Faith-in-Intuition scale, which measures enjoyment of and reliance on intuition. If deontology-consistent response is driven by emotional activation, one might expect that intuitive thinkers should demonstrate more deontology-consistent preference, while deliberative thinkers should demonstrate utilitarian preference.

2.4. Lay deontology, or unprincipled emotional response? Assessing whether deontology-consistent intuition is predicated on deontological constraints

Considering the popularity of accounts that stress the influence of affective processes (e.g., Greene et al., 2001; Haidt, 2001), it might be tempting to attribute a large share of deontology-consistent judgment to affective reactions that have little to do with moral constraints. However, other theorists have argued that constraints are necessary for deontological judgment.

Mikhail (2007), for example, notes that emotion-based accounts are in need of an appraisal theory – that merely noting that some perceived moral violations are associated with emotional responses misses, among other things, the important first step of interpreting the stimulus for evaluation. He manipulates the causal structure of trolley cases and finds that judgments are sensitive to relatively nuanced distinctions (e.g., doing/allowing harm, treating people as means/ends, intentional harm/harm produced as a side effect of good intentions). Sensitivity to these features is suggestive of the operation of relatively narrowly-directed constraints. And, while they do not offer a computational theory of appraisal, Nichols and Mallon's (2006) affect-backed normative theory postulates a set of rules that establish preconditions for judgments of impermissibility.

Instead of assuming constraints, Study 1 assesses participants' endorsement of deontological principles using the Idealism subscale of Forsyth's (1980) Ethics Position Questionnaire (EPQ). Participants are asked rate agreement with deontological principles (many drawn from Kant, 1966/1785), the majority of which concern actions that harm humans or otherwise violate people's rights (e.g., "One should never psychologically or physically harm another person," and "Risks to another should never be tolerated, irrespective of how small the risks might be"). One straightforward prediction is that people who indicate greater agreement with these principles should demonstrate more deontological preferences when asked to respond to ethical dilemmas. In other words, in the absence of rules proscribing harm, participants posed with the footbridge case should be likely to engage in a simple utilitarian calculus and view a five-lives-for-one trade-off permissible.

2.5. Hypotheses

Preference is expected to be most deontology-consistent for vividly-described dilemmas, and most utilitarian for the catastrophe dilemmas. The vivid descriptions of harmful actions make them seem more egregious, and because attention is directed to the act, I expect participants to view the actions as moral violations. Describing the consequences as especially grave serves to focus participants on the actions' ultimate consequences, thus directing attention away from the violation of a deontological constraint.

Deliberative thinkers are expected to exhibit more utilitarian preference than intuitive thinkers. Greene et al. (2001) suggest an "inhibition hypothesis" – they view some utilitarian judgments as the product of deliberatively overriding the affective intuitions that (they argue) fuel deontology-consistent response. Intuitive thinkers, who "trust their feelings," will not be motivated to override these feelings and will therefore demonstrate more deontology-consistent preference.

The difference between intuitive and deliberative thinkers predicted above is expected to be especially pronounced for responses collected from the affect-laden "vivid" condition. Because the emotional signal associated with the harmful action should be stronger in this condition, deliberative thinkers will have to work even harder to override the affectively pre-potent, deontological response.

Participants who indicate greater agreement with deontological constraints are expected to exhibit less utilitarian preference. Rather than explaining deontology-consistent preference entirely in terms of a lower-order emotional reaction, deontological preference may be principled. If participants' decontextualized normative perceptions (idealism) predict revealed preference, then moral preference may be shaped, in part, by adherence to moral rules.

2.6. Methods

2.6.1. Participants

Seventy-one undergraduates (45 females and 26 males) participated. They were tested in a small group setting (typically one to four participants per session). Another unrelated study was also run during these sessions. Eight participants did not supply data for one of the predictor variables (noted below) because of time constraints on the experimental session introduced by the duration of the unrelated study. All received partial course credit.

2.6.2. Materials and design

Participants in this study completed two assessments of individual differences and indicated preferences for 14 ethical dilemmas. First, participants responded to a randomized ordering of the Idealism subscale of Forsyth's (1980) Ethics Position Questionnaire (see Appendix A). Second, participants read and gave judgments for 14 ethical dilemmas similar to the one below (Greene et al., 2001; see also Alda et al., 1983, and see Appendix B for a full list of the 14 scenarios). The "vividness" manipulation appears in brackets; the "catastrophe" manipulation appears in parentheses:

> Enemy soldiers have taken over your village and will kill all remaining civilians. You and five (19) others are hiding in the cellar of a large house. Soldiers have come to search the house for valuables. A baby in your group begins to cry. So, you cover her mouth, but she cannot breathe. If you remove your hand, the baby can breathe, but her crying will summon the soldiers who will kill everyone in the cellar. [The baby is writhing around violently in your arms. With all her might, she desperately struggles to breathe].

In this situation, would you smother the baby?

Participants indicated their responses by clicking on one of the boxes. Responses were recoded from zero to one (coded 0.00, 0.25, 0.75, 1.00) so that higher numbers indicated more utilitarian preferences. (The results do not crucially depend on how these responses are scored. Analyses using different scoring methods – using just the

endpoints of the scale or the treating scale as a dichotomous measure – yield the same patterns as those reported below.)

Each participant participated in all three conditions (Standard, Catastrophe, Vivid) and responded to all 14 scenarios, but never responded to the same scenario twice (i.e., never saw more than one version of a scenario). For each participant, roughly one-third of the stimulus scenarios were from each of the three experimental conditions.

First, the order of presentation of the 14 dilemmas was randomized for each participant. Second, the assignment of conditions to trials was randomized for each participant, such that on every third trial, a participant responded to a standard/catastrophe/vivid dilemma. The permutation of conditions for each block of three trials – whether Ps saw a standard dilemma, followed by a catastrophe, followed by a vivid dilemma, versus one of the other five possible permutations, was randomly determined for each participant. This design ensured that each participant participated in each condition, that each participant responded to each of the 14 items, and that no participant saw more than one version of an item.

After responding to the ethical dilemmas, participants responded to a randomized ordering of a modified, 20-item version of the REI (Epstein et al., 1996; Pacini & Epstein, 1999). For each item, participants rated their level of agreement with statements like "I prefer to do something that challenges my thinking abilities rather than something that requires little thought" (where greater agreement suggests greater reliance on analytic-deliberative thought) and "Using my gut feelings usually works well for me in figuring out problems in my life" (where greater agreement suggests greater reliance on intuition; see Appendix A). Eight participants did not provide REI responses because of the time constraints placed on the experimental sessions in which they participated.

2.7. Results and discussion

2.7.1. Experimental results: Influence of vividness and catastrophe manipulations

Study 1 found that the experimental manipulations produced effects in the predicted direction. For the within-participants contrasts, I computed for each participant the average preference for the items they viewed in each of the experimental conditions. Overall, participants indicated less utilitarian preferences for the vivid condition items they viewed ($M = 0.37$, $SD = 0.19$) than for the standard condition items ($M = 0.45$, $SD = 0.19$, paired-$t(1,70) = -2.86$, $p < .01$). Forty-six of the 71 participants showed this effect. Also as predicted, participants indicated more utilitarian preferences for the catastrophe items they viewed ($M = 0.54$, $SD = 0.20$, paired-$t(1,70) = 3.82$, $p < .001$). Forty-eight participants showed this effect. As Table 1 shows, the within-subjects contrast yielded a large effect $F(2,140) = 23.19$, $p < .001$, $\eta_p^2 = .25$), consistent with expectations.

Most of the items exhibited similar effects. Recall that no participant saw more than one version of a scenario. So, the item comparisons shown in Table 1 are between-subjects. Even though the within-participants row of Table 1 suggests a very large effect, the between-subjects contrast, controlling for the effects of the experimental manipulations, yields an F-value$(1,70)$ of 663.69 and an effect size (η_p^2) of

Table 1
Study 1: Effects of condition (vivid, standard, and catastrophe) on utilitarianism in morally-motivated preference

Contrast	Vivid	Standard	Catastrophe	*F*-value (η_p^2)
Footbridge	0.16	0.10	0.27	2.99† (.08)
Submarine	0.50	0.49	0.68	2.25 (.06)
Hostages	0.19	0.53	0.50	9.45** (.22)
Surgery	0.06	0.13	0.18	1.63 (.05)
Trespassers	0.26	0.28	0.40	1.34 (.04)
Liferaft	0.31	0.56	0.64	5.90** (.15)
Spelunkers	0.49	0.55	0.53	0.25 (.01)
Derailment	0.44	0.52	0.63	1.85 (.05)
Baby	0.49	0.53	0.59	0.79 (.02)
Bystander	0.65	0.73	0.77	1.09 (.03)
Plane Crash	0.18	0.07	0.30	4.60** (.12)
Fumes	0.63	0.65	0.80	1.94 (.05)
Prisoners of War	0.49	0.49	0.69	3.41* (.09)
Soldiers	0.43	0.58	0.65	2.17 (.06)
Within-Ps	0.37	0.45	0.54	23.19** (.25)
Between-Ps				663.69**(.90)

* $p < .05$.
** $p < .01$.
† $p < .10$.

.90. Clearly, there was wide variation in preferences even within this (potentially restrictive) sample of undergraduates.

2.7.2. *Influence of individual differences*

All of the effects reported in this section appear in Table 2. The results for the average preference exhibited across all 14 responses were as predicted: the more a participant relied on intuition than deliberation, and the greater his or her endorsement of deontological principles, the more deontology-consistent were his or her preferences (*r*s .39 and −.32, both *p*s < .01).[1] Also consistent with expectations, the effect of thinking styles on preference was most pronounced for responses collected from the vivid condition ($r = .47$, $p < .01$). These results offer some support for an emotions-based account of morally-motivated preference something like Greene et al's (2001) inhibition hypothesis. That is, some utilitarian responses (especially those where "the heart" tugs in one direction and "reason" in the other) are produced by expending cognitive effort to inhibit the affectively pre-potent deontological response. In addition, the relationship between idealism and preference also suggests a role for deontological constraints.

The results of Study 1 suggest that there may be more than one type of morally-motivated decision maker, and that moral preference is flexible. The results

[1] Men were scored as more deliberative than women; men provided higher REI scores ($M = 0.98$, $SD = 1.93$) than women ($M = -0.42$, $SD = 2.21$, $t(1, 61) = 2.51$, $p < .05$, $\eta_p^2 = .09$). I had no theoretical basis for predicting this effect, and I have no explanation for it. There were no appreciable gender differences for any of the other variables measured in any of the current studies.

Table 2
Study 1: Correlations between individual differences indices and utilitarianism morally-motivated preference across experimental conditions (vivid, standard, and catastrophe)

	Experimental condition			
	Average	Vivid	Standard	Catastrophe
Thinking styles	.39**	.47**	.20	.25*
Idealism	−.32**	−.33**	−.26*	−.19

* $p < .05$.
** $p < .01$.

suggest that participants who affirm deontological principles, and that intuitive thinkers, exhibit preferences more consistent with deontology. Results also suggest that focusing participants' attention on violations of constraints – by exacerbating negative emotional reactions to these actions – promotes deontology-consistent choice. Study 1 also found that participants were willing to sacrifice human lives (in the catastrophe condition) if enough lives can be saved by doing so.

2.7.3. Why the generalizability of Study 1's results is limited

As noted earlier, researchers studying responses elicited by dilemmas sometimes exhibit what Tetlock et al. (1996) refer to as the anti-context (people as random replicates) and anti-content (items as random replicates) biases. Study 1 partially avoids the anti-context bias by accounting for within-sample variance: participants who are more likely to override emotional influence appear more utilitarian, and those who endorse moral rules demonstrate preferences more consistent with the operation of those moral rules. However, since Study 1 uses only trolley-like dilemmas, the generalizability of these results is limited.

Studies 2 and 3 use other types of scenarios. Recall Hare's (1981) wry comment about "playing trains," that is, the questionable relevance of intuitions and preferences elicited by those abstract – and, he argues – highly artificial stimuli. His contention was that intuitions generated for more natural stimuli are more likely to be utilitarian. At a minimum, researchers should be wary about generalizing the set of processing principles implicated in response to dilemmas to other choice contexts. Studies 2 and 3 are, in this way, conceptual replication and generalization studies, and use as contexts the types of policy-level decisions for which, utilitarians argue, their theory is especially well-suited (Goodin, 1993).

3. Protected values as affect-backed constraints: Moral judgment in separate (Study 2) and joint evaluation (Study 3)

Studies 2 and 3 examine whether and when we should expect moral values to engender nonconsequentialist decision principles by investigating PV-driven response in two evaluative contexts: absolute and comparative judgment.

Research on protected values suggests that when contemplating the exchange of a cherished resource (for which people have a PV), people may reason differently (invoking moral rules) than when reasoning about resources not tied to one's moral values (Baron & Spranca, 1997). In some cases, people react to proposed tradeoffs with moral outrage (Tetlock, 2002), and disavow consideration of costs and benefits. Baron and Spranca (1997) describe PVs as a subset of deontological rules that are tied to affect – rules that concern actions, like "do no harm", but not the consequences of those actions.

For present purposes, the PV framework offers a way to select the domains for which we might expect deontology-consistent judgment. To assess PVs, participants are presented with statements concerning the acceptability of tradeoffs, as below:

Causing the extinction of fish species.

(a) I do not object to this.
(b) This is acceptable if it leads to some sort of benefits (money or something else) that are great enough.
(c) This is not acceptable no matter how great the benefits.

People who endorse "c" are counted as having a PV for that resource (Ritov & Baron, 1999). People with PVs appear more sensitive to the distinction between doing and allowing harm and less sensitive to the consequences of their choices than people without PVs, resulting in what Baron et al. call an "omission bias" (Ritov & Baron, 1999; Spranca, Minsk, & Baron, 1991).

So, how rigidly nonconsequentialist are preferences for domains governed by PVs? Do people with PVs care less about the consequences than people without PVs? Perhaps. First, the measurement suggests a lack of concern with consequences – "no matter how great the benefits". Second, omission bias is consistent with commitment to moral prohibitions. That people with PVs prefer omission over a harmful action in contexts where the consequences favor action suggests PV-driven preference is consistent with a rigid deontology (Baron & Spranca, 1997).

Are people absolutist deontologists for domains governed by PVs? Perhaps not. People who care more about *not harming* the resource might also care *more* about the consequences of an act. Based on this logic, Bartels and Medin (2007) examined the PV-driven preference using two procedures. Using a procedure that focused attention on whether a harmful action should be taken to maximize net benefits, PVs were associated with nonconseqentialist response. Using a procedure that highlights the net costs averted by such actions, the trend reverses – people with PVs appeared *more* consequentialist than people without PVs. These results make sense if people with PVs for a resource care more about not harming the resource *and* about the consequences of actions in the domain.

Study 1 suggests moral cognition is influenced by rules and whether attention is directed to a harmful act or to its consequences. Studies 2 and 3 expand on these findings, using accounts of people's responses to dilemmas to investigate the context-sensitive role of PVs in moral judgment. In particular, PVs appear to share important properties with affect-backed constraints (Nichols & Mallon,

2006). Violating PVs elicits emotional response – tradeoffs of PVs can elicit extreme anger (Baron & Spranca, 1997). But just as even affect-backed rules can be overwhelmed if attention is directed to catastrophic consequences favoring infringing them, Bartels and Medin (2007) showed that people's willingness to accept tradeoffs of PVs depends on where attention is focused.

3.1. Protected values as affect-backed constraints

Studies 2 and 3 examine whether PVs operate like affect-backed constraints in moral judgment. In these studies, participants rate their approval of government administrators' decisions. In Study 2, they also judge whether the administrators' decisions violate a moral rule (judgments of "weak impermissibility").

For each scenario, each of two administrators chooses either to knowingly *do harm* to a resource to mitigate even greater harm or to *allow* the harm to happen. For example, participants read that 20 species of fish upstream would be made extinct unless a dam is opened, but that opening the dam will kill some species downstream (see Appendix C). One administrator desires to do no harm (and thus not to act), and so 100% of the anticipated harm results. For example, *"Paul does not want to kill any of the fish species downstream. So, the dam is not opened. The 20 species upstream die."* I refer to these choices as "omission."

The other administrator first calculates that by intervening, he or she will kill 80% of the resources, and based on this analysis, he or she chooses to intervene. For example, *"David wants to save the fish species upstream. He first calculates that opening the dam will kill 16 species downstream. Knowing that doing so will kill many fish, he chooses to open the dam."* I refer to these choices as "action." Note that the administrator's intent to save the species is given so that participants will not mistakenly assume malicious intent. Study 2 presents the decisions in separate evaluation: on a given trial, participants evaluate either the omission or the action, but not both. In Study 3, both are evaluated on a single trial.

Recall that in Nichols and Mallon (2006), weakly impermissible actions – those that violated rules but failed to elicit an emotional response – were less likely to be judged wrong, all things considered ("all-in impermissibility") than violations of affect-backed rules. By relating judgments of rule violation and something akin to all-in impermissibility – (dis)approval of a government administrator's decision – Study 2 offers a test of whether PVs operate like affect-backed constraints. The relationship between these judgments is predicted to be stronger for domains governed by PVs.

This test relates to Feldman and Lynch's (1988) approach to "self-generated validity" – an explanation of how judgments elicited by one probe can influence judgments elicited by a subsequent probe. They argue that an earlier response will be used as the basis for a subsequent judgment if the former is accessible and perceived to be more diagnostic than other accessible inputs. In Study 2, when one's judgment of whether a rule has been violated is diagnostic for whether or not one approves of a decision, one need not weigh other considerations. Especially considering how constraints preclude other considerations – like the goods to be promoted

by violating the rule – we might expect people with affect-backed constraints (PVs) to treat judgments of rule violation and decision approval as the same judgment. In normative deontological theory, these judgments collapse into one: actions that violate deontological constraints are wrong, *simpliciter*.

3.2. Joint versus separate evaluation preference reversals

Study 1 showed that utilitarian preference can be promoted or diminished by task-induced attentional effects. Studies 2 and 3 examine whether moral value-driven focus on rules and consequences is subject to a different set of task constraints. Study 2 asks participants to evaluate decisions in separate evaluation. Study 3 presents decisions in joint evaluation, inviting a comparison between options before rendering judgment.

Previous research demonstrates that attributes that appeal to one's intuitive sensibilities, and attributes that are otherwise easy to evaluate, drive preference in separate evaluation (where a number of otherwise useful comparisons are not made available), whereas attributes that appeal to "colder", more logical sensibilities drive preference in joint evaluation (Bazerman, Moore, Tenbrunsel, Wade-Benzoni, & Blount, 1999). For example, Hsee and Leclerc (1998) asked three groups to assign buying prices to an ice cream product. One group was asked to evaluate a 7-oz serving presented in a 5-oz cup, a second group evaluated an 8-oz serving in a 10-oz cup, and a third assigned buying prices to both.

Participants in the first condition were willing to pay more for the 7-oz serving than participants in the second condition were willing to pay for the 8-oz serving. In separate evaluation, participants incorporated feelings about the cup being overfilled/underfilled into their evaluative judgment. In the joint evaluation condition, where participants are able to select the most important attribute for setting a buying price, buying prices were greater for the 8-oz serving than the 7-oz serving.

One interpretation of this preference reversal is that people discard intuitive reactions when comparison highlights a more *sensible* basis for evaluation, and this is one basis of argumentation in normative ethics. Deontology tries to systematize our moral intuitions in the context of an over-arching theory, accommodating the "outrage" we feel when contemplating some forms of sacrifice in the name of promoting utilitarian considerations (e.g., Donagan, 1977). In other words, good normative theory should accommodate moral intuitions – they are honored as meaningful data in the development of normative ethical theory. Some utilitarians suggest we should discard intuitions and instead rely on very basic moral values, like promoting total well-being. Unger (1996) does this by inviting comparison between scenarios that illustrate to the reader that her intuitive judgments for one case are premised on a factor that is clearly irrelevant in a structurally similar case.

Joint evaluation offers a strong test of the normative status people give to deontology-relevant and utilitarian-relevant attributes (Bazerman & Messick, 1998). Studies 2 and 3 assess whether disapproval for violations of PVs is overwhelmed when the consequences favoring violating a PV are made salient in joint evaluation. That is, Study 3 tests whether people invest the doing/allowing harm distinction with

enough normative significance to outweigh utilitarian considerations for decisions involving PVs.

3.3. Intuitive and deliberative thinking styles

Like Study 1, Study 2 measures differences in thinking styles using the Rational versus Experiential Inventory. Individual differences in the propensity to allow intuitive reactions to influence judgment are expected to influence rated approval in Study 2. In separate evaluation – a context that promotes intuitive judgment – I expect PV-driven judgment for participants who "trust their feelings," to be more focused on the impermissibility of knowingly doing harm than on utilitarian considerations. In contrast, deliberative thinkers might be more likely to ignore or override intuitions generated by violations of their affect-backed PVs and, thus, render more consequentialist judgments (regardless of context).

3.4. Hypotheses

For separate evaluation judgments, I predicted that actions would be evaluated more negatively for domains governed by PVs than for domains not governed by PVs. For these items, the administrator is described as knowingly doing harm, violating a PV. Violations of PVs evoke anger (Baron & Spranca, 1997). Moreover, the administrator reaches a decision on the basis of cost-benefit analysis. Cost-benefit reasoning about moralized goods can itself elicit moral outrage (Tetlock et al., 2000; Viscusi, 2000). Because attributes that appeal to intuitive faculties constrain judgment in separate evaluation, I predicted that a negative affective reaction would contribute to judgments of disapproval.

I expected that the effect predicted above would be moderated by individual differences. Deliberative thinkers will be more likely to over-ride the pre-potent response to render strong disapproval of PV violations, and thus show a smaller effect. In contrast, the effect should be more pronounced for those who report greater reliance on intuition, because these participants will be less likely to minimize the influence of emotional reaction on judgment.

The relationship between rule violation and approval is expected to be stronger for domains governed by PVs than for other domains. When participants perceive that a PV has been violated, they should more strongly disapprove of the decision than when they perceive no violation or perceive that some other (non-affect-backed) moral rule has been violated. That is, violations of PVs (as affect-backed constraints) should be sufficient for strong disapproval. For other domains, participants may be willing to support a decision that violates a moral rule if the benefits brought about are great enough.

In joint evaluation, actions are expected to be met with more approval than omissions. Consistent with the predictions of Bazerman and Messick (1998), I expected the utilitarian attributes – 80% are lost with the action; all 100% are lost with the omission – would be large enough to sway even participants whose judgments might be consistent with a rigid deontology in other contexts.

3.5. Methods

3.5.1. Participants

Forty-eight undergraduates (25 women and 23 men) participated in Study 2 for partial course credit. They were tested in a small group setting (typically one to four participants per session). Those participants who wrote their contact information on a sign-up sheet for Study 3 were contacted about participating in the second round of data collection, conducted 61 to 71 days later. Thirty-two of the original 48 (18 women and 14 men) participated in Study 3 in exchange for $5 compensation. (Those who returned to participate in Study 3 did not differ from those who did not return on any of the predictor or criterion variables – all $ts < 1$).

3.5.2. Materials and design

In each session, participants completed one of three packets that differed only in the pseudo-randomized ordering of items within each type (PV items, Rational-Experiential Inventory items [REI completed once, in Study 2], judgment scenarios). First, participants responded to 30 PV items – seven that corresponded to the judgment scenarios intermixed with 23 unrelated PV items. Then, participants responded to a modified, 20-item version of the REI (see Appendix A). Finally, participants evaluated two governmental administrators' choices for seven problems. The 14 judgment scenarios crossed two types of decisions (omission, action) with seven problems (birds, children, dolphins, fish, jobs, poor, trees). The two versions of the "Children" problem appear below:

> (Name) is considering a vaccination program. Epidemiologists estimate that vaccinating 600 children will prevent them from dying from an epidemic of a new infectious disease. The vaccine itself will kill some number of children because it sometimes causes the same disease. Because this disease progresses rapidly, a decision must be made quickly, and the government's options are severely constrained.

Julie does not want to kill any of the children with the vaccine. So, the vaccine is not administered. The 600 children die.

Rich wants to save the children from the disease. He first calculates that administering the vaccine will kill 480 children. Knowing that doing so will kill many children, he chooses to vaccinate the children.

In Study 2, after reading about the administrator's decision, participants were asked to assess whether or not the administrator broke a moral rule. The item read, *"By (not) administering the vaccine, (Julie) Rich broke a moral rule."* Participants indicated agreement on a -3 (Strongly Disagree) to +3 (Strongly Agree) scale. Then, in Study 2, participants were asked, *"How do you feel about (Julie's) Rich's decision?"* Participants indicated approval or disapproval by circling a partitioning mark on a scale ranging from "Strongly Disapprove" (coded as 1 for the analyses that follow) to "Strongly Approve" (coded as 8).

In Study 3, participants read about both decisions before being asked to evaluate each. Before the first decision presented on a page, participants read *"Suppose this problem has been assigned to (Name)"*, and then read *"Now suppose, instead, that this problem has been assigned to (Name)"* between the first and second decision. Participants then rated their approval or disapproval of each decision as they did in Study 2.

For each of the three packets, the order of problems (i.e., birds, children, etc.) was randomized. For Study 2, the assignment of action type to the problems was randomized so that on every other trial, participants evaluated an omission (or an action).

3.6. Results

To assess whether moral judgment differs according to whether the domain is governed by protected values, I first report analyses of within-subjects effects where I separated the items for which each participant endorsed a PV (referred to as "PV") from the items for which he or she did not (referred to as "No PV"). I also report analyses conducted for each item.

3.6.1. Rule violations and approval (Study 2)

For each participant, I computed correlations between judgments of moral rule violation (i.e., "weak impermissibility") and approval ratings across PV and No PV items. I predicted that violations of moral rules would elicit strong disapproval ratings for domains governed by PVs. I found that, in general, when participants perceived rule violations, they disapproved of the administrators' decisions. This relationship held for No PV items ($M = -.66$), and as predicted, was stronger for PV items ($M = -.84$; Wilcoxon signed ranks test for related samples $Z = -2.42$, $p < .05$; both sets of within-Ps correlations reliably negative by signed ranks tests -360.5, -388, $ps < .001$). This finding is consistent with the notion that PVs function like affect-backed constraints in influencing moral judgment.

As a reviewer pointed out, the overall relationship between judgments of rule violation and disapproval is also consistent with an "emotional reaction hypothesis" whereby participants indicate their negative emotional response by registering negative evaluative judgments on both probes. After all, both Haidt, Koller, and Dias (1993) and Cushman et al. (2006) report evidence that people sometimes appeal to moral principles that cannot explain their behavior. Although this alternative explanation cannot be ruled out by these data, there may be reasons to favor taking people's judgment of rule violation in Study 2 at face value over the emotional reaction hypothesis. Previous studies demonstrate that participants behave in a manner consistent with the idea that harmful actions, but not necessarily omissions, violate moral rules (Baron & Spranca, 1997). As noted, Cushman et al. (2006) find that participants sometimes invoke principles that cannot explain their behavior, but notably, when participants in their studies were asked to justify why actions and omissions with equivalent consequences elicited discrepant judgments from them, participants *were* able to reliably gener-

ate the action/omission distinction (but not other principles) as the normatively relevant factor.

3.6.2. Approval ratings – Separate evaluation (Study 2)

For each participant, I calculated four averages: one each for the participant's approval ratings for omissions/actions on items for which he or she endorsed/did not endorse a PV. Recall that in the current design, actions resulted in better consequences (80% loss) than omissions (100% loss). In Study 2, this comparison was unavailable to participants, leaving emotional reactions to drive (dis)approval.

I expected that for domains governed by PVs, decisions to *knowingly do harm* on the basis of explicit cost-benefit reasoning would be considered offensive (violations of PVs – rules like "do no harm" – elicit anger), eliciting greater disapproval from participants than similar choices made for other domains. I also predicted that this tendency would be more pronounced for intuitive thinkers, who might be less likely to override their emotional reaction than deliberative thinkers. The results of a 2 (Decision: Action/Omission) × 2 (Domain: No PV/PV) repeated-measures ANOVA reveals effects of each factor (F's$(1,43) = 19.98$ and 19.87, p's $< .001$, η_p^2's $= .32$) and a reliable interaction ($F(1,43) = 12.58$, $p < .001$, $\eta_p^2 = .23$).[2]

Fig. 1 depicts the pattern of results obtained for Studies 2 and 3, presenting the average of participants' average approval ratings for acts and omissions by the presence and absence of PVs. As predicted, actions were evaluated more favorably for No PV domains than for PV domains (M's $= 5.39$ vs. 4.30), as is evidenced by the negative slopes apparent for the two solid lines in the left half of Fig. 1. Also as expected, the correlation between this difference score and participants' REI scores was strongly negative ($r(43) = -.54$, $p < .001$), indicating that deliberative thinkers showed a smaller effect. Fig. 1 shows that the negative slope supporting the expectation that PV-violating actions should be met with less approval is more pronounced for those scored as "intuitive" thinkers (bottom solid line in the left half of Fig. 1).

Approval ratings for omissions did not appreciably differ across these domains (M's $= 4.17$ vs. 4.07; depicted by the solid lines in the right half of Fig. 1), nor did difference scores for omissions relate to thinking styles ($r(43) = .15$, $p > .10$).

Table 3 presents approval ratings for each item as a function of decision and the presence or absence of PVs. As expected, actions are met with more approval by participants without PVs than participants with PVs for every item, though the contrast is only reliable for four of the seven items used (see column "Hyp SE" for "separate evaluation"). The results of ANOVAs run for each item, using Decision (Action vs. Omission) as a repeated-measures factor and the presence or absence of a PV as a between-Ps factor are also presented in Table 3.

[2] Four of the 48 Ps in Study 2 endorsed zero PVs. The test of rule violation and approval, as well as the tests run on Ps averages exclude these participants, but their responses are included in the item analyses summarized in Table 3. Similarly, two of the 32 Ps in Study 3 endorsed all seven PVs, and another three Ps endorsed zero. These five Ps' responses are counted in the item analyses summarized in Table 4, but not for the analyses run on Ps' averages.

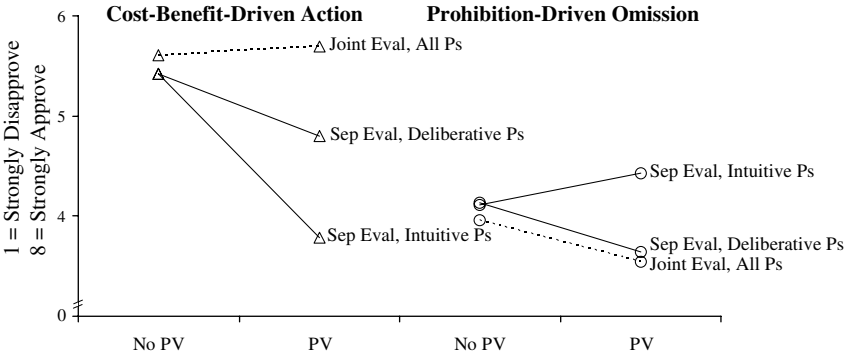

Fig. 1. *Studies 2 and 3:* Approval ratings as a function of Decision (Action vs. Omission) and presence or absence of PVs. Results are presented separately as a function of Thinking Styles (Intuitive vs. Deliberative) and Separate Evaluation (Study 2) versus Joint Evaluation (Study 3).

A reliable preference for action over omission is present in five of the seven items used. Also, for four of the seven items, the effect of PVs on judgment depends on which type of action is being evaluated; for these items, the relatively larger differences in (dis)approval for actions (between people with and without PVs) appears to produce these interaction effects.[3]

3.6.3. Approval ratings – Joint evaluation (Study 3)

Actions were evaluated more favorably than omissions in joint evaluation, regardless of the presence/absence of PVs and differences in thinking styles. The results of a 2 (Decision: Action/Omission) × 2 (Domain: No PV/PV) repeated-measures ANOVA reveals a large effect for Decision ($F(1,26) = 44.12$, $p < .001$, $\eta_p^2 = .63$) consistent with expectations, a marginal effect of the presence/absence of PVs ($F(1,26) = 3.36$, $p = .08$, $\eta_p^2 = .11$) and no reliable interaction ($F(1,26) = 2.26$, $p > .10$, $\eta_p^2 = .08$). The marginal effect of PVs in the ANOVA is influenced by an unanticipated difference in approval ratings for omissions across No PV and PV domains.

[3] Participants in Study 2, in particular, responded to many items (30 PV items, 20 REI items, and two blocks of 7 judgment scenarios). I thank a reviewer for directing me to this concern. It is difficult to assess the effects of the large battery of items on participants' responses (without a referent data set), but one might guess, for example, that fatigue may have dulled the emotional response triggered by perceived violations, which may have led to more utilitarian responding over the course of the judgment scenarios. Or perhaps information overload may have increased or decreased response variance over the block of 14 scenarios. However, no reliable order effects on the dependent measures are present in these data. More generally, researchers should devote more attention to the influence of exogenous factors on moral judgment, because the contexts in which responses are collected vary widely from study to study. For example, some trolley studies have presented to participants seated in a quiet classroom as few as one scenario for evaluation (e.g., Nichols & Mallon, 2006), while others have posed 32 scenarios to participants over the internet (Cushman et al., 2006), while others have posed as many as 60 scenarios to participants in an fMRI scanner (Greene et al., 2001).

Table 3
Study 2: Proportion of Ps endorsing PVs, Approval ratings for each item as a function of Decision (Act vs. Omission), and presence or absence of PVs and Effects of each factor (Decision and PV) and their interaction

Item	% PV	Action		Hyp SE	Omission		Decision	PV	Interaction
		No PV	PV	*t*-value (η_p^2)	No PV	PV	*F*-value (η_p^2)	*F*-value (η_p^2)	*F*-value (η_p^2)
Birds	0.31	5.6	3.7	4.12**(.27)	4.4	3.9	2.26 (.05)	15.99**(.26)	3.56† (.07)
Children	0.52	4.3	2.2	4.20**(.28)	3.5	3.6	<1	7.74**(.14)	7.07*(.13)
Dolphins	0.56	5.7	5.3	<1	4.7	4.4	10.51**(.19)	1.32 (.03)	<1
Fish	0.25	5.9	5.0	2.27* (.10)	4.4	3.8	24.41**(.35)	5.34* (.10)	<1
Jobs	0.19	5.5	4.8	1.31 (.04)	4.4	3.2	13.67**(.23)	6.26* (.12)	<1
The poor	0.44	5.4	4.1	2.97**(.16)	3.7	3.9	7.77**(.15)	3.82† (.08)	4.96*(.10)
Trees	0.40	5.5	4.9	1.41 (.04)	4.0	4.4	17.28**(.27)	<1	3.32† (.07)

* $p < .05$.
** $p < .01$.
† $p < .10$.

Omissions received higher approval ratings for No PV domains ($M = 3.96$, $SD = 1.00$) than for PV domains ($M = 3.55$, $SD = 1.03$, paired-t $(1, 26) = 2.09$, $p < .05$). This effect, though unexpected, is consistent with Bartels and Medin's (2007) finding that in a procedure that highlighted net costs and benefits, people endorsing PVs appeared *more sensitive* to utilitarian considerations than people without PVs. Study 3's joint evaluation context allows for comparisons of both deontology-relevant and utilitarian considerations. Given this more enriched evaluative context, attention given to the contrast in consequences appears to overwhelm constraints against doing harm that participants would otherwise treat as having normative significance. Table 4 presents approval ratings for each item as a function of decision and the presence or absence of PVs and ANOVAs run for each item, using Decision (Action vs. Omission) as a repeated-measures factor and the presence

Table 4
Study 3: Proportion of Ps endorsing PVs, Approval ratings for each item as a function of Decision (Act vs. Omission), and presence or absence of PVs and Effects of each factor (Decision and PV) and their interaction

Item	% PV	Action		Omission		Decision	PV	Interaction
		No PV	PV	No PV	PV	*F*-value (η_p^2)	*F*-value (η_p^2)	*F*-value (η_p^2)
Birds	0.38	6.0	5.6	4.0	3.6	30.00** (.50)	1.78 (.06)	<1
Children	0.56	6.1	5.2	2.9	3.6	15.32** (.34)	<1	1.85 (.06)
Dolphins	0.56	6.1	5.5	3.7	3.4	33.53** (.53)	2.37 (.07)	<1
Fish	0.25	5.3	6.0	4.2	3.6	14.12** (.32)	<1	2.15 (.07)
Jobs	0.19	5.6	4.8	4.1	3.8	9.80** (.25)	1.95 (.06)	<1
The poor	0.44	5.8	5.2	3.7	3.4	27.46** (.48)	1.77 (.06)	<1
Trees	0.34	5.5	5.9	4.0	3.2	47.00** (.61)	<1	4.16† (.12)

** $p < .01$.
† $p < .10$.

or absence of a PV as a between-Ps factor. As predicted, for each of the seven items, cost-benefit-influenced actions are strongly preferred to omissions when action and omission are jointly evaluated.

3.7. Discussion

These studies were motivated by the idea that a better understanding of moral judgment can be achieved through more thorough scrutiny of the processes that moral values motivate. Results suggest that the processes implicated in responses to ethical dilemmas (as in Study 1) also generalize to PV-driven judgment rendered for public-policy-level decisions. That is, judgments are influenced by whether rule-violations evoke strong affective reactions, by whether attention is directed to utilitarian considerations, and by individual differences in propensity to incorporate emotional reactions in judgment.

Previous theory suggested PVs motivate rigid, nonconsequentialist judgment and choice. By demonstrating the context-sensitivity of PV-motivated judgment, the present findings qualify previous conclusions, suggesting a more flexible PV-driven judge. Moreover, Studies 2 and 3 account for some of this flexibility by suggesting PVs operate as constituents of an affect-backed normative theory, offering some theoretical synthesis across unnecessarily disparate literatures.

4. General discussion

The studies reported here test a set of predictions that relate moral judgment and preference to underlying cognitive process. The results demonstrate the interactive influences of (a) differences in the judgment-elicitation context, (b) the presence of strong moral attitudes or deontological constraints, and (c) reliable individual differences in tendency to engage in intuition and deliberation.

The current approach places a greater emphasis on the flexibility of moral cognition than some other approaches have. For example, some descriptive frameworks of moral cognition stress its affective underpinnings (Haidt, 2001) and/or the reflexive, emotional nature of deontology-consistent response (Greene et al., 2001). While emotion plays a crucial role in moral cognition, the current studies suggest that deliberation and moral rules play a role, too. *Some* moral judgment is intuition-based, and the current studies show that participants who "trust their feelings" better resemble these reflexive moral judges. Participants more prone to deliberation respond differently.

Further, the current studies show that participants with and without moral rules (PVs) respond differently. Previous views of how PVs influence preference suggested a kind of rigidity in moral cognition: that protected values engendered rigid deontological decision strategies (Baron & Spranca, 1997; Ritov & Baron, 1999). The current studies suggest, instead, that PV-driven responding is a function of the

information highlighted by the situation (e.g., deontological versus utilitarian attributes) and by individual differences in thinking styles.

The approach of these studies is to treat the context-sensitivity of judgment as diagnostic of psychological process that compete to promote or inhibit deontological and utilitarian response. Study 1 shows that participants who affirm deontological principles in the abstract and participants who rely more on intuition than deliberation have preferences more consistent with deontology. Results also suggest that focusing participants' attention on actions that violate moral rules promotes deontological preference, while focusing on the consequences favoring violating the rules promotes utilitarian preference. The results of Bartels and Medin (2007) reviewed earlier, however, shows that this latter effect is seen only in those participants who ascribe moral relevance to the domain under consideration, as assessed by the endorsement of a deontological rule for the domain.

Finally, Studies 2 and 3 demonstrate that the adherence to rules evident in the work on omission bias (Ritov & Baron, 1999) and studies using dilemmas (e.g., Study 1) can be overwhelmed by the comparisons made available in an evaluative context. Participants appeared more utilitarian when asked to compare constraint-violating actions that produce better outcomes to constraint-affirming actions that produce worse outcomes prior to rendering judgment.

4.1. Revisiting the models of judgment elicited by ethical dilemmas

Much debate within psychology over the past few years, and within philosophy for the last few centuries, focuses on the role of emotion in moral cognition. It is clear that emotions play an important role in moral judgment, but it also appears that emotions cannot do *all* the work. First, as Mikhail (2007) and others have pointed out, accounts that attribute deontological judgment to an emotional reaction (e.g., Greene et al., 2001) are missing an appraisal theory. Without first addressing the question of *what about the stimulus* is upsetting, one can provide an incomplete sketch of moral judgment at best.

Second, as Nichols and Mallon (2006) argue, these accounts have missed the importance of moral rules. If we do not have a moral rule forbidding some action, the action is not treated as a candidate for being judged morally wrong. This even goes for actions that are disgusting in some respects, and harmful in an immediate and "personal" (by Greene et al's., 2001; standards) sense (e.g., circumcision). The studies reported here suggest that a *strictly* emotion-based account would be insufficient (a point stressed in Greene et al., 2004). Such an account might explain *Humean* morality (see Hume, 1969/1739), but it could not explain human morality.

It would seem that combining rule-based and emotional process accounts, as these studies do, might be a good approach. The recognition of a rule violation might be a good first ingredient for the development of a working appraisal theory, for example.

Studies 1 and 2 tested for and found an influence of deontological constraints on moral judgment. Study 1 found that people who affirmed deontological constraints were more likely to exhibit deontological preference. Study 2 found that

endorsement of deontological constraints predicted disapproval for constraint-violating actions in a context where the emotional signal generated by these violations was *not* expected to be overwhelmed by utilitarian considerations (i.e., separate evaluation). One might consider these results trivial – people who indicate they are less willing to make moral compromises (by endorsing a constraint) are less willing to do so when presented with dilemmas. However, it should be noted that participants expressed their endorsement of constraints by responding to an abstract, relatively context-free statement (at least in comparison to the dilemmas in which they might be invoked), and that afterwards, participants performed a "filler" task, responding to the (substantively unrelated) 20-item REI before advancing to the judgment scenarios. One might have expected moral judgment to bear only a weak relation to endorsement of constraints for some of the same reasons that decision researchers distinguish between stated and revealed preference and linguists distinguish competence and performance. I thank a reviewer for raising this concern.

In another test, Study 2 found that judgments of rule violation and disapproval were more highly correlated in domains for which people endorsed moral rules than in domains for which they did not. So, rules, emotional process, and deliberation each seem important for shaping moral judgment and preference.

The studies reported here are also consistent with the hypothesis that constraints are "affect-backed" – that they are intuitively available, and that moral judgment and preference is mediated by emotional process that can be exacerbated (as in the Vivid condition of Study 1) or diminished in one of two ways explored here, one opportunistic and one inhibitory. First, the opportunistic: adherence to constraints could be overwhelmed by utilitarian considerations when the stimulus highlighted utilitarian considerations (as in Catastrophe condition of Study 1, and the joint evaluation context of Study 3). Second, the inhibitory: these studies found that people prone to deliberative thinking could, as Greene et al. (2001) argue, over-ride the response consistent with adhering to these affect-backed constraints (Studies 1 and 2).

One view of the processes involved in moral judgment that seems consistent with Nichols and Mallon (2006) is the following: A participant forming a moral preference or judgment reads about some hypothetical situation and (1) feels a flash of negative affect triggered by a perceived rule violation (or not, if he or she does not possess the rule), (2) assesses utilitarian considerations if the situation makes them salient and if she is motivated to engage her computational resources to do so, and (3) registers a judgment or preference. This kind of approach recognizes the powerful influence of emotional activation on moral cognition, but treats it *as an input* to a response, rather than the response. It also suggests that many deontology-consistent responses are principled (i.e., sentiment predicated on rules). Finally, it suggests that while rules are important, people may reluctantly forego their proscriptions.

It should be noted that the theorizing in Nichols and Mallon (2006) relies heavily on the presence of moral rules and processes that make use of them. However, since they never actually test for the presence of moral rules, they might interpret the relationship between deontological response and idealism (Study 1) and PVs (Study 2)

and the test of the relationship between judgments of rule violation and disapproval (in Study 2) as stronger support for their framework than they adduce.

4.2. Revisiting the protected values framework

The protected values framework tests for the presence of moral rules but has been somewhat less focused on psychological process that make use of those rules than some of the accounts intended to capture people's responses to trolley problems. One purpose of Studies 2 and 3 was to test whether some of the competing processes identified in the trolley problem literature are at work in PV-driven judgment and decision making. Studies 2 and 3 developed an interpretation of PVs as affect-backed constraints – intuitively available, emotionally-charged moral rules that can be overwhelmed in a variety of contexts (as noted in the section above). So, PVs could be thought of as a crucial constituent in a simple process-based approach to understanding morally-motivated judgment and decision making.

In my view, this is not so much a reconceptualization of the construct as it is a process-informed supplement to the protected values literature. Indeed, Study 2 and Bartels and Medin (2007) offer a great deal of support for predictions motivated by previous empirical work on the role of protected values in decision making. As noted earlier, one part of Bartels and Medin (2007) was a replication of Ritov and Baron (1999): in one condition, PVs were associated with quantity insensitivity, which is taken as evidence that PVs motivate deontological preference. Similarly, in the current Study 2, people endorsing PVs indicated less approval for decisions made by third parties who decided to knowingly do harm to a resource on the basis of cost-benefit analysis. This disapproval is consistent with the idea that PVs motivate nonconsequentialist judgment and preference.

Deontology evaluates actions with respect to constraints, and those contexts that highlight the relationship between actions and moral rules are the contexts for which the PV framework appears most descriptively adequate. However, when consequences are highlighted, either by the preference elicitation procedure (Bartels & Medin, 2007) or the comparative nature of the choice context (Study 3), people with PVs no longer appear to be absolutist deontologists.

It seems reasonable that people who care more about not harming a resource (people with PVs) might also tend to care a great deal about the ultimate consequences realized in a domain (i.e., the good to be promoted). This rationale makes sense of the finding that PV-driven preference sometimes appears at least as utilitarian (in Study 3) or even *more utilitarian* (Bartels & Medin, 2007) than non-PV-driven preference.

The split in philosophy between deontologists and utilitarians is clear, but it is clearly not as pronounced a dissociation in participants' minds. I am, in effect, arguing that for some domains, a given person can be both more deontological and more utilitarian than his or her dispassionate counterpart. The findings reviewed in this section suggest a link between PVs and utilitarianism. But to be clearer, the argument I present here is not that PVs motivate utilitarianism, but rather that

affect-backed constraints (i.e., PVs) are present for some domains in which people care immensely about utilitarian considerations.

4.3. On combining experimental and individual differences-based approaches

The current approach placed a greater emphasis on individual differences than some other approaches have. Researchers who investigate processes involved in moral judgment tend to neglect the variance attributable to individual differences. Some research programs aim to identify universal principles of moral cognition (Hauser, 2006; Mikhail, 2007). Because parsimony is valued, many research programs develop general-function models of how a randomly-chosen individual (maybe a member of a specific culture, SES group, or gender) produces a moral judgment. The current studies demonstrate how an individual-differences based approach that accounts for variance *within a sample* can complement experimental tests of psychological process.

Underwood (1975) argues for the importance of individual differences in the development of psychological theory. "If we include in our nomothetic theories a process or mechanism that can be measured reliably outside the situation for which it is serving its theoretical purpose, we have an immediate test of the validity of the theoretical formulation, at least a test of this aspect of the formulation. The assumed theoretical process will necessarily have a tie with performance which reflects (in theory) the magnitude of the process. Individuals will vary in the amount of this characteristic or skill they 'possess.' A prediction concerning differences in the performance of the individuals must follow. If the correlation is substantial, the theory has the go-ahead signal, that and no more. If the relationship between the individual differences measurements and the performance is essentially zero, there is no alternative but to drop the line of theoretical thinking (p. 130)."

The studies reported here treated individual differences as instructive, using differences between people to inform an account of moral cognition. These studies suggest that perhaps the general-function models of moral cognition that do not make different predictions for different people ought to do so: Different models may better characterize different people, and the fit between model and behavior is predictable by reliable individual differences. The current studies were made more informative by combining experimental and individual differences-based methods to account for variability in moral cognition, both across contexts *and* across individuals.

5. Conclusions

The studies reported here find that morally-motivated judgment and preference (a) makes use of intuitive and deliberative process, (b) is influenced by the judgment-eliciting context, and (c) recruits representations of both deontological constraints and utilitarian considerations. These studies implicate a number of processes that combine to produce context-sensitivity in morally-motivated judgment and preference, suggesting that moral cognition is a hodge-podge of sorts.

That moral cognition is "messy", in this way, has led others to be dubious about *even the possibility* of adequate normative and descriptive theory. For example, Nichols and Mallon (2006) write:

> "It is probably unrealistic to expect a tidy processing account of how these factors interact to generate judgments of all-in impermissibility. But the fact that multifarious psychological factors impact judgments of all-in impermissibility brings us back to the difficulty philosophers have had in reaching a unified normative theory that captures our intuitions about moral dilemmas. If judgments of all-in impermissibility arise from the interaction of a diverse collection of psychological mechanisms – representations of prohibitions, utilitarian assessments, and emotions – then it is probably misguided to expect that there is a single normative criterion that can capture our intuitions about moral dilemmas."

Nichols and Mallon may be right on the normative point—that a set of normative principles that fully capture our moral intuitions may be hard to come by, but I do not share the opinion that the descriptive principles underlying moral cognition will be as hard to identify and account for. Studies like the ones reported here can form the basis of a reasonably well-constrained process-based explanation that accounts for much of the flexibility of morally-motivated cognition.

Acknowledgements

I thank Doug Medin, Lance Rips, Reid Hastie, Chris Bauman, Russ Burnett, Eugene Caruso, Jason Jameson, and especially Craig Joseph for comments on previous papers in this line of work. Thanks also to Steven Sloman, Joshua Greene, and two anonymous reviewers for constructive and generous criticism of an earlier draft. This work was partially supported by Grant No. 3000633806 from the National Science Foundation and Grant No. 10001864 MOD 02 from the Air Force Office of Scientific Research.

Appendix A. Individual differences items used in Study 1

(*Note:* Rational–Experiential Inventory also used in Study 2)

A.1. Ethics position questionnaire

1. It is never necessary to sacrifice the welfare of others.
2. Moral behaviors are actions that closely match ideals of the most "perfect" action.
3. Risks to another should never be tolerated, irrespective of how small the risks might be.

4. People should make certain that their actions never intentionally harm another even to a small degree.
5. One should never psychologically or physically harm another person.
6. The dignity and welfare of the people should be the most important concern in any society.
7. The existence of potential harm to others is always wrong, irrespective of the benefits to be gained.
8. If an action could harm an innocent other, then it should not be done.
9. One should not perform an action which might in any way threaten the dignity and welfare of another individual.
10. Deciding whether or not to perform an act by balancing the positive consequences of the act against the negative consequences of the act is immoral.

A. 2. Rational–Experiential Iventory

Need for Cognition – "Deliberative" Items

1. I enjoy intellectual challenges.
2. I enjoy solving problems that require hard thinking.
3. I prefer complex to simple problems.
4. I am much better at figuring things out logically than most people.
5. I try to avoid situations that require thinking in depth about something. (neg)
6. I'm not that good at figuring out complicated problems. (neg)
7. I am not very good at solving problems that require careful logical analysis. (neg)
8. I don't like to have to do a lot of thinking. (neg)
9. Thinking is not my idea of an enjoyable activity. (neg)
10. I am not a very analytical thinker. (neg)

Faith in Intuition – "Experiential," or "Intuitive" Items

1. I like to rely on my intuitive impressions.
2. Using my gut feelings usually works well for me in figuring out problems in my life.
3. I believe in trusting my hunches.
4. Intuition can be a very useful way to solve problems.
5. I often go by my instincts when deciding on a course of action.
6. I trust my initial feelings about people.
7. I don't have a very good sense of intuition. (neg)
8. I don't like situations in which I have to rely on intuition. (neg)
9. If I were to rely on my gut feeling, I would often make mistakes. (neg)
10. I think it is foolish to make important decisions based on feelings. (neg)

Appendix B. Ethical dilemmans from Study 1

(*Note 1:* Catastrophe manipulation in parentheses, vividness manipulation appears in bracket at end of item)

Footbridge. (Thomson, 1985)

In the path of a runaway train car are five(nineteen) railway workmen who will surely be killed unless you, a bystander, do something. You are standing on a pedestrian walkway that arches over the tracks next to a large stranger. Your body would be too light to stop the train, but if you push the stranger onto the tracks, killing him, his large body will stop the train. [You try to block out the terrible images of him falling and of his body being impacted by the train].

In this situation, would you push him?

Submarine. (Greene et al., 2001)

You are the captain of a small military submarine traveling underneath a large iceberg. An onboard explosion has dramatically decreased the oxygen supply and has injured a member of the crew. He may die from his injuries. The remaining oxygen is not sufficient for the entire six(twenty) – man crew to make it to the surface. If you shoot the injured crewmember, there will be just enough oxygen for everyone else to survive. [As you load your gun, the injured man desperately pleads for you not to kill him].

In this situation, would you shoot the crewmember?

Hostages. (Greene et al., 2001)

You are traveling with five(nineteen) other peace advocates in a war-torn country when your group is taken hostage by a group of terrorists. After a week, the terrorists' demands have not been met, and they offer you a choice: either they will kill all of you, or if you execute one of your fellow hostages, you and the others are free to go. [You are handed a knife, and your fellow hostages scream in terror, pleading for their lives].

In this situation, would you execute one of your fellow hostages?

Surgery. (Foot, 1967)

You are a surgeon with a number of patients. Five(Nineteen) of them need organ transplants. Each of them needs a different organ or they will surely die. You have another patient who is healthy and would be an ideal organ donor for the others. If you transplant his organs (against his will) into the bodies of the other patients, they will live but he will die. [You approach him with anesthesia and a scalpel. He pleads with you not to take his life].

In this situation, would you perform this transplant?

Trespassers. (Greene et al., 2001)

You and a group of friends are backpacking, and while crossing a rocky plain in the Guatemalan highlands, you inadvertently camp on a local clan's burial ground. The leader of the clan says that his clan's laws command him to kill all six(twenty) of you as a sacrifice to the gods, but that if you kill one of your own, the gods will be appeased, and the rest will be let go. [He hands you a dagger. Your friends' eyes fill with terror. They beg you to not to kill them].

In this situation, would you kill one of your friends?

Liferaft. (Regan, 1983)

You are on a small ship, a fire breaks out, and the ship has to be abandoned. Because your tiny liferaft is carrying more than its capacity, it is sitting dangerously low in the water. The seas get rough and the raft begins to fill with water. Unless you do something, all six(twenty) of you will drown. There is an injured person onboard who may die either way. If you throw him overboard, everyone else will be saved. [Lying there almost helpless, he whimpers, knowing he will drown because he is unable to swim in his condition].

In this situation, would you throw him overboard?

Spelunkers. (http://news.bbc.co.uk/1/hi/magazine/4954856.stm)

You and five(nineteen) others are exploring a seashore cave. A rock falls and blocks your exit. The tide is rising. You spot a hole elsewhere and let a large man in your group out first. He gets stuck, and unless you get out soon, everyone but this man (whose head is sticking out of the cave) will drown. You have a stick of dynamite that will not move the rock, but will blast the man out of the hole. He does not want to die; neither do you or the others. [He is overcome by terror. Shaking in fear, he begs and pleads for you to not to kill him].

In this situation, would you blast him out?

Derailment. (Unger, 1996)

In the path of a runaway train car are five(nineteen) railway workmen who will surely be killed unless you, a bystander, do something. If you flip a switch, the train will be diverted onto a set of tracks in need of repair. The train will be derailed and go down a hill, across a road, and into a man's yard. The owner, sleeping in his hammock, will be killed. [You try block out the image of his body being shred to pieces as it flies through the patio door].

In this situation, would you flip the switch?

Baby. (Greene et al., 2001; see also Alda et al., 1983)

Enemy soldiers have taken over your village and will kill all remaining civilians. You and five(nineteen) others are hiding in the cellar of a large house. Soldiers have come to search the house for valuables. A baby in your group begins to cry. So, you cover her mouth, but she cannot breathe. If you remove your hand, the baby can breathe, but her crying will summon the soldiers who will kill everyone in the cellar. [The baby is writhing around violently in your arms. With all her might, she desperately struggles to breathe].

In this situation, would you smother the baby?

Bystander. (Foot, 1967)

In the path of a runaway train car are five(nineteen) railway workmen who will surely be killed unless you, a bystander, do something. If you flip a switch, the train will be diverted onto another track, where it will kill a single railway workman. [You do everything in your power to block out the terrible image of the train impacting his body].

In this situation, would you flip the switch?

Plane Crash. (Greene et al., 2001)

Your plane has crashed in the Himalayas. The only survivors are you, some other men, and a young boy. The six(twenty) of you travel for days, battling extreme cold

and wind. Your only chance of survival is to make it to a village a few days away. The boy cannot move very quickly. Without food, you and the other men will surely die. One of the men suggests killing the boy and eating his remains over the next few days. [The boy is shocked and terrified. He does not want to die. He starts crying and begging for your mercy].

In this situation, would you sacrifice the boy?

Fumes. (Thomson, 1986)

You are the late-night watchman in a hospital where an accident has occurred in one of the on-site testing labs, and now there are deadly fumes rising up through the hospital's ventilation system. The fumes are headed to a certain area where there are five(nineteen) patients who will surely die. If you flip a switch, the ventilation system will cause the fumes to bypass this room and enter a room containing a single patient, killing him. [You try not to imagine how the person in this room will writhe violently in his gurney, gasping for air].

In this situation, would you flip the switch?

Prisoners of War. (Baron, 1992)

You and some other soldiers were captured. After a year in a prison camp, your group tried to escape but was caught. The warden has decided to hang your group in front of the other prisoners of war. At the gallows, he releases the noose from your neck and announces that if you pull the chair from underneath one man in your group, the remaining five(nineteen) will be set free, otherwise you all die. He means what he says. [As you approach the chair, you try block out the image of your cell-mate's body writhing violently as he hangs].

In this situation, would you remove the chair?

Soldiers. (Greene et al., 2001)

You are leading a group of soldiers returning from a completed mission in enemy territory when one of your men steps in a trap. He is injured, and the trap is connected to a device that alerts the enemy to your presence. If the enemy finds your group, all six(twenty) of you will die. If you leave him behind, he will be killed, but the rest of the group will escape safely. [You hear him crying, desperately in need of help, begging you not to leave him there to be killed].

In this situation, would you leave him behind?

Appendix C. Stimuli used in Studies 2 and 3

(*Note:* Omission appears first, Action appears second below.)

Birds. During the final stages of constructing an amusement park, an area in which a species of endangered birds nests will be disturbed. Scientists estimate that 100 endangered birds on the northwest end of the site will die as a result. Scott (Steve) is considering building some barriers that will save these birds, but the barriers will cause some other birds of this endangered species on the south-east end of the site to die as a result. Because the construction is so far along, a decision must be made quickly, and the government's options are severely constrained.

Scott does not want to kill any birds in the southeast end of the site. So, the barriers are not built. The 100 birds in the northwest end of the site die.

Steve wants to save the birds in the northwest end of the site. He first calculates that putting up the barriers will kill 80 birds in the southeast end of the site. Knowing that doing so will kill many birds, he chooses to build the barriers.

Children. Julie (Rich) is considering a vaccination program. Epidemiologists estimate that vaccinating 600 children will prevent them from dying from an epidemic of a new infectious disease. The vaccine itself will kill some number of children because it sometimes causes the same disease. Because this disease progresses rapidly, a decision must be made quickly, and the government's options are severely constrained.

Julie does not want to kill any of the children with the vaccine. So, the vaccine is not administered. The 600 children die.

Rich wants to save the children from the disease. He first calculates that administering the vaccine will kill 480 children. Knowing that doing so will kill many children, he chooses to vaccinate the children.

Dolphins. An area off the southeast coast of Florida is heavily populated with dolphins and tuna. Tuna fisherman accidentally catch a number of dolphins in this area each year. The dolphins that are caught in the tuna nets drown. If nothing is done, scientists estimate that 60 dolphins in this area will drown in the next year. Linda (Laura) is considering forcing the boats to fish in a different area where they will catch just as many tuna, but some dolphins will drown in the second area as a result. Because the tuna fishing season is about to start, a decision must be made quickly, and the government's options are severely constrained.

Linda does not want to kill dolphins in the second area. So, the fishermen are not forced to switch areas. The 60 dolphins in the first area drown.

Laura wants to save the dolphins in the first area. She first calculates that making the fishermen switch areas will kill 48 dolphins in the second area. Knowing that doing so will kill many dolphins, she makes the fishermen switch areas.

Fish. A flash flood has changed the water levels upstream from a dam on a nearby river. Scientists estimate that 20 species of fish upstream from the dam are threatened with extinction. Paul (David) is considering opening the dam, which will save these species, but some species downstream will become extinct because of the changing water level. Because this flood has rapidly changed water levels, a decision must be made quickly, and the government's options are severely constrained.

Paul does not want to kill any of the fish species downstream. So, the dam is not opened. The 20 species upstream die.

David wants to save the fish species upstream. He first calculates that opening the dam will kill 16 species downstream. Knowing that doing so will kill many fish, he chooses to open the dam.

Jobs. An economic downturn has caused job cuts at manufacturing plants. Joe (Mary) is considering cutting some financial support from one plant and reallocating those funds to a second plant. Economists estimate that reallocating these funds will save 300 people from losing their jobs in the second plant, but some number of workers in the first plant will be laid off as a result. Because the downturn was unexpected,

a decision must be made quickly, and the government's options are severely constrained.

Joe does not want to terminate the jobs of anyone in the first plant. So, the funds are not reallocated. The 300 people in the second plant lose their jobs.

Mary wants to save the people's jobs in the second plant. She first calculates that reallocating the funds will terminate the jobs of 240 people in the first plant. Knowing that doing so will terminate the jobs of many people in the first plant, she chooses to reallocate the funds.

The Poor. Funds for treating poor people afflicted with cancer are limited. Liz (Mike) is considering withdrawing funds that subsidize an expensive treatment for one kind of cancer to subsidize a less expensive treatment for a second kind of (equally-bad) cancer. Medical experts estimate that reallocating these funds will cure 300 poor people afflicted with the second kind of cancer, but some number of people suffering from the first kind of cancer will die because they will not be able to afford treatment. Because these cancers progress rapidly, a decision must be made quickly, and the government's options are severely constrained.

Liz does not want to kill people afflicted with the first kind of cancer. So, the funds are not reallocated. The 300 people afflicted with the second kind of cancer die.

Mike wants to save people suffering from the second kind of cancer. He first calculates that reallocating the funds will kill 240 people afflicted with the first kind of cancer. Knowing that doing so will kill many of people afflicted with the first kind of cancer, he chooses to reallocate the funds.

References

Alda, A., et al. (1983). Goodbye, farewell, and amen [Television series episode]. In B. Metcalfe (Producer), *M*A*S*H*. Los Angeles: 20th Century, Fox Television.

Amir, O., & Ariely, D. (2007). Decisions by rules: The case of unwillingness to pay for beneficial delays. *Journal of Marketing Research, 44*, 142–152.

Baron, J. (1992). The effect of normative beliefs on anticipated emotions. *Journal of Personality and Social Psychology, 63*, 320–330.

Baron, J. (1996). Do no harm. In D. M. Messick & A. E. Tenbrunsel (Eds.), *Codes of conduct: Behavioral research into business ethics* (pp. 197–213). New York: Russell Sage Foundation.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes, 70*, 1–16.

Bartels, D. M., & Medin, D. L. (2007). Are morally-motivated decision makers insensitive to the consequences of their choices? *Psychological Science, 18*, 24–28.

Bazerman, M. H., & Messick, D. M. (1998). On the power of a clear definition of rationality. *Business Ethics Quarterly, 8*, 477–480.

Bazerman, M. H., Moore, D. A., Tenbrunsel, A. E., Wade-Benzoni, K. A., & Blount, S. (1999). Explaining how preferences change across joint versus separate evaluation. *Journal of Economic Behavior & Organization, 39*, 41–58.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgments: Testing three principles of harm. *Psychological Science, 17*, 1082–1089.

Darwall, S. (Ed.). (2003a). *Consequentialism*. Oxford: Oxford University Press.

Darwall, S. (Ed.). (2003b). *Deontology*. Oxford: Oxford University Press.

Davis, N. (1993). Contemporary deontology. In P. Singer (Ed.), *A companion to ethics* (pp. 205–218). Malden, MA: Blackwell Publishing.

Donagan, A. (1977). *The theory of morality*. Chicago: University of Chicago Press.

Epstein, S., Pacini, R., DenesRaj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology, 71*, 390–405.

Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421–435.

Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: Reactions to transactions that transgress the spheres of justice. *Political Psychology, 18*, 255–297.

Fried, C. (1978). *Right and wrong*. Cambridge, MA: Harvard University Press.

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

Forsyth, D. R. (1980). A taxonomy of ethical ideologies. *Journal of Personality and Social Psychology, 39*, 175–184.

Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busemeyer, D. L. Medin, & R. Hastie (Eds.). *The psychology of learning and motivation: Decision making from a cognitive perspective* (Vol. 32, pp. 83–136). San Diego, CA: Academic Press.

Goodin, R. E. (1993). Utility and the good. In P. Singer (Ed.), *A companion to ethics* (pp. 241–248). Malden, MA: Blackwell Publishing.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An FMRI investigation of emotional engagement in moral judgment. *Science, 293*, 2105–2108.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*, 389–400.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*, 1144–1154.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J., Koller, S., & Dias, M. (1993). Affect, culture, and morality, or is it wrong to eat your dog? *Journal of Personality and Social Psychology, 65*, 613–628.

Hare, R. M. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Oxford University Press.

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

Hsee, C. K., & Leclerc, F. (1998). Will products look more attractive when presented separately or together? *Journal of Consumer Research, 25*, 175–186.

Hume, D. (1969). *A treatise of human nature*. London: Penguin (Original work published 1739).

Kant, I. (1966). *The moral law: Kant's groundwork of the metaphysics of morals*. London: Hutchinson University Press (Original work published in 1785).

Medin, D. L., Schwartz, H. C., Blok, S. V., & Birnbaum, L. A. (1999). The semantic side of decision making. *Psychonomic Bulletin & Review, 6*, 562–569.

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

Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition, 84*, 221–236.

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

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972–987.

Pettit, P. (1993). Consequentialism. In P. Singer (Ed.), *A companion to ethics*. Malden, MA: Blackwell publishing.

Quinn, W. S. (1989). Actions, intentions, and consequences: The doctrine of double effect. *Philosophy & Public Affairs, 18*, 334–351.

Regan, T. (1983). *The case for animal rights*. Berkeley: University of California Press.

Rettinger, D. A., & Hastie, R. (2001). Content effects on decision making. *Organizational Behavior and Human Decision Processes, 85*(2), 336–359.

Rettinger, D. A., & Hastie, R. (2003). Comprehension and decision making. In S. L. Schneider & J. J. Shanteau (Eds.), *Emerging perspectives on judgment and decision making* (pp. 165–200). Cambridge, UK: Cambridge University Press.

Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes, 79*, 79–94.

Shweder, R. A., Mahapatra, M., & Miller, J. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1–90). Chicago: University of Chicago Press.

Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology, 27*, 76–105.

Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly, 44*, 684–707.

Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109*, 451–471.

Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences, 7*, 320–324.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853–870.

Tetlock, P. E., Peterson, R., & Lerner, J. S. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. Olson, & M. Zanna (Eds.), *Ontario symposium on social and personality psychology: Values*. Hillsdale, NJ: Earlbaum.

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

Thomson, J. J. (1986). *Rights, restitution, and risk*. Cambridge, MA: Harvard University Press.

Underwood, B. J. (1975). Individual differences as a crucible in theory construction. *American Psychologist, 30*, 128–134.

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

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science, 17*, 476–477.

Viscusi, W. K. (2000). Corporate risk analysis: A reckless act? *Stanford Law Review, 52*, 547–597.

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.