NORTHWESTERN UNIVERSITY

Accounting for Some of the Flexiblity of Morally-Motivated Judgment and Decision Making

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology

By

Daniel M. Bartels
EVANSTON, ILLINOIS
December 2007

ABSTRACT

Accounting for Some of the Flexiblity of Morally-Motivated Judgment and Decision Making

Daniel M. Bartels

Why would someone's judgments and choices disregard the consequences he or she cares about most? Considerable research (Baron & Spranca, 1997; Baron & Leshner, 2000; Ritov & Baron, 1997; Tetlock, 2002, 2003; Tetlock, Kristel, Elson, Green, & Lerner, 2000) appears to show that in many contexts, people show precisely this paradoxical tendency. Researchers interpret these results as showing that people sometimes make choices on the basis of moral factors — factors that may have little to do with the likely consequences of a choice alternative.

In other words, previous research in judgment and decision making has documented a number of departures from expected utility theory, which is generally taken to be the "gold standard" for rationality. Utilitarianism as a normative ethical theory attempts to justify moral judgments with the demands of rationality in a similar way. Because "morally-motivated" judgment and decision making sometimes disavows any consideration of utilitarian concerns (i.e., costs and benefits), it has been viewed as especially *rigid*.

The studies in this thesis attempt to probe deeper than previous research into how the processes underlying morally-motivated judgment and decision making differ from those underlying *non-moral* judgment and decision making. They show that a number of variables exert powerful influences on moral cognition, and they suggest a different conclusion than that arrived at by previous research. Namely, they suggest moral cognition is flexible and context-sensitive, and that these properties of moral judgment and preference are diagnostic of a number of competing influences and psychological processes.

The studies reported here examine people's use of utilitarian and non-utilitarian judgment and decision making strategies across a variety of experimental paradigms and contexts of judgment and choice. The studies demonstrate the flexibility of value-driven judgment and preference, explaining a number of different ways in which utilitarian and non-utilitarian judgment and decision principles are promoted.

Part One casts doubt on the simplistic picture of morally-motivated judgment and preference as characteristically rigid. Some previous views (e.g., Baron & Spranca, 1997) suggest that when moral values are brought to bear on a choice situation, we should expect a single pattern of responses that is uncharacteristic of non-moral preference. Studies 1 and 2 suggest that, in fact, what differentiates morally-motivated from non-moral preference is that the former is especially *flexible* in some contexts. Whereas non-moral preference is relatively invariant across experimental conditions in Studies 1 and 2, morally-motivated preference is *more malleable,* showing greater effects of manipulations that target attentional (Study 1) and representational processes (Study 2).

Part Two reiterates the point made in Part One, that moral cognition is flexible because it is constituted by multifarious psychological processes. Moral cognition appears to be part emotion-laden intuition *and* part deliberation, and appears to make use of both moral rules *and* assessments of the costs and benefits associated with different courses of action. Studies 3 and 4 account for some of the flexibility of moral judgment and preference by developing a process-based framework that incorporates methodologies and insights from a number of perspectives in moral psychology and philosophical ethics.

Taken together, the current studies offer initial steps toward synthesis across theoretical perspectives in moral psychology and generalization of when we should expect utilitarian and non-utilitarian judgment and decision making across a variety of contexts.

Acknowledgements

I would like to thank Doug Medin for his invaluable advice throughout my graduate training and for being a great role model. His discretion, modesty, tact, and concern for others are more exemplary than he can appreciate. I would also like to thank Lance Rips and Reid Hastie for their perspicacity. Nothing ever gets by either of them, and I have benefited immensely from their incisive feedback on this project and others. I also owe a debt of gratitude to each of the following graduate students and post-docs who endured at least one presentation or write-up of at least one of these studies: Russ Burnett, Will Bennis, Jason Jameson, Chris Bauman, Sam Day, Rumen Iliev, Sonya Sachdeva, Sara Unsworth, Jennifer Asmuth, Eyal Sagi, and Winston Chang. Special thanks go to Craig Joseph, whose comments made the writing of this thesis markedly better. Thanks to Florence Sales and Tomeka Bolar for making the Psych Department run smoothly throughout my tenure at Northwestern. I am more grateful than I can express to the three most important people in my life. A million thanks to my wife, Tab, and to my mom, Carol, for approximately two thirds of the unrelenting support that makes all things possible. And finally, these and all of my efforts are dedicated to my dad, Doug Bartels, the most selfless, virtuous person who anyone who has ever met him has ever known.

Table of Contents

1 General introduction

1.1 "Morally-motivated decision making" in the judgment and decision making literature

Moral choice is pervasive: many of the decisions people face could be construed as "moral" ones. Baron (2000) writes that "Moral thinking is important for decision making as a whole, since most real decisions involve moral issues. We often are not aware that moral issues are involved in our everyday decisions, but wherever our choices affect the utilities of others, a moral decision must be made" (p. 381). Whether or not one agrees with this characterization of moral choice (and the close association of morality with utility), understanding the processes underlying value-driven judgment and preference is an important goal for the behavioral sciences.

Psychologists interested in the judgment and reasoning processes underlying choice assume that in most contexts, people choose in such a way as to bring about the best consequences (broadly construed) from their point of view. Descriptive models of decision making show that important components of the choice process (for example valuation processes and tradeoffs between attributes or alternatives) vary across social contexts and across domains (e.g. financial, personal and legal, to name a few) (Fiske & Tetlock, 1997; Goldstein & Weber, 1995; Rettinger & Hastie, 2001, 2003). Most importantly for this thesis, the processes underlying decision making have been found to differ depending on whether or not moral values are brought to bear on the choice (Medin, Schwartz, Blok, & Birnbaum, 1999; Tenbrunsel & Messick, 1999).

The influence of moral values on judgment and decision making could hardly be a more important or relevant research topic. Values influence how people make sense of their social

worlds, and they ground many of the judgments and decisions we find meaningful in our everyday lives. If we want to comprehend people's actions, both commonplace and extraordinary, we must understand the values that inspire them.

"Extreme" actions, such as selfless heroism and suicide terrorism, are only the most vivid examples of the influence that values may have on behavior. Even seemingly mundane decisions — like what kind of coffee to drink, which type of furniture to purchase, or whether or not to make a purchase at all — may be driven by moral values (e.g., Ehrich & Irwin, 2005). Indeed, some psychologists have suggested that "all attitudinal and behavioral decisions should be traceable to personal value priorities" (Rohan, 2000, p. 270).

The literature on value-driven choice suggests that moral values influence and sometimes conflict with processes (e.g., tradeoff calculus and valuation processes) ascribed to *homo economicus*. The dominant view is that in most choice situations, people seek to maximize "utility" — that a person's choices are driven by the value he or she seeks to attain through realization of the outcomes brought about by his or her choice. *Normative* models judge the quality of choice by the degree of utility maximization. *Descriptive* models both attempt to explain why people's choices fail to maximize utility, and they share the view that even people's suboptimal choices are a reflection of the outcomes they desire. *Morally-motivated* choice is a more complicated phenomenon for the *homo economicus* view, because it is guided by moral principles that may have little to do with the relationship between consequences and self-interest (revisit selfless heroism, suicide terrorism).

Recent empirical work seems to find a link between moral values and a lack of concern for consequences (Tetlock, 2002; Baron & Spranca, 1997), and some of the conclusions reached in this research are startling. For decisions involving cherished resources, we

sometimes appear to be driven less by the costs and benefits associated with a course of action than by moral rules concerning the ways these resources should be treated (Tetlock, 2002; Baron & Spranca, 1997). These sorts of resources are often bestowed a kind of "protected" status, and people react strongly to proposed tradeoffs of these resources for other kinds of resources (especially money) on moral grounds. One formulation that is likely to be recognizable is "You can't put a price on a human life!" (e.g., Tetlock, Kristel, Elson, Green, & Lerner, 2000).

More startling is the paradoxical suggestion raised in the literature on "protected values" (PVs) (Baron & Spranca, 1997) that morally-motivated decision makers sometimes disregard outcomes *entirely*. People with PVs for an issue (e.g., endangered species) — those who view causing its extinction as impermissible, no matter what the benefits — are more likely to exhibit "quantity insensitivity" than people without PVs for the issue. For example, they are more likely to view the extinction of one species as being equally wrong as the extinction of five species (Baron & Spranca, 1997).

A second, more compelling example comes from Ritov and Baron (1999). Participants in these studies read about scenarios where the only way to promote the best outcomes for the resources in question was to engage in an action that caused harm to the resource. For example, participants read about a situation where the only way to save 20 species of fish upstream from a dam on a river was to open the dam. However, opening the dam would kill two species of fish downstream. Participants were then asked whether they would open the dam, and for the maximum number killed downstream at which they would open the dam.

In this situation, some participants say they would not open the dam. Some even say they would not want to cause the loss of a single species, *even though not opening the dam*

leads to the loss of all 20 species. This kind of "quantity insensitive" response is more likely to be made by people who endorse a PV for the relevant resource.

The link between PVs and a lack of concern for consequences is both paradoxical and disquieting. If we assume that people hold PVs for things they cherish (e.g., family), we would expect them to care more about consequences for these cherished resources than about consequences for other, less valued resources. Furthermore, if PVs motivate non-consequentialism, then not only do people who really care about an issue fail to maximize their utility, they appear comparatively ignorant — *they might not be taking stock of the consequences at all.*

How can this be? Why would someone make choices that disregard the consequences he or she cares about most? This thesis offers a partial answer.[1] When people care immensely both about moral rules and about consequences, they can be induced to afford greater weight to *either* factor in forming preferences. The studies that follow investigate the processes underlying moral judgment and choice, and along the way, they undermine a few over-simplifications made by frameworks that purport to explain moral judgment and preference.

1.2. Demonstrations and explanations of the *flexibility* of moral choice

[1] I stress the word *partial* — there are likely *many* answers. Some of my colleagues are pursuing other potential explanations of (apparently) non-consequentialist morally-motivated choice. One line of research examines a kind of lay rule-consequentialism, where people say their choices accord with moral rules of right and wrong because they believe that relying on rules will ultimately result in greater long-term benefits. A second line of research tests whether people appear to disregard consequences because they doubt the plausibility of scenario presented to them — where morally-motivated decision makers, in particular, wish or believe that fish upstream from the dam on the river will not actually die if the dam is not opened, for example. Instead of discussing these interesting possibilities, this thesis reflects the foci of the empirical and philosophical work that motivates it.

The protected values framework seems to charge specifically moral cognition with a kind of rigidity. When moral values are implicated in a decision they induce the decision-maker to adopt an automatic and highly informationally-constrained decision procedure, and the resulting choice will be blind to consequences (see Hauser, 2006, p. 11). The implication, that people might be less likely to take stock of precisely those consequences about which they care the most, is baffling — and probably incorrect. At the least, this seems like a somewhat over-simplified account of how moral values influence decision making.

The working hypothesis developed in Part One of this thesis is that what is distinctive about morally-motivated choice is *not* that it is especially unconcerned with consequences. In fact, in some contexts decision making driven by moral values does not produce any *one* response pattern. The two studies discussed in Part One demonstrate that morally-motivated decision-making can be *more* flexible — more penetrable by experimental manipulations of the choice context — than non-moral preference.

But simply to conclude that the picture is more complicated than previously thought is not enough. Therefore, in Part Two, I report the findings of further studies that identify specific processes that interact to shape judgment and preference, and thereby account for some of the malleability of morally-motivated choice.

1.3. Consequentialism and deontology as psychological processes

The studies in this thesis examine the processes motivated by moral values and attitudes, and their findings warrant a critical examination of the explanatory power of contemporary frameworks that intend to explain moral cognition (e.g., Baron & Spranca, 1997; Greene, Nystrom, Darley, & Cohen, 2001; Nichols & Mallon, 2006). These frameworks

reference two modes of moral reasoning: (1) *consequentialism* (choosing based on costs and benefits of actions) and (2) *deontology* (in which choice is dictated by invariant rules of rightness and wrongness of actions themselves).

Although many perspectives in moral psychology and philosophy use "deontology," "consequentialism" and the like as explanatory constructs, little theoretical synthesis has been achieved. The theoretical picture is messy, perhaps because few studies aim at descriptive generalization across contexts (i.e., different types of judgment and choice situations). The studies reported here aim to bring together insights from different research programs in moral psychology to better understand decision makers' use of utilitarian and non-utilitarian principles in forming judgments and preferences.

The current studies proceed by demonstrating how morally-motivated decision making is malleable, apparently more sensitive to manipulations intended to influence attentional process (Study 1) and mental representation (Study 2) in some contexts. The second group of studies (Studies 3 and 4) builds on the first. They provide an account of how contexts that direct attention to consequences or to rules can interact with the presence of situation-specific values and individual differences to shape moral judgment and choice.

Because these studies examine the use of "deontological" and "consequentialist" judgment and decision principles, some preliminary remarks on these concepts are in order. Deontology and consequentialism are two qualitatively distinct positions in normative ethics. They attend to different features of moral choice scenarios, thereby arriving at different notions

of the "right" way to make moral choices.[2] Rather than discuss the relative merits of these positions,[3] I will focus on describing the reasoning processes they imply, presenting strong versions of each position to draw out contrasts for illustrative purposes.

Normative ethical theories offer bases for judging acts as morally forbidden, permissible, or obligatory.[4] Both deontology and consequentialism hold that the goodness of the consequences produced by an act are relevant for determining its moral status, but they differ in key respects. Strict consequentialism acknowledges only one morally-relevant feature of situations — the goodness of consequences, as evaluated from an impartial point of view. In contrast, deontology holds that some actions are simply obligatory or impermissible because they are enjoined or forbidden by moral rules[5], and such rules, when applicable, outweigh considerations of the goodness of consequences (Kagan, 1998).

---

[2] Although other psychologists have argued for the normative superiority of deontology (e.g., Kohlberg, 1969) or consequentialism (e.g., Baron, 2000), I propose to bracket the normative questions of what people are morally required to do and how people ought to reason.

[3] Though philosophy has not solved the problem of which is right and/or under which circumstances people should reason one way or the other, philosophers who disagree on this issue (and on more fundamental ontological and epistemological issues), agree on one point that justifies a neutral approach. Descriptively, people behave as if their moral cognition is objective and valid (Smith, 1994), and so "Ethical thought and feeling have 'objective purport'. From the inside, they apparently aspire to truth or correctness and presuppose that there is something of which they can be true or false" (Darwall, 1998, p. 25).

[4] As Pidgen (1993, p. 421) notes: "In philosophy, not only the doctrines, but the definitions of the doctrines are subject to dispute." I will try to avoid using problematic terms and cite the sources to which I attribute each potentially-controversial description of a philosophical construct.

[5] Judgments of action as per se impermissible are predicated on constructs that are variously referred to as constraints, laws, limitations, morals, norms, principles, prohibitions, proscriptions, rationales, rules, etc. Following Davis (1993) and Kagan (1998) I will use the terms "constraints" and "rules" interchangeably to refer to the constructs on which deontological judgment is predicated. "Constraint" is probably the preferable term, because the absolutist "rules" referred to are, by definition, exclusionary—actions that violate these moral rules are wrong,

Strict deontology attributes intrinsic significance to rules classifying acts as per se impermissible, setting up simple (or apparently simple) constraints on action, usually in the form of prohibitions. These constraints forbid agents from doing certain types of things — such as *knowingly harming* an innocent person — regardless of the goodness of the consequences that might result. For consequentialism, what matters is only that the best consequences be produced. In many contexts, consequentialism and deontology yield the same judgments regarding harmful acts: doing harm leads to worse consequences, other things being equal. But strict consequentialism treats prohibitions of harmful acts as akin to rules of thumb, which must be broken in cases where doing so would produce better consequences.

What is important for the purposes of this thesis is that these two normative positions imply different cognitive processes. Absolutist deontological reasoning checks certain qualities of actions (but not their consequences) against a set of rules that must be honored. The output of this process is that some acts are judged wrong in themselves, and thus are morally unacceptable (even as means to morally-obligatory ends; see Darwall, 2003a; Davis, 1993). Conversely, strict consequentialist reasoning is focused on ends — whatever values an individual adopts, this perspective mandates that one brings about the best consequences *by any means* (Darwall, 2003b; Pettit, 1993). Because the only inputs to this reasoning process are the consequences (and not their causes), the morally right course of action is the one that produces the best outcome.

In the studies reported here, I focus on the factors that *elicit* or *suppress* judgments and preferences consistent with consequentialism. More specifically, each study examines

---

simpliciter—precluding any consideration of elements of a situation that count for or against violating the rule (see also the discussion of rules, mandatory norms, and exclusionary reasons in Raz, 1975).

preferences consistent with and at odds with *utilitarianism*, a specific kind of consequentialist position. Utilitarianism, in a nutshell, has been interpreted to mandate the promotion of "the greatest good for the greatest number." Its critical features are that goodness be assessed from an impersonal point of view, and relatedly, that each person's goodness is to be counted equally.[6]

Because utilitarianism requires summing the degree of satisfaction of welfare interests (i.e., basic needs) across individuals, counting each equally[7], it mandates a simple-to-execute decision principle for the situations posed to participants in the current studies. That is, count up the number of lives saved by each alternative and choose accordingly. Choices that maximize the number of lives saved are consistent with utilitarianism. Otherwise, they are not.

1.4. Ethical dilemmas

[6] More specifically, utilitarianism "is the result of combining consequentialism with *welfarism*. Since consequentialism holds that an act is right if and only if it leads to the best consequences, and welfarism holds that the goodness of an outcome is ultimately a matter of the amount of individual well-being, counting everyone equally, it follows that utilitarianism is the view that an act is right if and only if it leads to the greatest total amount of well-being" (Kagan, 1998, p. 52), where "well-being", or "*welfare interests*", are "abstracted from actual and possible preferences. Welfare interests consist in just that set of generalized resources that will be necessary for people to have before pursuing any of the more particular preferences that they might happen to have. Health, money, shelter, sustenance, and such like are all demonstrably welfare interests of this sort, useful resources whatever people's particular projects and plans." (Goodin, 1993, p. 242)

[7] In applying this principle to the domain of health care and related public policy, sometimes the degree of satisfaction of welfare interests is calculated in terms of quality-adjusted life years — the product of (a) the life expectancy and (b) the "quality" of the remaining life years to be produced by a medical intervention. So, in practice, two or more people are only counted equally when neither of these factors is known to appreciably differ across them. For a criticism of this general approach to medicine, see La Puma and Lawler (1990), and for a critique of how quality-adjusted life years are measured, see Wakker (1996).

Philosophers often discuss morality by constructing ethical dilemmas intended to distill real world problems to their "essential" features. Ethical dilemmas typically pit moral rules against consequences: less acceptable actions result in better consequences; more acceptable actions result in worse consequences (for a host of examples, see http://www.friesian.com/dilemma.htm).[8] Whereas philosophers have used dilemmas to develop normative arguments, moral psychologists have used responses elicited by these dilemmas to develop descriptive accounts of moral cognition. (e.g., Cushman et al., 2006; Green et al., 2001; Mikhail, 2007; Nichols & Mallon, 2006; Waldmann & Dieterich, 2007).

Recent work has used variations on the "trolley problems" (one variety of ethical dilemma) in attempts to elucidate some psychological underpinnings of moral judgment. The most widely studied trolley problems are the "bystander" and "footbridge" cases (adapted from Foot, 1967 and Thomson, 1976).

The bystander case:

> *In the path of a runaway train car are five railway workmen who will surely be killed unless a bystander does something. If he flips a switch, the train will be diverted onto another track, where it will kill a single railway workman.*

The footbridge case:

> *In the path of a runaway train car are five railway workmen who will surely be killed unless a bystander does something. He is standing on a pedestrian walkway that arches over the tracks next to a large stranger. His body would be too light to stop the train,*

[8] Philosophers sometimes use the term "dilemma" in a more technical sense to refer to cases where two absolute duties, or constraints *all things considered*, conflict in an irresolvable way. For example, cases where the normative ethical theory under scrutiny mandates that an agent perform some act while also prohibiting the act would make it impossible to meet all of the moral obligations proposed by the theory. Such a case thus qualifies as a (technically-defined) dilemma and poses a problem for the principle-based ethical theory that judges the agent's behavior morally wrong no matter what she does (because, for example, a case like this violates deontic logic).

but if he pushes the stranger onto the tracks, killing the stranger, the stranger's large body will stop the train.

People tend to judge that the action in the bystander case — causing one death to save five — is acceptable, but that the action in the footbridge case — pushing a man to his death — is wrong (Hauser, 2006, Cushman, et al., 2006; Greene et al., 2001; Nichols & Mallon, 2006; Waldmann & Dieterich, 2007). If someone were strictly following a broad enough deontological rule, like "It is absolutely forbidden to intentionally kill someone (regardless of the consequences of not killing him)," she would respond "no" in both cases. (Of course, deontological constraints are usually more narrowly-defined and narrowly-directed than this rule; more on this point later.) Weighing costs (one life) and benefits (five lives) in a utilitarian manner (i.e., bring about the greatest good for the greatest number) would lead to a "yes" response in both cases.

These examples show that most people's intuition — that these cases should be treated differently — cannot be characterized by a strict version of either normative strategy. This normative status of the distinction drawn between these cases has been the subject of much philosophical debate (e.g., Foot, 1967; Quinn, 1989; Thompson, 1976), and the divergence in judgments rendered for these problems has motivated some clever recent experiments (Cushman et al, 2006; Greene et al., 2001; Mikhail, 2007; Nichols & Mallon, 2006). Part Two uses process-based theories of the judgments elicited by ethical dilemmas to explain the context-sensitivity of morally-motivated decision making demonstrated in Part One.

1.5. Utilitarianism as the normative standard in judgment and decision making research

Psychologists and behavioral economists who study decision making have largely adopted utility theory as a normative model (Savage, 1954; von Neumann & Morgenstern, 1947) and consequentialist theories as descriptive models (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). Most of the research has focused on how people assign utility to consequences, how they judge the likelihood of these consequences, and how they integrate this information when making decisions. These well-studied processes all presume that the decision maker's main consideration is the consequences of her choices. It is difficult to overstate the similarity between models of decision making and normative consequentialist philosophy. Amir & Ariely (2007, p. 150) write of the assumed default decision making process: "In summary, 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."

One strategy that has been used in judgment and decision making research is to (1) identify a normative model, (2) demonstrate ways in which people's responses systematically diverge from the predictions of the normative model, and (3) treat these "errors" as diagnostic of mental processes (see Kahneman, 2000). Sunstein (2005) uses this method to identify what he calls "moral heuristics." Although it is true that in many cases, consequentialism and deontology render the same judgment concerning an act's moral status (e.g., for sadistic, unprovoked aggression towards an innocent), sometimes their assessments diverge.

Recall that consequentialism treats deontological constraints as rather quaint "rules of thumb" that must be broken in cases where infringing the constraint promotes the good. Sunstein points to a number of contexts in which people's behavior appears more consistent

with non-consequentialist principles. “People use… moral short-cuts, or rules of thumb, that lead to mistaken and even absurd moral judgments. In [many] contexts, rapid, intuitive judgments make a great deal of sense, but sometimes produce moral mistakes... One implication is that moral assessments ought not to be made by appealing to intuitions about exotic cases and problems; those intuitions are particularly unlikely to be reliable. Another implication is that some deeply held moral judgments are unsound if they are products of moral heuristics” (Sunstein, 2005, p. 531).

Many of the themes in Sunstein’s article are extensions of utilitarianism to the domains of law and public policy. Critics argue that this approach is grounded in an inadequate normative theory — that choices serve more functions than just seeking utility (e.g., Tetlock, 2005), that the characterization of principle-constrained behavior as “error” is unfounded (e.g., Mikhail, 2005), or they question the necessity of adopting a normative model for understanding moral psychology (e.g., Pizarro & Uhlmann, 2005).

In the studies that follow, I will use utilitarianism as the primary standard of comparison — more like a default than a normative model — to which I will compare people’s responses. So, I examine influences that promote utilitarian judgment or undermine it (in Study 1) and influences that promote deontology-consistent judgment as opposed to utilitarian-consistent judgment (in Studies 2-5) without arguing for the superiority of one ethic over another.

1.6. Values as motivators of deontology

1.6.1. The sacred values framework

As noted above, people sometimes appear to reason and choose non-consequentially when they truly cherish the resources at issue. The literatures on "sacred values" (e.g. Tetlock, 2002) and "protected values" (e.g. Baron & Spranca, 1997) focus on the restrictive tradeoff rules participants appear to have for cherished resources, and suggest that strongly held, situation-specific values engender deontological decision strategies.

Many moral decisions, according to these frameworks, are more about adherence to moral rules than they are about maximizing benefits or minimizing costs. These frameworks intend to explain when and why moral values appear to drive people to decisions that, while they may not violate one's moral rules, appear almost blind to the ultimate consequences of one's choice. The sacred values framework (Fiske & Tetlock, 1997; Tetlock, 2002) explains such non-consequentialism as the result of a person having internalized a set of culturally-defined norms that constrain the manner in which different types of goods can be permissibly exchanged for each other.

For the sacred value framework, the decision principle to be followed is determined by the moral significance attached to the goods involved in the exchange. Certain cherished goods, like lives, health and nature, are treated by some communities as having moral, intrinsic or transcendental value. In all but the most extreme circumstances, these *sacred values* are not to be exchanged for *secular values* — goods that can be bought or sold. Such exchanges are judged to be morally reprehensible (Tetlock et al., 2000). Tetlock's (2002) framework suggests that when posed with "taboo tradeoffs" — sacred-for-secular tradeoffs, those where *only one of two or more resources is treated as morally significant* — decision makers view utilitarian considerations (i.e., costs and benefits of alternative courses of action) as strictly off-limits.

When posed with taboo tradeoffs, people instead adhere to deontological constraints, affirming their culture's proscriptive moral rules.

When choices are viewed as moral (because they involve sacred values), stringent norms that classify types of tradeoffs as morally permissible and impermissible are invoked. Sometimes, people who either lack these norms or do not cringe at infringing them propose trade-offs between sacred values and money because, they argue, the practice would have good consequences. Such proposed tradeoffs of sacred values for money (e.g., auctioning of body parts, selling pollution credits, selling futures betting on acts of terrorism) are met with strong disapproval (Tetlock et al., 2000). In some cases, these tradeoffs are met with moral outrage and an outright refusal to consider costs and benefits (Tetlock, 2002, 2003). Even knowing that a third party *merely contemplated* such a tradeoff is aversive; eliciting contempt, disgust, and the judgment that such contemplation is unforgivable from participants. "[Taboo tradeoffs] are… morally corrosive: the longer one contemplates indecent proposals, the more irreparably one compromises one's moral identity. To compare is to destroy." (Tetlock et al., 2000).

According to the sacred values framework, the application of utilitarian decision principles for taboo tradeoffs is impermissible, but for other types of tradeoffs, utilitarianism is more likely to be viewed as permissible. In situations where *only* morally significant goods are at issue the space of permissible cognition and decision principles expands. In these situations, people think it is okay, maybe even a good idea, to weigh utilitarian considerations. For example, in one study, participants read about a hospital administrator agonizing over a decision about which of two dying children should be given the one life-saving organ available (Tetlock et al., 2000). Judgments rendered about protagonists facing such "*tragic tradeoffs*" — however he or she ultimately chooses — are more moderate. Third parties' contemplation

of tragic tradeoffs elicits sadness from participants, and far more forgiving assessments of the protagonists' decisions (Tetlock et al., 2000).

The sacred values framework thus links moral values (as internalized norms that afford certain goods moral significance) to deontology in some contexts: utilitarian considerations are off-limits when contemplating a sacred-for-secular exchange. The sentiment is that some goods are not exchangeable for money, no matter what good might be brought about by the exchange. This framework also identifies situations where consequentialist cognition is permissible (as in tragic tradeoffs, like the transplant example and the ethical dilemmas constructed by philosophers, and secular-secular tradeoffs, like purchasing a computer).

1.6.2. Values as very strict deontological constraints: The protected values framework

One major motivator of the current studies is the protected value (PV) framework (Baron & Spranca, 1997). This framework describes morally-motivated choice as constrained by an *even more restrictive* set of tradeoff rules. Protected values, like sacred values, are protected from tradeoffs with other values. But PVs are defined as exhibiting *absolute* tradeoff resistance — in theory, they cannot be traded off or harmed for any reason, no matter what the consequences are that favor the tradeoff or harm.

Thus, for scenarios entailing the exchange of a resource for which people have a PV, people are expected to reason differently (i.e., they make use of moral rules) than when reasoning about less moralized resources. This is because PVs are thought to be associated with deontological rules — rules that concern actions (e.g., ‘‘do no harm’’; Baron, 1996), but not the overall consequences of those actions.

In this thesis, I focus more attention on the PV framework than the sacred values framework because it offers a straightforward way to select situations in which deontological decision strategies should predominate over utilitarian ones. Note that the sacred values framework only picks out situations in which utilitarian cognition is *permissible*, but it does not necessarily make predictions about how people ultimately will judge or decide in those situations. For example, both the bystander and footbridge cases described earlier would qualify as "tragic tradeoffs" — both involve the very same sacred value (i.e., human life). While it may be that utilitarian cognition is *permissible* for both cases, people render deontology-consistent judgment for the footbridge case.

The predictions of the PV framework are straightforward, in part, because the very way PVs are measured suggests they are non-consequentialist constructs akin to deontological constraints. Whether a participant has a PV for a given resource is typically assessed by presenting him or her with statements concerning the acceptability of tradeoffs for some resource, 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.*

This dichotomous measure classifies participants as having a PV for the resource if they respond "c" (Ritov & Baron, 1997). Recall the previous scenario, where weighing utilitarian considerations (saving 20 fish species instead of two) mandates that you perform a harmful action (open the dam and kill the two species downstream). Participants who endorsed PVs

were especially reluctant to *cause any harm* by opening the dam — an effect Baron and colleagues refer to as "omission bias" (Spranca, Minsk, & Baron, 1991) — suggesting that they were less sensitive to quantity (Ritov & Baron, 1999).

These results linking moral preference to non-consequentialism are important to the field of judgment and decision making for at least two reasons. First, as I discussed earlier, the field adopted consequentialist models to describe how people should choose (e.g., Savage, 1954) and how they actually do choose (Kahneman & Tversky, 1979). For these theories, quantity sensitivity is simply part of rationality: more of a good thing is preferable to less, other things being equal. That PVs are associated with quantity *insensitivity* calls into question the explanatory power of the prevailing (consequentialist) descriptive models of choice. The second reason is that, as I noted earlier, the people who we might expect to care most about outcomes (i.e., those people whose choices should be most consistent with utilitarianism) are the people who appear least sensitive to outcomes.

A *complete* lack of concern for the consequences about which one cares the most seems implausible. For example, consider a parent who is reluctant to vaccinate her daughter because she worries that she might harm her daughter with the vaccine (e.g., that her daughter might suffer from improbable side effects). If the mother's focus is shifted to the much greater risk posed to her daughter by not vaccinating, the parent may now feel a very strong (moral) imperative to vaccinate her daughter. The PV framework argues that constraints like "do no harm" prevent people from giving the "appropriate" weight to the consequences favoring a potentially harmful action — in this case, the risk mitigated by vaccination (Ritov & Baron, 1999).

The protected values framework has tested and found support for a number of provocative principles governing moral choice. However, a reader of this literature might conclude, erroneously, that value-driven choice is rigid, non-consequentialist, and error-prone — that morally-motivated decision makers are rigid deontologists who do not *or cannot* attend to consequences. For example, after spending a few paragraphs 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."

Part One of this thesis shows that to expect non-consequentialism — or indeed any *one* pattern of results — from morally-motivated decision makers is to underappreciate how malleable moral choice really is. The first two studies focus on this issue and show that in some contexts, moral choice is *more flexible* than non-moral choice; it can be both more non-utilitarian *and more utilitarian* than non-moral choice.

1.7. General structure of the thesis

The studies reported in this thesis show that morally-motivated choice is context-sensitive in a number of respects, and it treats these demonstrations of context-sensitivity as diagnostic of a number of cognitive processes. Each of the studies examines factors that push

people's moral sentiments toward or away from utilitarianism (the primary standard of comparison adopted for these studies). Study 1 focuses on the paradox identified above: why, and under what circumstances, people exhibit non-utilitarian preference in situations where we might expect them to care the most about the consequences of an action.

Study 2 examines situations in which we might expect *only* utilitarian preference, because utilitarianism and deontology are not brought into conflict. In contrast to ethical dilemmas where someone must be killed to save others, a simple choice between saving 60 of 240 lives at risk or saving 50 of 100 seems less conflict-ridden. However, people do not demonstrate exclusively utilitarian preferences in these situations. Sometimes, they prefer saving *fewer*, rather than more, lives (Bartels, 2006) — a pattern of preferences that is decidedly at odds with the utilitarian prescription to count lives saved and choose accordingly. Study 2 is motivated, in part, by a prominent contemporary utilitarian theory that makes predictions about ways to promote or undermine an important evaluation process on which utilitarian judgment is predicated.

The first two studies find that when choices involve moral values, preference is not as rigid or non-utilitarian as previously supposed, but is instead more malleable than non-moral preference. Morally-motivated decision makers show greater effects of experimental manipulations that target attentional processes (Study 1) and representational processes (Study 2) underlying preference formation.

Part Two of the thesis examines the interacting influences of a number of competing processes assumed to be at work when people are reasoning about ethical dilemmas. These two studies argue that judgment and decision making in these contexts is predicated on moral rules, emotional reactions, and assessments of costs and benefits. They suggest that contexts

that direct attention to violations of moral rules generate deontology-consistent emotional reactions, but also that deontological response can be outweighed by contextual factors that direct attention to utilitarian considerations.

These studies aid in the development of a better process-based framework for understanding one important kind of judgment and decision making, addressing some deficiencies in the literature on moral judgment and choice along the way. Behavioral decision research, in particular, has often been chastised for being more interested in effects than in psychological process. An unfortunate choice of approaches employed by some researchers is to identify systematic departures from normative or optimal responses, label people's responses "errors" or "biases," and leave the discussion there.

The studies in this thesis take a decidedly different approach. This thesis brackets normative issues and aims to explain moral decision making in terms of psychological processes. This work demonstrates the powerful influence of moral values on judgment and preference by relating processes implicated in philosophical ethics and studies of people's responses to ethical dilemmas to behavioral decision theory and vice versa.

2 Study 1—Protected values and utilitarian quantity sensitivity

2.1. Introduction

2.1.1. Protected values as deontological constraints

PV-driven choice is often depicted as unique because of its apparent rigid adherence to deontological constraints. The suggestion is that for decisions involving cherished resources (for which people have a PV), people decide differently (i.e., make use of moral rules) than for decisions about resources not tied to their moral values. And, in fact, participants endorse a PV by affirming a non-consequentialist, deontological constraint — that doing harm to their cherished resource is forbidden, no matter how great the benefits. So, the expectation that PVs should exert a main effect, producing a single pattern of (non-consequentialist, constraint-affirming) preferences, seems straightforward enough.

But can it be that PVs are supposed to play out in exactly the same way across a variety of choice contexts? The answer here is far from clear, in part because PVs exhibit their defining property — absolute tradeoff resistance — only in theory. Recall that PVs are cannot be traded off, not even for more of the same thing (c.f., Tetlock et al.'s, 2000 "tragic tradeoffs"). Baron and others note, however, that people implicitly trade off these cherished resources on a constant basis. Valuing human life above all else and not accepting any tradeoffs, for example, would entail a variety of impossibly expensive public policies. As an extreme example, such a stance might dictate a national speed limit of 15 miles per hour and a requirement that cars be made out of Nerf.

In addition, Baron and colleagues have shown empirically that PVs can be overwhelmed by utilitarian considerations. Baron and Leshner (2000) show that participants with PVs are

sometimes willing to accept some minuscule harm if it yields massive benefits. Baron and Leshner (2000) characterize PVs as errors, but present no psychological-process-informed account of how these "errors" are to be corrected. So, we are left with no theory specifying the conditions under which we should expect people to accept tradeoffs to PVs. Nor is there an adequate explanation of how utilitarian features can acquire the moral affordance to overwhelm PVs. (This is not a problem unique to this paradigm, however. Philosophy has been debating for decades about why the utilitarian solution seems appropriate in some cases and not others.)

Baron and Leshner (2000) seem satisfied to conclude that "PVs are strong opinions, weakly held" (p. 193). Also, because they demonstrate that some PVs are "amenable to challenge" (p. 183), they characterize PVs as "unreflective overgeneralizations" (p. 184).[9] This type of rhetoric is similar to that which is found in utilitarian philosophy (e.g., Hare, 1981; Unger, 1996), in part, because the PV framework accepts utilitarianism as the paradigm of genuinely *rational* judgment and choice.

Each of the studies that follow share a focus on utilitarian judgment and decision principles, but intend to characterize moral judgment and choice in terms of its cognitive underpinnings and not its normative status. Identifying an effect (in the present context, an

[9] One might be tempted to interpret Baron and Leshner's findings as evidence that PVs do not *really* exist, or to assume that what is typically measured in these studies is a pseudo-PV. This thesis will present findings where PVs are upheld and where they are overwhelmed by utilitarian considerations, but they explain this flexibility in terms of the processes that make use of PVs, as opposed to the more problematic task of asserting ontological claims. In addition to this concern, there are reasons to believe that using a single item to index the presence or absence of a moral issue is methodologically unsound. In fact, in one study I ran, participants responded to a battery of 23 PV items and then responded to the same set of items later in the academic quarter. Endorsement rates for each item were compared across time periods. The average correlation between responses for a given item at time 1 and time 2 was only .35. Considering all the reasons why we might not expect PVs to influence judgment and preference, the effects attributed to PVs in this thesis are all the more surprising.

apparent unwillingness to make tradeoffs) and labeling it an “error” or “fallacy” is not the same thing as providing a process-based account of it.

Values, even very important ones, may seem labile because of the many processes they influence and by which they are influenced. For example, Fischhoff (1991; Fischhoff et al., 1980) cogently argues that although people have a set of values, responses to value-eliciting probes are more likely to be computed than retrieved. The computational component produces the context-sensitivity that makes up the major focus of Studies 1 and 2 in this thesis (and, more generally for Fischhoff, introduces discrepancies across probes and between stated and revealed preference). Each study in this thesis demonstrates that moral preference is malleable, rather than rigid, and explains the malleability in terms of psychological process.

2.1.2. The omission bias paradigm

The literature focuses less attention on the processes PVs motivate and more attention on policy ramifications and defining the construct. For example, the literature is relatively silent on the attentional processes involved in PV-driven judgment and preference, saying little about the cues that evoke these constraints or how they might be overwhelmed. The focus, instead, has been directed toward a number of testable hypotheses implied by the definition of the construct. Recall that PVs are associated with, indeed practically defined in terms of, trade-off avoidance.

Rigid tradeoff avoidance is thought to drive patterns of preference that reflect two non-consequentialist properties: *omission bias* and *quantity insensitivity* (Baron & Spranca, 1997). Omission bias is a preference for indirect harm caused by omissions (i.e., failures to act) over equal or lesser harm caused by acts (e.g., Spranca, Minsk, & Baron, 1991; Ritov & Baron, 1992, 1995). Because deontological constraints can be formulated as very narrowly-directed and

defined moral prohibitions, lying (an explicit, overt act) and failing to tell the truth (an omission) are not viewed the same way. PVs reflect this aspect of deontological constraints; Baron and his colleagues have amassed evidence that PVs are associated with a large omission bias (Ritov & Baron, 1990, 1999). Their paradigm typically involves presenting participants with problems like dam example mentioned earlier and included below (Ritov & Baron, 1999):

> *As a result of a dam on a river, 20 species of fish are threatened with extinction. By opening the dam for a month each year, you can save these species, but 2 species downstream will become extinct because of the changing water level.*
>
> *Would you open the dam? Yes/No*
>
> *What is the largest number of species made extinct by the opening at which you would open the dam?*_____

Participants with PVs are less likely to open the dam, and some even say they would not want to cause a single species to die. This response affirms their constraint but produces the worst consequences possible for each of the scenarios tested. Ritov and Baron (1999) use the last of the questions above to calculate a continuous index of a participant's "quantity sensitivity": the value supplied is divided by the risk associated with omission (in this case, 20), yielding an index ranging from zero to one. The smaller this threshold value, the less "quantity sensitive" a participant is said to be. Low thresholds are interpreted as reflecting relative insensitivity to the consequences of one's choices. In other words, higher thresholds are more consistent with utilitarianism, quantity insensitivity, as indexed by relatively lower thresholds, is more consistent with adherence to deontological constraints.

Participants with PVs — those who (in this example) judge the extinction of fish species unacceptable no matter what the consequences — provide lower threshold values than participants without PVs, which indicates that PV-driven preference is quantity insensitive

(Ritov & Baron, 1999). This result seems intuitive if not circular: people indicate unwillingness to trade off on two different measures, one abstract and one contextualized.

If we are safe to assume that people have PVs for the resources about which they care the most (e.g., family), why do their choices show apparent disregard for the consequences realized by those resources? As I have already suggested, the idea that people refuse to consider the consequences they care about most is probably an over-generalization. And, in fact, Baron and Leshner (2000) have already shown that people are willing to accept some tradeoffs of PVs.

Study 1 starts with the idea that in domains where people indicate protected values, we should expect relatively greater sensitivity to both violations of moral rules *and* to utilitarian considerations. The current study shows that judgment and preference influenced by moral values can be responsive either to utilitarian or to deontological considerations, depending on which types of considerations are made salient by the context.

2.1.3. An alternative paradigm

There are reasons to believe that omission bias and quantity (in)sensitivity are malleable, able to be switched on or off with the judgment context. Recently, Connolly and Reb (2003) examined the effects of modest changes to the omission-bias paradigm. In their Study 2, they varied the risks associated with acts and omissions in a repeated measures design. Consider an adapted version of the previous scenario:

> *As a result of a dam on a river, 20 species of fish are threatened with extinction. By opening the dam for a month each year, you can save these species, but some species downstream will become extinct because of the changing water level.*
>
> *Would you open the dam if it would kill 2 species of fish downstream as a result? Yes/No*
> *Would you open the dam if it would kill 6 species of fish downstream as a result? Yes/No*
> *Would you open the dam if it would kill 10 species of fish downstream as a result? Yes/No*

Would you open the dam if it would kill 14 species of fish downstream as a result? Yes/No
Would you open the dam if it would kill 18 species of fish downstream as a result? Yes/No

Rather than asking participants to generate their own threshold value, this method provides them with a range of options. It also does not begin with an anchor. Using this alternative paradigm, Connolly and Reb examined decisions concerning whether or not to vaccinate (the vaccine sometimes had bad side effects) and found no evidence for omission bias. Although there has been debate concerning the relative complexity and merits of Ritov and Baron's procedure (1999) and Connolly and Reb's procedure (Baron & Ritov, 2004; Connolly & Reb, 2004), I employed both methods as a means of clarifying the nature of PVs and their role in decision making.

2.1.4 Hypotheses

The theoretical notion guiding Study 1 is that PV-driven preference involves attentional processes and that the two procedures may influence attention differently. Specifically, Ritov and Baron's (1999) procedure may direct attention to the question of *whether one should knowingly do harm* to a cherished resource. Violations of deontological constraints (if PVs are deontological constraints) are usually judged impermissible. The RB procedure may engender this sentiment.

In contrast, in Connolly and Reb's (2003) procedure, participants are asked the same question at different levels of risk entailed by the action. This procedure may convey the presupposition that some tradeoff is expected, directing attention away from the tradeoff itself and directing participants' attention on balancing risks and consequences. Utilitarians treat deontological constraints as rule of thumb that must be broken if the consequences favor doing

so. Connolly and Reb's procedure may engender this sentiment. (However, because they did not assess PVs, it is unclear how people with PVs would respond in their paradigm.)

Consistent with the findings of Ritov and Baron (1999) participants with PVs are expected to exhibit less quantity sensitivity than participants without PVs. The RB condition is expected to direct attention to the permissibility of a harm-producing action — "Would you open the dam?" — leading to less consideration of the ultimate consequences of action, and thus, less quantity sensitivity for people possessing the deontological rules (PVs).

Participants with PVs are expected to exhibit *greater* quantity sensitivity than participants without PVs in the CR condition. The CR condition is expected to direct attention to the ultimate consequences of action by varying the consequences of the action within-Ps. And, if people who endorse PVs care more than other people about the resource at risk, they should give greater consideration to consequences (and demonstrate more quantity sensitivity) than participants who care less about the consequences entailed in the scenario.

There has been enough research on PVs and decision making to establish that this domain is theoretically and practically rich, but there has been too little research aimed at establishing generality across paradigms and social contexts. A primary goal of Study 1was to contribute to the literature by examining the generality of results across two closely related procedures. This study assessed the relation between PVs and quantity sensitivity for three scenarios, using a replication of Ritov and Baron's (1999) procedure with some participants and a procedure inspired by Connolly and Reb's (2003) study with others.

In addition to examining response formats and quantity sensitivity, I assessed whether PVs exert domain-general or domain-specific influences by collecting responses for three additional, unrelated PVs. If quantity sensitivity or insensitivity is predicted by endorsing many

PVs, then the relation between PVs and quantity sensitivity may reflect individual differences in generalized deontology, rather than use of different reasoning processes depending on whether cherished or uncherished resources are at risk.

2.2. Method

2.2.1. Participants

Seventy-four undergraduates (44 women and 30 men) enrolled in an introductory psychology course participated. Each completed the questionnaire at his or her own pace. They were tested individually but in a small-group setting (typically one to four participants per session). Usually, other participants were completing the study at their own pace in the same room. All received partial course credit (completing 30 minutes of their required 10 hours of research participation to receive full credit for the course).

2.2.2. Materials and design

After reading the instructions, participants were asked to read and respond to three scenarios from Ritov and Baron (1999): River Diversion (given earlier), Starvation, and Cutting Forests. The latter two scenarios were worded as follows:

> *Starvation. A convoy of food trucks is on its way to a refugee camp during a famine in Africa. (Airplanes cannot be used.) You find that a second camp has even more refugees. If you tell the convoy to go to the second camp instead of the first, you will save 1000 people from death, but 100 people in the first camp will die as a result.*

> *Cutting Forests. A logging company has the rights to 1000 square miles of old-growth forest. The company is willing to trade this land for 100 square miles of similar land, now part of a national park. You can give the smaller area to the company and make*

the larger area into a national park. The trees and scenery in the two areas are much the same. The logging company will cut all the trees in whichever area it owns.

The three scenarios were included in random order within a packet. Two versions of the questionnaire were constructed. Half of the sample received the ‘‘RB’’ version, which used the items and procedure from Ritov and Baron (1999). The other half received the ‘‘CR’’ version (modeled after Connolly & Reb, 2003), in which participants were not asked for a threshold value, but instead were asked whether or not they would act if acting entailed 10%, 30%, 50%, 70% and 90% of the risk entailed by the omission.

After responding to the three scenarios, participants’ PVs for six items from Baron and Spranca (1997) — three corresponding to the scenarios and three unrelated items — were assessed. The additional actions participants judged as acceptable or unacceptable were the following: ‘‘*Selling products for profit made by strike breakers*,’’ ‘‘*Putting people in jail for expressing nonviolent political views*,’’ and ‘‘*Aborting normal fetuses in the last three months of pregnancy*.’’ Again, participants who endorse “this is not acceptable no matter how great the benefits” are counted as having a PV for an issue.

2.3. Results

2.3.1. By item

For the RB procedure, each threshold value — the highest level of harm caused by action that a participant viewed as permissible — was converted to a proportion by dividing this value by the harm caused by omission. This proportion is taken as a measure of quantity (in)sensitivity: the higher the value, the more sensitive to quantity (i.e., utilitarian) participants appear to be; the lower the value, the less quantity sensitive the participant is.

Similarly, the highest level of harm caused by action that a participant endorsed for each item was used to index quantity sensitivity in the CR procedure (values ranged from 0 to .9). If a participant circled zero "Y" responses, it was coded as 0. Levels of quantity sensitivity and their relation to PVs were compared across paradigms.

As expected, the threshold results in the RB procedure replicated those of Ritov and Baron (1999). Participants with PVs showed less quantity sensitivity than participants without PVs, providing lower threshold values (see Table 1). This difference was evident for each of the three items used, but reliable for only two. The analyses for the Starvation item had a lack of power (because so many of the participants had a PV for this item).

Strikingly, but as predicted, the opposite pattern was observed in the CR condition. Participants with PVs demonstrated *greater* quantity sensitivity than participants without PVs, providing higher thresholds (see Table 1). Again, this difference was evident for each item, but not reliable for the Starvation item. The implication is that the CR method — intended to highlights the ultimate consequences of action — induces people with PVs to appear *more* utilitarian than people without PVs.

2.3.2. Individual differences

A second set of analyses examined the relationship between the number of PVs endorsed (i.e., one, two, or three of the three relevant and three irrelevant items) and the average level of quantity sensitivity exhibited across all three items for each participant. Analyses for the three relevant items mirror the by-item analyses above: the more PVs a participant endorsed in the RB version, the less sensitive he or she was to quantity ($r(35) = -.57$, $p < .001$); conversely, the more PVs a participant endorsed in the CR version, the *more*

sensitivity to quantity was expressed ($r(35) = .38$, $p = .02$). Endorsement of the three irrelevant items correlates only moderately with quantity sensitivity (r's(35) = .22 and .23, *n.s.*), suggesting the effects observed in Study 1 were more attributable to the presence of specific moral rules than to general individual differences in deontological tendency.

2.4. Discussion

Whereas previous work on PVs portrayed morally-motivated decision makers as rigid deontologists who tend *not* to attend to consequences, Study 1 suggests the opposite. Morally-motivated decision makers were *more* influenced by whether the preference-elicitation procedure directed their attention to utilitarian or deontological considerations. The two paradigms – the Connolly/Reb and the Ritov/Baron – yielded diametrically opposing results. As assessed by the RB paradigm, people endorsing PVs appear less quantity sensitive than people not endorsing PVs, but as assessed by the CR paradigm, they appear *more* quantity sensitive than people not endorsing PVs.

The RB procedure appears to direct attention to an action that violates a moral rule (held by those endorsing PVs, absent for those without PVs). In other words, the distinction made salient by this procedure is between *doing harm* and *allowing harm* – a discrete, binary distinction that deontologists invest with considerable normative significance. In contrast, the CR procedure appears to direct attention towards net benefits, which utilitarians argue is the only sensible basis for moral choice.

Because all of the questions for each scenario ask for judgments about tradeoffs, participants may interpret the CR procedure as suggesting that a tradeoff is appropriate. Conversely, the RB procedure may focus the participant's attention toward the acceptability or

unacceptability of the tradeoff itself. By analogy, although one may be reluctant to sell an heirloom at any price, if one decides to sell it, the same respect for the heirloom now may demand that one get the best price possible.[10]

So, what should we make of protected values? There was good reason to expect an association between PVs and non-consequentialist, deontological decision principles, and there still is. Having a PV means endorsing the statement "this is not acceptable no matter how great the benefits." The measurement itself gives one good reason to think that people who endorse a PV should care less about utilitarian considerations (i.e., costs and benefits) than people who do not. Also, the evidence linking PVs to a large omission bias is consistent with a commitment to a kind of rigid deontological position. A preference for omission over action in contexts where action produces better consequences suggests these constraints are not treated like a utilitarian would view them — as rules of thumb that can be permissibly infringed in these cases.

Are people absolutist deontologists for domains governed by PVs? Probably not. PVs tend to be about particularly important issues — those issues where it seems like consequences should matter most. For example, most people likely bristle at the idea of harming a family member, finding it more offensive than the idea of harming a stranger. But it is also likely that most people care more about the ultimate outcomes realized by a family member than the outcomes realized by a stranger. The context-sensitivity revealed in Study 1 seems reasonable

[10] I thank Daniel Kahneman (personal communication, February 13, 2006) for this example.

if we assume the people who care more about not harming the resource also care more about the ultimate consequences of an action in those moralized domains.[11]

Study 1 is a preliminary demonstration that morally-motivated preference is malleable, that attentional processes exert a powerful influence on moral choice. When attention is directed to utilitarian considerations, they are more heavily weighted in determining choice than when attention is directed to deontological considerations. (Of course, other processes contribute, too. Studies 3 and 4 offer a more complex account of how attentional processes interact with affectively-charged intuition, and "colder" deliberation in preference formation.) Similarly, Study 2 tests whether moral preference is more malleable than non-moral preference, but Study 2 targets the influence of representational, rather than attentional, processes in utilitarian and non-utilitarian preference formation.

[11] Study 1's results might be interpreted as offering qualitative support for the sacred values framework (Tetlock, 2002), which argues that the application of consequentialist decision principles is viewed as impermissible *only* for "taboo" (secular-for-sacred) tradeoffs. Each choice presented to participants in Study 1 involved only a single resource. The same is true for each study in this thesis. Consistent with theorizing about "tragic" (sacred-for-sacred) tradeoffs, participants thought it was *okay* to trade off some of a PV for more of a PV *in some cases*. And, to the extent that the sacred values framework allows for greater flexibility in moral cognition than the PV framework, demonstrations of flexibility might be interpreted as lending support to the sacred values framework. However, because participants were not presented with taboo tradeoffs in the current studies, no pattern of results could refute the predictions of the sacred values framework. Therefore, none of these studies should be considered a *targeted* test of its descriptive adequacy.

3 Study 2—Proportion dominance and futility thinking in moral judgment

3.1. Introduction

Studies 1, 3, and 4 all pose a simple question about non-utilitarian preference: Why would someone's judgments and choices disregard the consequences he or she cares about most? A related, but somewhat different question addressed by Studies 1, 3, and 4 is "Why do people refuse to do things that will cause harm, even when as a result of not acting, *more* harm will result?" Study 1 suggests a partial answer to these questions: when attention is focused on the question of whether or not to sacrifice some lives (i.e., to do harm) to save others, morally-motivated preference seems deontological. Results also demonstrate that when attention is directed to the others-to-be-saved, morally-motivated preference is more consistent with utilitarianism.

What about when the choice *only* concerns others-to-be-saved and *not* whether to harm? Study 2 focuses on the question of *which* lives to save in *non-sacrificial* contexts. Based on Study 1's findings, we might expect that removing harm-doing from the equation should promote *rigidly utilitarian* preference for everyone (because utilitarianism and deontology are not brought into conflict). Because utilitarianism requires summing the satisfaction of welfare interests (i.e., basic needs) across individuals, counting each equally, its prescription — that decision makers simply count the others-to-be-saved and choose — seems especially apposite for these contexts. But, expecting any one pattern of morally-motivated preference, even for these contexts, turns out to be a mistake.

Some argue that understanding morally-motivated judgment and decision making in these contexts is of utmost importance due to issues pertaining to ecological validity. "As utilitarians

tirelessly and rightly point out, very rarely should ordinary agents (as opposed to trolley operators) think they can produce large net benefits only by harming innocent others." (Hooker, 1999, p. 177)

So, the current study begins with a different — but equally morally-weighty — question about non-utilitarian decision making in non-sacrificial contexts: "Why would people ever choose to save *fewer*, rather than more, lives?" In the present context, when posed with a choice between saving 60 of 240 lives at risk or saving 50 of 100, why would people ever choose the latter?

A partial answer is that when people view these situations as a choice between saving 25% of a group versus saving 50% of a group, their preferences depart from the prescriptions of utilitarianism (that *individuals* be counted). In these contexts, the moral intuitions and resultant behavior triggered by individual (save 60 or 50) versus group construal (save 25% or 50%) are *widely* divergent. This point is most forcibly made in a quote attributed to Mother Teresa: "If I look at the mass I will never act. If I look at the one, I will."

3.1.1. Proportion dominance

Choices, whether moral or non-moral, often lend themselves to description in both relative and absolute terms. In the marketplace, for example, an offer may be made in the form of a $50 rebate or 10% off. Similarly, public policies often involve costs and benefits that can be viewed in absolute or relative terms e.g., saving 450 lives versus saving 3% of 15,000 people who suffer from a certain disease.

Even in contexts where a utilitarian focus on absolute numbers might seem appropriate, decisions are often influenced by relative considerations. In one study, for example, participants

evaluated a program that would save the lives of two pedestrians annually at a Pittsburgh intersection (Jenni & Loewenstein, 1997). For one group of participants, the pedestrians were described as 2 of 4 people who die at that intersection annually. For a second group, they were described as 2 of 1,700 people who die in auto-related accidents in Pennsylvania annually. The first group evaluated the program more favorably. The program's consequences are identical in both cases, and utilitarians argue that all that matters is that 2 lives are saved (Unger, 1996), but relative considerations — the proportion of the reference group saved — make the first description more compelling. Other experiments employing similar between-participants designs have revealed similar effects (Baron, 1997; Fetherstonhaugh, Slovic, Johnson, & Friedrich, 1997; Friedrich et al., 1999).

Relative considerations sometimes trump absolute considerations even when they are pitted against each other. For example, participants in Bartels (2006) read that anthrax had been weaponized and released into the air above two cities. They then chose between saving 225 of the 300 people expected to die in one city versus saving 230 of the 920 expected to die the other city. Nearly half of participants preferred the first option — saving a greater proportion, even though this meant saving fewer lives. Following Slovic et al. (2002), I call this phenomenon *proportion dominance* (PD). Further, while on reflection most participants felt that absolute savings should be maximized (at the expense of relative savings), they endorsed this strategy more strongly for problems concerning human life than for other problems (e.g., involving sea otters).

3.1.2. Futility thinking

Previous research on proportion dominance has investigated simple preference, making little connection to research on moral reasoning and judgment. This seems like a major oversight, considering that the resources under consideration (e.g., human life, natural resources) are typically drawn from domains that are ascribed moral relevance by many people. Baron (1997) first pointed out that effects like proportion dominance are discussed by a prominent utilitarian ethical theory proposed by Unger (1996).

Unger argues that moral intuitions are subject to a fallacious "futility thinking" — for example, saving lives is considered less obligatory when they are construed as a few among overwhelmingly many at risk. Unger argues that futility thinking underlies the common intuition that letting a child drown in a nearby pond is less permissible than letting a child die of malnutrition in a famine-struck country. Unger's book *Living High and Letting Die: Our Illusion of Innocence* provides partial motivation for the hypotheses tested in Study 2. Unger writes:

"Usually, you're very much in the grip of a doubly misleading sort of moral thinking, fallacious futility thinking. On one side of this habitually confusing coin, when so gripped you're greatly influenced by a consideration that's morally irrelevant: the vastness of the serious losses that will be suffered even after you do your utmost to lessen such suffering. And, on the other side, when so gripped you're only slightly influenced, or perhaps even influenced not at all, by a consideration that's morally weighty: the lessening in serious suffering you can effect. So it is that, usually, you erroneously think that, since you can make only a small dent in the vast mass of all the serious suffering, there's no strong moral reason for you to... give what's your own, to lessen the serious suffering. So, these badly misleading factors serve to promote fallacious forms of moral thinking, prominently including what I'll call futility thinking." (p. 63)

Unger describes five components of the psychological process by which futility thinking undercuts utilitarian moral intuition. Especially relevant for the current study is the non-

utilitarian moral intuition (fourth component), which is triggered by group construal (third component):

"We take futility thinking's pattern to have these five steps: (1) First, when *all* you know is that others are in great need, you (correctly) think there's *strong* moral reason for you to help them… (2) Second, often you also know that no matter *what* you do, very many of the greatly needy people *still* won't have their needs met, and, so, *they'll all suffer anyway.* (3) Third, often your most *powerfully operative* thought about the needy is, then, one that *primarily presents* the individuals as *members of just such a hopelessly overwhelming group.* (4) Fourth, even behavior that's *successful* in meeting some of the people's great needs then *seems to be futile*; it seems like successfully removing a mere drop of trouble from a whole sea of suffering. (5) Fifth, and finally, you (incorrectly) think there *isn't* strong moral reason for you to help meet any of the great needs." (pp. 74-75, emphasis in original)

Unger also considers factors that work against futility thinking. One involves presenting a life as belonging to a smaller reference group, perhaps a group of one:

"…we see how there may be something to upset the pattern just roughly related: When some folks' needs are *highly conspicuous to you*, often you're liberated from futility thinking's grip. And, when that occurs, often *your most powerfully operative thought* primarily presents *those* folks as *greatly needy people*, not as members of such an overwhelming group. And, when that happens, behavior that meets their great needs *doesn't* seem futile. And, then, your last thought about the matter is just the same as your first: You (correctly) think there's *strong* moral reason for you to help them." (p. 75, emphasis in original)

Here I adopt an approach to understanding proportion dominance, motivated, in part, by Unger's speculation about processes underlying futility thinking. For Unger, utilitarian thinking is undercut by deindividuating the people at risk, making one's efforts to alleviate serious suffering seem futile. The current study uses individuation and deindividuation to examine the processes at work in proportion dominance. I hypothesize that people construct mental

representations of choice scenarios in which resources can be construed more or less as forming groups and that group construal promotes proportion dominance. Proportion dominance follows naturally from group construal: In the extreme case where distinctions among individuals are overridden entirely and groups are units with no internal structure, relative considerations are the natural (indeed, the only) basis for choice or evaluation.[12] This approach also provides one intuitive explanation of the content-domain differences found by Bartels (2006): participants tended to construe humans as individuals and therefore focused on absolute considerations, whereas they construed nonhumans as more group-like and therefore gave greater weight to relative considerations.

The current experiment tests this approach. The goal was to manipulate participants' construal of resources on the individual-group dimension and to assess the effect on the weights given to relative and absolute considerations on people's moral judgment.

3.1.3 Concepts and category extension

To manipulate construal, I adopted a method from studies that investigated conditions under which adults and children treat groups as single units. In one study (Bloom, 1996; Bloom & Kelemen, 1995), participants were shown a static display of 15 unfamiliar-looking objects arranged in 3 groups of 5 and were told, for example, "these are fendles." Participants tended to interpret the name as referring to the objects, and when asked, reported that there were 15 fendles. In another condition, however, the 3 groups moved as units; each group followed a distinct path across the display. In this condition, participants interpreted the novel name as

[12] And indeed, group construal can exert effects elsewhere. In a study examining the effects of interdependence on choice under uncertainty, for example, Bloomfield et al., 2006 report a framing effects study showing that risk preference changes when groups are viewed as meaningful units.

referring to the groups, reporting that there were 3 fendles. In another study (Bloom & Veres, 1999), participants saw groups of objects moving along distinct paths and interacting with one another. When asked to describe these animations, participants described the groups, not the objects, as individual agents with intentions to move in certain ways. Joint motion, then, is an effective cue to "groupness."

In Study 2, participants saw resources (people, otters, etc.) depicted as arrays of objects. These objects came into view via computer-presented animations. In the *individuals* condition, objects emerged from different, randomly chosen off-screen locations and followed independent paths to their final locations in the array. In the *groups* condition, objects moved in concert. These animations were accompanied by verbal descriptions of scenarios in which absolute and relative considerations were pitted against each other, and participants rated their preference for one alternative or the other.

I expected group construal to undercut utilitarian intuition in a way that is similar to what Unger (1996) suggests. I expected greater proportion dominance in the group condition than in the individuals condition. That is, participants viewing joint motion would judge maximizing proportion saved (at the expense of absolute number saved) as morally better than participants in the individuals condition.

I also used a different measure to assess whether participants viewed the choice they faced as moral than the one employed in Study 1 (i.e., endorsement of PVs). In Study 2, I measured participants' moral conviction — a straightforward measure where participants indicate the degree to which their attitudes about the topic are distinctly moral — for each of the resources under consideration. I anticipated that these predicted effects on moral judgment

would be *more pronounced* for participants who ascribe *more moral relevance* to the choice situation than for participants who ascribe less moral relevance to the situation.

3.1.4. Hypotheses

Participants in the group condition are expected to exhibit greater proportion dominance — exhibit greater preference for saving a larger proportion (e.g., 22 of 26) over saving a larger number, but smaller proportion (e.g., 24 of 176) — than participants in the individuals condition. The group condition is expected to induce more futility, or "drop in the bucket," thinking precisely because it induces participants to attend to the size of the bucket, meaning each individual drop is afforded less weight in judgment of the act's moral status. In other words, participants in the group condition will be more likely to treat groups — and not individuals — as the relevant unit of analysis, and so they should exhibit greater proportion dominance.

Moral conviction is expected to moderate this effect, interacting with individual versus group construal in shaping preference. Participants who ascribe *less* moral relevance to a given issue will show *smaller* effects of group versus individual construal because the already low weight they afforded to the individuals (the "drops") is less likely to be undermined by deindividuation and is less likely to be promoted by individuation.

3.2. Method

3.2.1. Participants

Fifty undergraduates enrolled in an introductory psychology course participated in the main experiment, and 115 undergraduates enrolled in an introductory psychology course (taught during a previous academic quarter) participated in the manipulation check. All participants were tested individually. Usually, other participants were completing the study at

their own pace in the same room. All received partial course credit (completing 30 minutes of their required 10 hours of research participation to receive full credit for the course).

3.2.2. Materials and design: Main experiment

The experiment was administered by computer. After some initial instructions, participants advanced to a screen where they read a scenario posing a tradeoff between relative and absolute savings. The scenario involved lives or other resources at risk, and two alternatives were described: one saving a larger number of individuals and another saving a larger proportion of an at-risk group. For example:

> *Recent developments in Zaire have marginalized a significant minority of the population. These refugees are clustered in two camps, struggling to survive because little clean water is available. A plane with water treatment capabilities will be sent. This treatment plane will save a number of refugees, but there is only enough fuel and time to visit one camp. Program A would treat enough water to save 22 refugees in a camp of 26. Program B would treat enough water to save 24 refugees in a camp of 176. These programs are mutually exclusive and the only two options available.*

Participants then advanced to a screen where the experimental manipulation took place. Elements appeared on this screen in the following sequence.

(1) On the left side of the screen, a frame labeled “Program A” appeared. Gray objects representing Program A’s reference group appeared in this frame. For example, if Program A would save 22 of 26 people, then 26 stick figures appeared. In the individuals condition, these 26 figures followed distinct paths from locations around the edges of the frame (see Individuals A in Figure 1) and assembled into a rows-and-columns array. In the groups

condition, the individuals moved together into the frame, like an army marching in formation (see Groups A). The final rows-and-columns arrangement was the same in both conditions.

(2) A description appeared (e.g., “Program A saves 22 of 26”) beneath the frame, followed by the text “To see this depicted, click on the figure above.” Participants had to click for the task to proceed, and when they did so, the resources lost (e.g., 7 figures) remained gray, while the resources saved (e.g., 22 figures) came into color.

(3) A frame labeled “Program B” appeared on the right side of the screen. Step 1 was then repeated for Program B. For example, if Program B saved 24 of 176 refugees, 176 figures appeared, by either independent or joint motion depending on condition (see Individuals B and Groups B).

(4) Step 2 was repeated for Program B.

(5) With the end-state depictions of Program A and Program B on screen, the dependent measure for Study 2 appeared at the bottom of the screen: “Choosing to implement Program A (instead of Program B).” Participants then registered their responses on this measure by moving a slider any point along a continuum with five partitioning labels: “Morally forbidden”, “Morally impermissible”, “Not a moral issue”, “Morally permissible”, and “Morally obligatory”, and clicking “Continue”. (see Screen C).

(6) Finally, participants were asked for the same judgment about “Choosing to implement Program B (instead of Program A)”

Importantly, only steps 1 and 3 differed between conditions. The end-state depictions that participants saw while registering preferences were identical in the two conditions. Consequently, any difference in preferences between conditions can be attributed to whether

the resources had moved jointly or independently. Such a difference would constitute strong evidence that people construct mental representations in which resources are more or less individuated and that degree of individuation influences proportion dominance.

There were five scenarios — two involving human lives and three involving nonhumans (otters, fish, dolphins) — presented in random order (stimuli are shown in Appendix A). Whether Program A maximized absolute savings and Program B maximized relative savings or vice versa was determined randomly for each trial. Each participant was randomly assigned to the individuals or group condition.

Ratings were scored as follows: Morally forbidden, Morally impermissible, Not a moral issue, Morally Permissible, and Morally Obligatory were scored -100, -50, 0, +50, and +100. So, participants' responses on the dependent measure could take any value between -100 and +100)

After responding to the five scenarios, a kind of individual differences variable was collected. Participants' moral conviction (Skitka, 2002; Skitka, Bauman, & Sargis, 2005) was assessed for the topics referenced in each the five scenarios. This measure was chosen because it is the most straightforward approach to assessing whether the participant treats a domain as moral or amoral. Participants were presented with a topic — "Threat of potential harm to (resource)" — and asked to rate agreement with the statement "My attitude about this topic is closely related to my core moral values and convictions" ranging from "Strongly disagree" (scored -100) to "Strongly agree" (scored +100).

Participants who indicated greater agreement (i.e., higher numbers, subsequently referred to as "high conviction" participants) were expected to show greater effects of individual versus group construal on moral judgment than "low conviction" participants. The

prediction is that greater moral conviction will lead to relatively greater proportion dominance in the groups condition and relatively less proportion dominance in the individuals condition.

3.2.3. Materials and design: Manipulation check

To assess whether independent versus joint motion induced individual versus group construal, a separate group of participants viewed the opening segment of an animation used in the main experiment (where resources emerged on screen), drawn from either the groups or individuals condition (between-subjects). As the animation looped, participants rated "the degree to which the people in this animation seem like individuals or like a group." The scale was continuous and was explained in this way: "A rating of –3 means that they are individuals with distinct identities. A rating of +3 means that they are a tight group with a single identity. A rating of 0 means that they are individuals and a group to equal degrees."

3.3. Results

3.3.1. Manipulation check

Joint motion promoted group construal. Participants who saw group motion rated the resources as more group-like (M = 1.84, SD = 1.13) than participants who saw independent motion (M = 0.86, SD = 1.37, $t(113) = 4.19$, $p < .001$, $\eta_p^2 = .13$).

3.3.2. Main experiment

For each item analysis, I computed a difference score, subtracting the rating given to the absolute number-maximizing action (e.g., save 24 of 176) from the rating given to the proportion-maximizing action (e.g., save 22 of 26) for each item. Since these ratings given to

each action ranged from -100 (morally forbidden) to +100 (morally obligatory), the resultant index of proportion dominance could take any value between +200 (strongest possible preference for maximizing proportion saved, rather than number saved) and -200 (strongest possible preference for maximizing number saved, rather than proportion saved). For the analysis of participant averages, I compared the average proportion dominance exhibited across items for each individual.

As predicted, participants in the individuals condition exhibited less proportion dominance, (M = -26.42, SD = 69.81) than participants in the group condition, (M = 17.20, SD = 71.08, $t(48) = 2.18$, $p < .05$, $\eta_p^2 = .09$). Also as expected, this difference is qualified for each analysis by a reliable or marginally-reliable interaction with moral conviction for both the participants' average comparison just reported, and for each item (See Figure 2 and Table 2). Participants higher in moral conviction — those who agree with the statement "this topic strongly relates to my core moral convictions" — exhibit more proportion dominance than those lower in moral conviction in the groups condition, but the reverse holds for the individuals condition. Figure 2 depicts that pattern of participants' average preferences broken up by high and low moral conviction and by assignment to experimental condition. Each of the items follows a similar pattern.[13] Table 2 reports the results of ANCOVAs run for each item and across the participants' averages.

[13] Results for the Zaire item does not exhibit the main effect of individual versus group construal like the other human life-related scenario (Anthrax). Results of a previous study that was similar in design, but where proportion dominance was measured on a 0 to 1 scale (Bartels & Burnett, submitted), found a reliable effect for the Anthrax scenario (M_{group} =.61, M_{indiv} = .35, $t(28) = 2.29$, $p < .05$, $\eta_p^2 = .16$), but only a medium-sized, non-reliable effect for the Zaire scenario (M_{group} = .59, M_{indiv} = .39, $t(28) = 1.64$, $p = .11$, $\eta_p^2 = .09$). These items are qualitatively similar to each other and to the other items, but in both studies, the Zaire item shows weaker effects, and I have not yet come up with a satisfactory explanation for the difference in the strength of the effect.

These results suggest that construal of resources as groups (versus individuals) causes greater weight to be given to relative (versus absolute) considerations in morally-motivated judgment and preference, especially for those people who moralize the domain under consideration.

3.4. Discussion

When decisions or evaluations permit both relative and absolute considerations, the weights given to these considerations depend on the degree to which resources are construed as groups versus individuals. Unger (1996) suggested that presenting a life as belonging to a smaller reference group reduces futility thinking, bringing moral judgment more in-line with a utilitarian perspective. Study 2 suggests a related point, that when the reference group is fixed, futility thinking is reduced when its members are construed more as individuals and less as a group.

Study 2 suggests that when people are induced to think about groups (i.e., "the bucket"), the moral weight afforded to individual lives saved (i.e., "the drops") is reduced, inducing (non-utilitarian) proportion dominance. Participants induced to think about resources as a collection of individuals, instead, exhibited less proportion dominance. That is, their preferences were more consistent with utilitarianism. Further, the divergence in preference introduced by group versus individual construal is greatest for participants who think of the choice situation as *especially* morally relevant.

The combined results of Studies 1 and 2 suggest that, in stark contrast to prevailing views, morally-motivated preference may be *more* flexible than amoral preference (in some circumstances). Study 1 found that morally-motivated preference was *both less* consistent with

utilitarianism (in the RB condition) and *more* consistent with utilitarianism (in the CR condition). Similarly, Study 2 finds nearly identical response patterns across experimental conditions for participants who do not view the situation as an especially moral one. But participants who ascribe moral relevance to the situation indicate preferences *both less* consistent with utilitarianism (in the groups condition) and *more* consistent with utilitarianism (in the individuals condition).

If proportion dominance follows naturally from group construal, this raises the question of whether proportion dominance is necessarily non-normative, as some (e.g., Bartels, 2006) have argued about specific cases of proportion dominance. Yet to the extent that proportion dominance is caused by group construal, the claim that it is fallacious amounts to the claim that groups are never the appropriate unit of analysis for thinking about resources at risk. There may be room for reasonable people to disagree on this point.

Some people prefer to see burdens like unemployment and military service distributed equitably rather than concentrated on particular communities. Some might judge an equitable distribution as morally better even if this means a slightly greater overall burden (J. R. Friedrich, personal communication, March 13, 2006).[14] These egalitarians may view this as a

[14] People have been shown to have robust egalitarian values, acting on these motives even in contexts that discourage it. For example, Dawes et al. (2007) show that people will promote egalitarian considerations (in a cooperation game) by redistributing resources at personal cost, even when doing so does not promote utilitarian considerations (i.e., does not reinforce future cooperative behavior). The participants in Dawes et al. (2007) might best be characterized as egalitarian consequentialists, but not utilitarians, because the total material good was not increased by promoting equality.

reasonable way of resolving the tension between avoiding "harm" (broadly construed), on one hand, and unbiased distribution of "harm", on the other.[15]

In the current study, treating the group — or "the bucket" — as the appropriate unit of analysis increases futility thinking, but all that can be said about preferences collected in the group condition of Study 2 is that they are decidedly non-utilitarian. Studies 3 and 4 avoid this negative definition (non-utilitarian) by posing conflicts between utilitarianism — and consequentialism, more generally — and deontology, and investigating the influences that promote one strategy versus the other. So, in discussing Studies 3 and 4 the terms "utilitarian" and "consequentialist" can be used interchangeably, and distinctions between different varieties of deontology do not affect the conclusions reached.

3.5. Transition from Part One (demonstration of flexibility) to Part Two (accounting for this flexibility) of this thesis

Part One of this thesis demonstrates flexible moral choice; Part Two examines the competing processes that produce this flexibility. In demonstrating that moral choice is malleable, these preceding studies implicated attentional processes (Study 1) and representational process (Study 2) as key components of moral preference formation.

After having noted the importance of understanding morally-motivated judgment and decision making in non-sacrificial contexts, I now return to those sacrificial contexts. I do so because the dominant framework for understanding morally-motivated preference in these

[15] Of course, utilitarians will argue that equality can have instrumental value (because of diminishing marginal utility, the greatest good can be promoted by giving more to the least well-off, so the argument goes), but that there is no *intrinsic* value in equal distribution. They will also require that the egalitarian specify (a) an assessment of how much equality matters, and (b) a way to assess degrees of inequality so as to produce (c) defensible principles for trading off certain amounts or degrees of equality in cases where equal distribution results in a smaller total well-being.

contexts (the PV framework) predicts rigid deontology, and thus is in need of some process-informed augmentation.

So, like Study 1, Studies 3 and 4 scrutinize utilitarian and deontological preference in situations where they are brought into conflict, with the result that non-utilitarianism is explained in terms of adherence to moral constraints. Study 3 develops a process-based approach to understanding morally-motivated preference formation in sacrificial contexts that is then re-examined as a possible account of PV-driven moral judgment in Study 4.

Studies 3 and 4 add a layer of complexity, suggesting that processes compete to shape moral judgment and preference. They also incorporate new methods, examining the influence of reliable individual differences in deontological tendency (Study 3) and differences in thinking styles (Studies 3 and 4) on moral judgment and preference. While what follows is somewhat less parsimonious, adding new processes and individual differences to the mix serves in the development of a more descriptively adequate process-based explanation of morally-motivated judgment and decision making.

Study 3, in particular, makes extensive use of methods and concepts borrowed from philosophy. Study 3 also aims to incorporate theoretical viewpoints from different research perspectives in moral psychology that have not received as much attention in judgment and decision making research. Study 4 aims to import some insights offered by these perspectives to judgment and decision making (and the PV framework, in particular).

4 Study 3—Affect-laden rule violations and catastrophe cases promote deontology-consistent and utilitarian preference

4.1. Introduction

Study 3 returns to the question of why people's preferences sometimes appear to disregard very highly-valued outcomes. Rather than varying the resources at issue, like in Studies 1, 2, and 4, participants in the current study were faced with choices about whether or not to sacrifice one human life to save others. Removing this factor allows for a focused initial test of the influence of a number of processes at work in resolving conflict between deontological and utilitarian response. In doing so, Study 3 develops a partial answer to the question "Why do people refuse to do things that will cause harm, even when as a result of not acting, *more* harm will result?"

4.1.1 Trolley problems, intuitive deontology, and catastrophe cases

As I mentioned in the general introduction, moral philosophers — and in particular, deontologists — have developed variants of the "trolley problem" and other similar dilemmas to elicit intuitions that support their arguments about morality.

Singer (1999, p. 187) succinctly conveys this method of developing a normative argument:

> "The appeal to intuitions is often used as a positive argument for a normative theory:
>
> > In the following circumstances, we all think it would be right to do B; normative theory T explains, better than any of its rivals, why it would be right to do B in those circumstances. Therefore we ought to accept T.
>
> Similarly, the appeal to intuitions has often been used as a negative argument against consequentialism:

> If consequentialism were correct, then in the following circumstances (a description of a case follows) we ought to do A. But we all think that it would be wrong to do A. Therefore consequentialism is false."

Deontologists use this method (in addition to other lines of argument) to suggest that for cases where a harmful action is judged impermissible, as in the footbridge case, consequentialism is an inadequate moral theory. More generally, Donagan (1977, p. 183) notes (as do several others): "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 argue that actions, and not their outcomes, are the proper unit of analysis. They further argue that because proscriptive rules can be narrowly framed and directed, actions with logically identical outcomes can, nonetheless, represent different kinds of actions and elicit different judgments (Davis, 1993). For example, unprovoked killing can be judged impermissible, while letting die and killing in response to aggression can be judged permissible.

Because constraints can be formulated to be *very* specific, some deontologists uphold even more nuanced distinctions, judging foreseen but unintended harm as more permissible than harm intended as a means to promote good ends. (Mikhail, 2007 reviews evidence that laypeople's moral judgment appears sensitive to *very* nuanced distinctions between types of actions.) One basic motivation common to these constraint-based approaches (as well as other approaches) is to try to systematize and account for our varying moral intuitions.

Utilitarians, on the other hand, treat many of these distinctions as irrelevant and suggest that people, on a sober second thought, would agree. They sometimes argue that the intuitions generated by these artificial problems are not trustworthy. 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 (like the "futility thinking" induced in Study 2), they should be discarded, and rational, logical analysis should form the basis of normative ethical theory (for a book's worth of examples, see Unger, 1996). (Both these lines of argument are reflected in Sunstein 2005, discussed in Section 1.4)

Deontologists are not the only philosophers who construct cases to generate intuitions about right and wrong that are consistent with their favored theory. Consequentialists have used "catastrophe cases" to elicit intuitions that are difficult to square with deontology, as illustrated in a prominent 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, 1976, p. 10)

Normative ethical theories try to serve many purposes. They intend rationalize intuitive judgment, and offer a basis for deliberative moral reasoning. And, in cases where intuition and deliberation come into conflict, they aim for resolution: either discard the intuition (Unger, 1996)

or add a “catastrophe clause” to accommodate it (Fried, 1976). Study 3 examines preferences elicited by trolley-like and catastrophe cases. The aim is not to use the results to argue for the strength of one normative ethical theory or another, but rather to address descriptive issues. Specifically, Study 3 is motivated by contemporary dual process models that link utilitarian judgment to deliberation and deontology-consistent judgment to intuitive process.

In Study 3, participants were presented with 14 trolley-like ethical dilemmas. Each dilemma had a standard version — where six people are at risk and harming one reduces the number expected to die — and two modified versions: a “vivid” variant and a “catastrophe” variant. To construct the “vivid” version of each dilemma, I added a re-description of the action entailed in the scenario that was intended to make the scenario more (negative) affect-laden. This was done to elicit the kind of moral outrage described by Donagan (above) and studied in the context of “sacred values” by Tetlock et al. (2000). The expectation was that this outrage might trigger deontology-consistent reponding in Study 3. The theoretical basis for these predictions is discussed below and is a theme revisited and tested again in Study 4.

The “catastrophe” variant describes the group at risk as 20, rather than 6 people, making the cost of adhering to a deontological constraint more grave. So, for these dilemmas, participants faced the decision of imparting harm for a net savings of 19 (rather than 5). I expect participants to demonstrate preferences more consistent with utilitarianism for these scenarios.

4.1.2. Intuitive and deliberative influences on moral judgment

Whereas philosophical argumentation in ethics intends to be rational, logical, and the product of deliberation, moral judgments elicited from laypeople are not always of that character. Moral cognition relies on both controlled and automatic process. The historically-dominant

descriptive account of moral cognition, Kohlberg's (1969) developmental theory, emphasizes the rational, deliberative aspects of moral judgment. People start with an egocentric focus on obedience and punishment — a type of focus on consequences — and at the highest stage, reason by universal ethical principles — a kind of reasoned deontology. Each successive stage in Kohlberg's scheme is characterized by increasingly deliberate reasoning and reflection (among other qualitative differences). Mixing the normative and the descriptive, Kohlberg's theory argues that coherent, deliberative deontology is the gold standard for moral judgment.

Recently, psychologists have been elaborating a view of moral cognition as rapid, intuitive and "hot" – that is, affectively charged and even affectively driven. Haidt (2001) believes most moral cognition can be explained in terms of processes that happen with little effort (Wegner & Bargh, 1998) and are unavailable to introspection (Nisbett & Wilson, 1977; Wilson & Schooler, 1991). He suggests that people's moral reasoning is often post-hoc justification for judgments produced by lower-order, nonrational (i.e., not related to a normative ethical position) affective reactions. He offers as evidence for his claim studies that purport to show "moral dumbfounding," in which people become unable to justify or explain their moral judgments to others (Haidt, Koller, & Dias, 1993). As absolutist deontology and strict consequentialism serve as illustrative normative bookends in philosophy, Kohlberg's and Haidt's frameworks serve a similar orienting purpose in the description of moral judgment as the product of deliberation or intuition.

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

Judgments elicited by ethical dilemmas have proved useful for developing process-based theories of moral cognition. Researchers have begun using trolley cases as a tool to investigate

the contribution of automatic and controlled process to moral judgment. (Cushman, Young, & Hauser, 2006; Nichols & Mallon, 2006, to name a few). Participants are asked to judge the permissibility of acts that do harm to one person to prevent harm to others.

To revisit the examples considered earlier, researchers have compared reactions to the bystander and footbridge versions of the trolley problem. In the former, a protagonist may flip a switch to divert runaway train car threatening to kill five railway workers onto a track where it will kill only a single railway worker. In the latter, 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 flipping the switch permissible (consistent with utilitarianism), but pushing the fat man impermissible (consistent with adhering to a deontological constraint, like "do no harm"; Mikhail, 2007).

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, 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 most widely-publicized of the trolley problem studies (cited > 170 times as of 5/10/07), 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.

They 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, Wheatley and Haidt (2005) found that judgments of moral transgression were more severe for participants who had been subliminally primed to feel disgust. Also, 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.

It is clear that moral judgment involves at least *some* emotional processing. However, the recent emphasis placed on emotional functioning in moral cognition may lead some to the conclusion that moral judgment is emotional response *and nothing more* (such a view is maintained by Prinz, in press). There is good reason to be skeptical of this claim.

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

Theories like Haidt's and Greene's attribute moral judgment almost exclusively to affective responses to situations. Like the claim that value-driven choice is rigidly non-consequentialist, there is reason to think the claim that moral judgment *is* emotional response is likely an over-simplification in need of correction.

A more moderate claim is made by the philosopher Shaun 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 moral rules that codify moral and immoral behavior. These constraints are "affect-backed" because they are often accompanied by affect. Nichols and Mallon (2006) attribute an important influence to affect, but argue that other, strictly emotion-based accounts, like Greene et al.'s, 2001, neglect the crucial role that rules play in people's reasoning about dilemmas.

Nichols and Mallon (2006) argue that proscriptive rules serve the important role of establishing preconditions for actions being viewed as morally wrong. While the function ascribed to constraints may seem tautological to some, Nichols and Mallon (2006) note that Greene et al.'s model would predict that a parent's choice to circumcise her son would be viewed as morally wrong. The action certainly qualifies as "personal" by Greene et al.'s standards. But since our culture does not have a rule proscribing circumcision, this behavior is viewed as permissible (or even a good idea).

Nichols' account suggests three processes interactively shape moral judgment: cost-benefit analysis, checking to see whether the action violates a moral rule, and an emotional reaction. To support the claim that moral judgment is mediated by affective response, Nichols (2002; Nichols & Mallon, 2006) presents two kinds of evidence. First, Nichols (2002) found that conventional norm violations that elicited affective reactions (e.g. spitting at the table) were judged as less permissible than rule violations that did not (e.g. playing with your food). Importantly for the hypotheses tested in Study 3, this effect was moderated by individual differences: the effect was more pronounced for participants high in disgust sensitivity.

Second, Nichols and Mallon (2006) developed trolley-like and footbridge-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 of whether the action was morally wrong, all things considered — what they call "all-in impermissibility". They show that violations of affect-backed rules are 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 by good or bad consequences of great magnitude. 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 such an action violates a moral rule. However, only 24% judged that the action was morally wrong, all things considered.

Nichols and Mallon's (2006) results suggest moral judgment is influenced by whether violations of moral rules evoke affective reactions and by whether sufficient attention is directed to consequences favoring violating a moral rule. It appears that non-affect backed rules operate as a normative consequentialist theory might treat all commonsense moral rules: In cases where the consequences favor a harmful action, infringing a deontological constraint may be 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 sufficient consequences favoring the sacrifice are great enough (i.e., many lives to be saved; Nichols & Mallon, 2006) and (c) individual differences in propensity to allow emotional reactions to influence judgment (Nichols, 2002). Study 3 tests for the influence of each of these factors by having participants respond to ethical dilemmas. Some of these dilemmas are modified to elicit a stronger emotional reaction to action, while others are modified to be more like the catastrophe cases.

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

Linking utilitarian judgment too strongly with deliberation is probably oversimplifying the picture. For example, Study 2 found that morally-motivated decision makers could be made more and less utilitarian by promoting different mental representations of the good to be promoted. There is no reason to assume that people exerted a great deal of mental effort in actively constructing these representations. In fact, the suggestion is that individual construal triggered *intuitions* – that is, rapid and effortless responses – that promoted utilitarian preference.

The studies reviewed in the previous two sections suggest that both intuition *and* deliberation shape moral judgment and suggest ways of explaining the apparent flexibility of morally-motivated judgment and choice. Both frameworks intend to explain judgments in contexts where deontology and utilitarianism are brought into conflict. Both treat deontology-consistent judgment as intuitive: Greene et al.'s (2001) research, in particular, suggests that sensitivity to violations of moral rules is often more reflexive than reflective. Nichols' (2002; Nichols & Mallon, 2006) account suggests the importance of affect (triggered by violations of constraints) for judging an action impermissible, *all things considered*.

Also, both frameworks motivate hypotheses about how and when deontological response can be overridden. 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 3 uses an individual differences approach to test predictions about automatic and controlled processes in moral judgment. Asking whether deliberative thinkers will exhibit preferences more consistent with utilitarianism than intuitive thinkers offers one way of informing a process-based account of morally-motivated choice. Thus, Study 3 uses a modified version of Epstein'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 preference elicited by ethical dilemmas is driven by emotional activation, then one might expect that intuitive thinkers should demonstrate more deontology-consistent preference, while deliberative thinkers would demonstrate utilitarian preference.

4.1.6. 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 process (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 strictly 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 people's judgments are sensitive to relatively nuanced distinctions (e.g., doing/allowing harm,

treating people as means/ends — see also Waldmann & Dieterich, 2007, intentional harm/harm produced as a side effect of good intentions). Sensitivity to these features is suggestive of the operation of relatively narrowly framed and directed deontological constraints in moral judgment. And, while they do not offer a sophisticated 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 deontological constraints, Study 3 takes the straightforward approach of directly assessing participants' endorsement of deontological principles. The Idealism subscale of Forsyth's (1980) Ethics Position Questionnaire (EPQ) does just that. This measure asks people to 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 moral rules proscribing harm to humans, participants posed with the footbridge case should be likely to engage in a simple utilitarian calculus and view a five-lives-for-one tradeoff permissible.

4.1.7. Hypotheses

Preference is expected to be most consistent with deontology for vividly-described dilemmas, and most consistent with utilitarianism for the catastrophe dilemmas. The vivid descriptions of harmful actions make them seem more egregious, and because their attention is

directed to the act (and not its ultimate consequences, as in the RB condition of Study 1), I expect participants to view the actions as moral violations. Describing the consequences as especially grave (i.e., killing 19 people, rather than 5) serves to focus participants on the actions' ultimate consequences, thus directing attention away from the violation of a deontological constraint (as in the CR condition of Study 1).

Deliberative thinkers are expected to exhibit more utilitarian preference than intuitive thinkers.[16] Greene et al. (2001) suggest a testable "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 "correct" for 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 deontology-consistent response 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

[16] This prediction is also tested in Study 4. This prediction only extends to the sacrificial contexts tested. Whether this tendency will hold in other, non-sacrificial contexts, or for decisions involving moral issues other than suffering and harm — like hierarchy, reciprocity, or purity (see Haidt & Joseph, 2004) — is an empirical question not examined in the studies reported here.

principled. If participants' decontextualized normative perceptions (idealism) predict revealed preference, then moral preference may be shaped, in part, by adherence to moral rules.

4.2. Method

4.2.1. Participants

Seventy-one Northwestern University undergraduates (45 females and 26 males) participated. Each participant completed the study at his or her own pace. They were tested individually, but in a small group setting (typically one to four participants per session). Usually, other participants were completing the study at their own pace in the same room. 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 (completing 30 minutes of their required 10 hours of research participation to receive full credit for the course).

4.2.2. 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 B). Second, participants read and gave judgments for 14 ethical dilemmas similar to the one below (see Appendix C 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 (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?

NO **-2** **-1** **+1** **+2** **YES**

Participants indicated their responses by clicking on one of the boxes.[17] Responses were recoded from zero to one (coded 0.00, 0.25, 0.75, 1.00) so that higher numbers indicated more utilitarian preferences.[18]

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

[17] Pilot work suggested the 4-option format. Participants in a pilot study indicated they were frustrated by a simple yes/no choice and suggested the task would be more well-received and taken more seriously by others if the response format included in-between responses.

[18] The results do not crucially depend on the way responses are scored. At the suggestion of a reviewer, I also analyzed the data using just the endpoints of the scale (-2 and +2) and treating the scale as a dichotomous measure (scoring -2 and -1 zero, and scoring +1 and +2 one). None of the qualitative patterns change, with the one caveat that an analysis using just the endpoints (discarding 64% of the data) results in lower scores overall. This flat influence is due to participants' general reluctance to strongly endorse the utilitarian solution. When people responded "no", they used the endpoint (-2) 45% of the time, but when they responded "yes", they used the endpoint (+2) only 26% of the time (only 12% of the time overall).

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 catastophe, 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 B). Eight participants did not provide REI responses because of the time constraints placed on the experimental sessions in which they participated.

4.3. Results and discussion

4.3.1. Experimental results: Influence of vividness and catastrophe manipulations

Study 3 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 3 shows, the within-subjects contrast yielded a large effect F (2, 140) = 23.19, MSE = .02, $p < .001$, $\eta_p^2 = .25$), 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 3 are between-subjects. Even though the within-participants row of Table 3 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 0.90. Clearly, there was wide variation in preferences even within this (potentially restrictive) sample of undergraduates.

4.3.2. Influence of individual differences

All of the effects reported in this section appear in Table 4. 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).[19] 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

[19] 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.

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 inhibiting the affectively pre-potent deontological response. In addition, the relationship between idealism and preference also suggests a role for deontological constraints. However, because these thinking styles and idealism were not independent — deliberative thinkers indicated less agreement with deontological principles than intuitive thinkers (r = -.34, p < .01) — a more in-depth look at the data is warranted.

The left panel of Figure 3 presents the preference data conditioned on a median split of REI scores (Deliberative = participants above the median; Intuitive = participants below the median). First, observe the overall positive slope across conditions, offering support for the main prediction — that catastrophe condition responses would be the most utilitarian and vivid condition responses would be the least utilitarian. Note also that the "deliberative" line lies above the "intuitive" line, meaning that deliberative thinkers' preferences were more utilitarian than intuitive thinkers' preferences. The picture gets more complicated if we present the data for participants above and below the median on idealism, as shown in the middle and right panels of Figure 3. Thinking styles interacted with idealism scores to predict each participants overall average preference (Std $\beta_{\text{Thinking Styles}}$ = .25, p < .05, Std β_{Idealism} = -.20 , p = .10, Std $\beta_{\text{REI*Idealism}}$ = -.26, p < .01).

Thinking style had a larger effect on preferences for low idealism participants — those who indicated relatively little agreement with deontological constraints (right panel of Figure 3) — than for participants who endorsed deontological constraints (middle panel of Figure 3). Considering just low–idealism participants, deliberative participants appear more utilitarian than their intuitive counterparts. The reliable interaction noted above is largely attributable to

the difference between these low-idealism, deliberative participants and everyone else: they appear to be the (especially utilitarian) odd group out. They show little effect of the vividness manipulation, for example (see right panel of Figure 2). This group of participants may have actively corrected for an emotional reaction that (they believe) is not principled, in that they do not endorse adherence to deontological constraints.

In summary, Study 3 demonstrates that there may be more than one type of morally-motivated decision maker, and that morally-motivated choice is remarkably flexible. The results show that participants who affirm deontological principles, and those who rely more on intuition than deliberation, exhibit preferences more consistent with deontology.

The study also showed that that focusing participants' attention on violations of deontological rules — by exacerbating people's negative emotional reactions to these actions — promotes deontology-consistent choice. But at the same time, the results indicated that participants in Study 3 were willing to sacrifice human lives (in the catastrophe condition) if enough human lives can be saved by doing so.

4.3.3. Why the generalizability of Study 3's results is limited

Psychologists and philosophers have studied intuitions elicited by trolley-like dilemmas intending to explain *de facto* morally-motivated processes, describing sensitivities to features of dilemmas that generalize across people. One problem with the general approach taken by these researchers is that in trying to develop parsimonious theories postulating "fundamental" (i.e., context-independent) laws of human thought that generalize across a wide range of content domains, these designs 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 3 partially avoids the anti-context bias by accounting for within-sample variance. Results demonstrate that participants who are more likely to override emotional influence on preferences appear more utilitarian, and that people who endorse moral rules demonstrate preferences more consistent with the operation of those moral rules. However, since Study 3 uses only trolley-like dilemmas, the generalizability of these results is limited.

In light of this limitation, Study 4 tested whether similar processing principles work for other types of choice scenarios. Recall Hare's (1981) wry comment about "playing at 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-consistent. At a minimum, therefore, researchers should be wary about generalizing the set of processing principles implicated in response to trolley-like dilemmas to other choice contexts. Study 4 is, in this way, a conceptual replication and generalization study, and uses as contexts the types of policy-level decisions for which, utilitarians argue, their theory is especially well-suited (Goodin, 1993).

5 Study 4—Protected values as affect-backed constraints: Moral judgment in separate (4a) and joint evaluation (4b)

5.1. Introduction

Study 4 offers a second answer to the question "Why would people's judgments and choices disregard the consequences they care about the most?" Like Study 1, Study 4 examines whether and when we should expect moral values to engender non-consequentialist decision principles. Study 4 examines the processes underlying morally-motivated judgment and decision making by examining PV-driven response in two qualitatively different kinds of evaluative context.

Studies 1 and 3 suggest that moral preference is influenced by the presence of moral rules and by whether attention is directed towards the permissibility of rule-violating act or to the act's consequences. Study 4 expands on these findings, relating process-based accounts of people's responses to ethical dilemmas to an investigation of PV-motivated judgment. The current study uses the findings of Studies 1 and 3, and the approaches discussed in Study 3 to examine the context-sensitive role of moral rules in moral judgment across content domains and across individual differences.

The major motivation for Study 4 is cross-pollination between the moral judgment and morally-motivated decision making literatures. One idea motivating Study 4 is that PVs share important properties with "affect-backed rules" — the constraints that comprise Nichols' (2002; Nichols & Mallon, 2006) affect-backed normative theory. As I mentioned earlier, PVs are intimately tied to strong emotions — proposed tradeoffs of PVs can elicit extreme anger (Baron & Spranca, 1997; see Tetlock et al., 2000 for related effects). But just as even affect-backed rules can be overwhelmed if sufficient attention is directed to consequences favoring infringing them

(as in the catastrophe cases in Nichols & Mallon, and in Study 3), Study 1 showed that people's willingness to accept tradeoffs of PVs varies depending on where attention is focused, a factor that varies widely across real-world contexts. In particular, Study 1 found that in contexts that direct attention towards net benefits brought about by doing harm, people with PVs were more willing to engage in the harmful action than people without PVs.

5.1.1. Protected values as affect-backed constraints

Study 4 examines whether PVs operate like affect-backed constraints in moral judgment. In both Study 4a and Study 4b, participants rate their approval or disapproval of decisions made by government administrators. In Study 4a, they also judge whether the administrators' decisions violate a moral rule (judgments of "weak impermissibility" in Nichols & Mallon, 2006).

For each scenario, each of two administrators is described as facing a choice about whether to knowingly *do harm* to a resource to mitigate even greater harm or to merely *allow* the harm to happen. For example, participants read that 20 species of fish upstream from a dam would be made extinct unless the dam is opened, but if the dam is opened, some species downstream will be made extinct (see Appendix C). One administrator is described as motivated by a desire to do no harm (and thus not to act), and so 100% of the anticipated harm to the resources occurs. For this 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 this class of choices as "omission."

The other administrator is described as first calculating that by intervening, he or she will kill 80% of the resources under consideration, and based on this analysis, he or she chooses to

intervene. For this 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."*[20] I refer to this class of choices as "action." Study 4a presents the decisions in separate evaluation: on a given trial, participants evaluate either the omission or the action, but not both. Study 4b presents the decisions in joint evaluation: both the omission and action are evaluated on a single trial.

Nichols and Mallon (2006) show that judgments of actions as morally wrong may be, but are not always, guided by judgments that a moral rule has been violated. They argue that rules are essential to moral cognition: in the absence of rules prohibiting an action, they argue, the action is not likely to be viewed as morally impermissible. However, while it may be *necessary*, judging that a rule has been violated is not always *sufficient* for judging an act morally wrong, all things considered. By their account, such judgments depend on both a rule violation *and* on the emotional reaction to the violation. They found that judgments that a rule had been violated ("weak impermissibility") were not sufficient to produce judgments that an action was morally wrong, all things considered ("all-in impermissibility") for emotionally pallid dilemmas.

Non-affect-backed constraints, it seems, are treated as utilitarians would treat any commonsense moral rule: if the consequences favor a harmful action, infringing them may be required and thus, morally justified, producing dissociation between weak and all-in

[20] In the preceding studies, the decisions were first-person. The switch to evaluation of third-party decisions was necessary for Study 4, because response variance on the dependent measures is anticipated to be contingent on the decision principles and related cognition that are imputed to these agents. It would be considerably more difficult to induce choices and decision principles on participants who consider the application of those principles morally impermissible.

impermissibility. The operation of affect-backed constraints is more consistent with a rigid deontology: violating these rules is forbidden except in the most extreme circumstances.

By collecting both judgments of rule violation and something akin to all-in impermissibility — judgments of approval or disapproval of a government administrator's decision — Study 4 offers a straightforward test of whether PVs are affect-backed constraints. Nichols and Mallon (2006) suggest that judgments of rule violation are sufficient for judgments of all-in impermissibility for affect-backed-constraints, but not for non-affect-backed constraints. So, Study 4 compared the degree to which judgments of rule violation predict decision approval for those items where participants do and do not endorse PVs. The relationship between these two types of 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" — a process-level 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 4, 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 deontological 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*.

5.1.2. Joint versus separate evaluation preference reversals

The preceding studies show that utilitarian judgment and preference can be promoted or diminished by the nature of the mental representation of the situation (Study 2) or by attentional effects introduced by the task (Studies 1 and 3). Study 4 examines whether moral value-driven focus on rules and consequences is subject to a different set of task constraints. Study 4a asks participants to evaluate decisions independently, in separate evaluation. Study 4b presents the same participants with two decisions in a joint evaluation format, inviting the participant to compare the two scenarios before rendering judgment.

Previous research demonstrates that attributes that appeal to one's affective or intuitive sensibilities, and attributes that are otherwise easy to evaluate, constrain preference in separate evaluation (where a number of otherwise useful comparisons are not made available). In contrast, attributes with greater normative significance that appeal to "colder," more logical sensibilities exert a larger influence on preference in joint evaluation (Bazerman et al., 1999; see also Hsee et al., 1999). For example, Hsee and Leclerc (1998) asked three groups of participants to assign buying prices to an ice cream product. One group was asked to evaluate a 7 oz. serving of ice cream presented in a 5 oz. cup, a second group was asked to evaluate an 8 oz. serving in a 10 oz. cup, and a third group assigned buying prices to both.

Participants in the first condition were willing to pay more for 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 overfilled/underfilled attribute of the product into their evaluative judgment. Of course, buying prices in the joint evaluation were greater for the 8 oz. serving than the 7 oz. serving. The joint evaluation condition affords participants with a richer evaluative context, where participants are able to pick out the most important attribute for setting a buying price.

One interpretation of this preference reversal is that people essentially discard intuitive reactions when comparison highlights an even more *sensible* basis for evaluation, and this is one way in which argumentation in normative ethics proceeds. Deontology often 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 be constructed around explaining "commonsense" moral intuitions, an aim Unger (1996) refers to as "preservationism." In this sense, our affective reactions to cases (perhaps produced by reliance on affect-backed constraints) are honored as meaningful data in the development of normative ethical theory.

Utilitarians wish to align our ethical cognition with their interpretation of the demands of rationality (Pettit, 1993; Goodin, 1993), and some argue the best way to do this is to endorse "liberationism" (Hare, 1981; Unger, 1996). That is, we should discard many of our moral intuitions and instead rely on very basic moral values, like promoting total well-being and/or lessening serious suffering. Unger (1996) does this by presenting pairs of moral choice scenarios that illustrate to the reader that her intuitive judgments in one type of case are premised on a factor that is clearly irrelevant in a structurally similar case. The intent of this method is to persuade the reader of the superior logic of utilitarianism (see also Hare, 1981).

Study 4a examines PV-driven responses to single cases, some of which involve the (outrage-producing) violation of a PV. Because previous work suggests separate evaluation judgment is more likely to be influenced by features that appeal to one's intuitive sensibilities, one might expect violations of PVs to be associated with deontology-consistent response for these cases.

Study 4b asks participants to engage in a comparison before rendering judgment (not unlike Unger's 1996 method). Because previous research suggests attributes that appeal to "colder" sensibilities are afforded more weight in joint evaluation, we might expect influence of affect-backed constraints (i.e., PVs) to be overwhelmed by utilitarian considerations in these contexts.

Bazerman and Messick (1998) argue that one way to assess the normative status people give to deontology-relevant vs. utilitarian-relevant attributes is to present scenarios in joint evaluation. Not unlike Unger and Hare, they suggest consequentialism better captures people's lay normative theory for a host of problems. So, by their account, sensitivity to deontological considerations (e.g., the doing/allowing harm distinction) should be diminished in joint evaluation.

The current study collects judgments in both separate (Study 4a) and joint evaluation (Study 4b) to assess whether disapproval rendered for violations of PVs is robust or overwhelmed when the consequences favoring violating a PV are made salient in joint evaluation. In other words, Study 4 tests whether people invest the doing/allowing harm distinction with enough normative significance to outweigh utilitarian considerations for decisions involving PVs.[21]

---

[21] Two points are worth mentioning here about any perceived privilege afforded to decision makers by joint evaluation contexts. First, the "distinction effect": Hsee and Zhang (2004) demonstrate that joint evaluation can lead people to *overweight* attributes that appeal to logical sensibilities, resulting in suboptimal choice. They use the example of a shopper who evaluates two sets of speakers for a home audio system: one that is pleasing to look at but almost imperceptibly poorer in sound quality in joint evaluation versus one that is ugly but almost imperceptibly higher in sound quality. The shopper chooses based on sound quality (the more "sensible" attribute), but when he gets home, he cannot bear to look at them, and so the speakers go unused. Hsee and Zhang (2004) argue that mismatches between evaluative context (in this case, joint) and consumption experience (in this case, separate) can produce suboptimal choices. Second, on base rates: Some have argued that ‘‘much of

5.1.3. Intuitive and deliberative thinking styles

Study 4 assesses whether the processing principles identified in Study 3 and postulated by Nichols and Mallon (2006) generalize to policy-level decision making (which utilitarians argue should promote utilitarian moral judgment; see Goodin, 1993). Like Study 3, Study 4 uses a modified version of Epstein's (1996) Rational versus Experiential Inventory (REI) to measure differences in intuitive/deliberative thinking styles.

The affect-backed rules account and the dual-process account offered by Greene et al. suggest a number of important, competing influences on moral judgment. First, both accounts suggest an association between emotional influence and deontology-consistent response. Greene et al. (2001) equate the two. Nichols and Mallon (2006) stress the importance of affect elicited by the violation of a constraint for producing deontological response.

Second, both accounts offer an explanation of utilitarian response. Greene et al. (2001) view some of these judgments as the product of deliberatively over-riding the affectively pre-potent, deontology-consistent response. Nichols and Mallon (2006) show that even affect-backed constraints are yielded when the utilitarian consequences favoring a sacrifice are great enough, as in catastrophe cases.

Individual differences in the propensity to allow intuitive reactions to influence judgment are expected to influence judgments of approval and disapproval in Study 4a. In separate

---

life resembles a between-subjects experiment'' (Kahneman, 2000, p. 682) — that many of life's experiences take place in separate evaluation contexts (e.g., Kahneman, Knetsch, & Thaler, 1986; Shafir, 2002). The point to be stressed here is not whether people's judgments elicited in either context are, in a sense, "better" (as utilitarians might argue), but rather that these two contexts can differentially trigger and inhibit some of the cognitive processes underlying moral judgment.

evaluation — a context that promotes intuitive judgment — we might 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).

5.1.4. 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 (outrage-driven) 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 moral outrage 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 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.

5.2. Methods

5.2.1. Participants

Forty-eight undergraduates (25 women and 23 men) participated in Study 4a for partial course credit (completing 30 minutes of their required 10 hours of research participation to receive full credit for the course). Participants completed the study tasks at their own pace. They were tested individually but in a small group setting (typically, 1 to 4 participants per session). Usually, other participants were completing the study in the same room. Those participants who wrote their contact information on a sign-up sheet for Study 4b 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 4b in exchange for $5 compensation. Sixteen of Study 4b's participants were run in one session; the others participated individually in a small group setting. No differences were observed on any of the variables between the large- and small-group setting participants in Study 4b.

5.2.2. 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 4a], 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 B). 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 4a, 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 4a, 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 4b, 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 4a.

For each of the three packets, the order of problems (i.e., birds, children, etc.) was randomized. For Study 4a, the assignment of action type to the problems was randomized so that on every other trial, participants evaluated an omission (or an action).

5.3. 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.

5.3.1. Rule violations and approval (Study 4a)

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, $p < .05$). This finding is consistent with the notion that PVs function like affect-backed constraints in influencing moral judgment.

5.3.2. Approval ratings—Separate evaluation (Study 4a)

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 4a, 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) x 2 (Domain: No PV/PV) repeated-measures ANOVA reveal 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 ($F(1,43) = 12.58$, $p < .001$, $\eta_p^2 = .23$)[22]. For those items for which participants endorsed PVs, they indicated less approval overall than they did for items for which they did not endorse PVs. This main effect likely reflects the sentiment

[22] Four of the 48 Ps in Study 4a 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 5. Similarly, two of the 32 Ps in Study 4b endorsed all seven PVs, and another three Ps endorsed zero. These five Ps' responses are counted in the item analyses summarized in Table 6, but not for the analyses run on Ps' averages.

that for resources bestowed a "protected" status, *any harm* elicits a negative reaction, driving approval ratings down.

Figure 4 depicts the pattern of results obtained for Studies 4a and 4b, 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 Figure 4. 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. Figure 4 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 Figure 3).

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 Figure 4), nor did difference scores for omissions relate to thinking styles ($r(43) = .15$, $p > .10$). I did not predict a difference in approval for omissions as a function of PVs for two reasons. First, administrators who *allow* harm do not necessarily violate a PV. Second, Haidt and Baron (1996) demonstrated that omission bias — a preference for allowing harm over doing harm — did not generalize to evaluations of third-party decisions made by people occupying especially responsible social roles, like that of a government administrator.

Table 5 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 4.1"). The results of ANOVAs run for each item, using 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 5.

A reliable preference for action over omission is present in five of the seven items used. Also, for five of the items used, participants with PVs indicated less approval overall than participants without PVs. (Again, contemplating *any harm* happening to a cherished resource may elicit a negative reaction.) Finally, 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.

5.3.3. Approval ratings—Joint evaluation (Study 4b)

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) x 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.

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 Study 1's 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 4b'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 6 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 using Decision (Action vs. Omission) as a repeated-measures factor and the presence 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.

5.4. Discussion

Studies 4a and 4b were motivated by the idea that a better understanding of moral judgment and choice can only be achieved through more thorough scrutiny of the processes that moral values motivate. Study 4 proposed that the competing processes identified in Study 3 would also be at work in PV-driven judgment.

Study 4 found evidence that the processes implicated in responses to ethical dilemmas also generalize to moral judgment rendered for public-policy-level decisions. That is, these judgments are influenced by whether rule-violations evoke strong affective reactions, by whether sufficient 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 moral judgment, the present

findings qualify previous conclusions, suggesting a more flexible PV-driven judge. Moreover, Study 4 accounts for some of this flexibility by suggesting PVs operate as constituents of an affect-backed normative theory, offering some badly needed theoretical synthesis across unnecessarily disparate literatures.

6 General discussion

Together, 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 mental representation of the resources under consideration, (b) the presence of strong moral attitudes or deontological constraints, and (c) reliable individual differences in tendency to engage in intuition and deliberation.

The approach embodied in these studies places a greater emphasis on the flexibility of moral cognition than some other approaches did. For example, some frameworks intended to capture people's responses to ethical dilemmas stress the reflexive, emotional nature of these responses. However, there is more to moral judgment than knee-jerk responding; 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 cold 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. Here again, the current studies suggest the picture is more complicated than previously assumed. They show 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 general approach of these studies is to treat the apparent context-sensitivity of moral judgment and preference as diagnostic of some of the psychological process underlying

people's responses. Study 2, for example, targets the influence of mental representation on moral preference. It shows that participants induced to think about resources as a collection of individuals (a key component of utilitarianism) render more utilitarian moral judgment than participants induced to think about resources as undifferentiated groups. Moreover, this effect is more pronounced for participants who ascribe moral relevance to the domain under consideration.

The other studies examine the processes by which utilitarian and deontological response was promoted or inhibited. Study 3 shows that participants who affirm deontological principles and participants who rely more on intuition than deliberation have preferences more consistent with deontology. In addition, results suggest that focusing participants' attention on actions that violate deontological rules promotes deontological preference, while focusing on the consequences favoring violating the rules promotes utilitarian preference. Study 1, 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, Study 4 demonstrates that the adherence to deontological constraints evident in Studies 2 and 3 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.

6.1. Revisiting proportion dominance and futility thinking

Utilitarianism requires summing the degree of satisfaction of welfare interests (i.e., basic needs) across individuals, counting each individual equally (Goodin, 1993; Kagan, 1998). In the scenarios tested in Study 2, this seems like a sensible and easy to execute strategy: you simply count up the number to be saved by each alternative and choose accordingly. However, the literature suggested people do not always choose this way; instead, people sometimes exhibit proportion dominance (Bartels, 2006), and futility thinking may underlie some of these choices (Unger, 1996).

The approach taken up in Study 2 emphasized the role of mental representation in constraining people's judgments and decisions, in general, and in moral judgment and preference in particular. Study 2 was motivated by the idea that a better understanding of proportion dominance might be achieved through more careful scrutiny of the effect's cognitive underpinnings. The findings suggest that proportion dominance is not limited to the domain of simple preference, but rather that the same phenomenon appears in moral judgment. This should come as no surprise, as the resources people reason about in proportion dominance studies are resources to which many people ascribe moral relevance (e.g., human lives and natural resources).[23]

Study 2 investigated whether deindividuating the people at risk would undercut utilitarian calculus, predicting and finding less utilitarian preference for participants assigned to the group condition. This method of undermining utilitarian judgment is, of course, not limited to studies of proportion dominance and not limited to the particular artificial means of manipulating individual versus group construal employed in Study 2. For example, social

[23] But it is surprising that work in philosophical ethics (Unger, 1996) presented sophisticated and testable descriptive hypotheses concerning the cognitive underpinnings of proportion dominance before the experimental psychologists researching the effect did.

psychologists have invoked the notion of entitativity, or the degree to which a social group constitutes a single entity (Campbell, 1958). Many studies have examined factors that promote *entitativity* (e.g., Brewer, Weber, & Carini, 1995) and the consequences of entitativity for memory of, and reasoning about, persons and groups (e.g., Hamilton & Sherman, 1996). Most relevant to the current studies are the findings that perceivers see greater entitativity in out-groups (Wilder, 1981), minority groups, and groups with whom the perceiver has a competitive relationship (Brewer et al., 1995). So, one might expect more proportion dominance for decisions concerning members of out-groups, minority groups, and competing groups.

Group versus individual construal also appears in cultural psychology, in several forms: holistic versus analytic cognition (Nisbett, Peng, Choi, & Norenzayan, 2001), interdependent versus independent construal of the self (Markus & Kitayama, 1991), and collectivism versus individualism (Hofstede, 1980). At the risk of oversimplification, this work suggests that cultures emphasize interconnectedness versus individuality to different degrees. One might expect cultural differences in construal to lead to corresponding tendencies in proportion dominance.

6.2. 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 judgment. It is clear that emotions play an important role in moral judgment, but it also appears that emotions cannot do *all* the work. In short, these approaches may explain *Humean* morality (see Hume, 1969/1739), but they cannot explain *human* morality. First, as Mikhail (2007) and others have pointed out, accounts that attribute moral judgment to an emotional reaction are missing an appraisal

theory. Without first addressing the question of *what about the stimulus* is upsetting (for Greene et al., 2001) or what is disgusting (for Haidt, 2001), 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). While the studies reported here do not test whether amoral emotion-eliciting actions are judged wrong, they do suggest *strictly* emotion-based accounts are insufficient.

It would seem that combining rule-based and emotional process accounts, as this thesis does, 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.

In fact, each study in this thesis that tested for the influence of deontological constraints on moral judgment and preference found it. Study 3 found that people who affirmed deontological constraints were more likely to exhibit deontological preference, and that with greater endorsement of deontological constraints, differences in thinking styles exerted a smaller influence on moral preference.

Study 4a 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). In another test, Study 4a found that judgments of rule violation and disapproval were more highly correlated for 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 3) 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 the CR condition of Study 1, the Catastrophe condition of Study 3, and the joint evaluation context of Study 4b). 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 (Study 3 and 4a).

The view developed in Studies 3 and 4 is a relatively simple one, largely borrowed from Nichols and Mallon (2006). 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 his or her computational resources to do so, and (3) registers a judgment or preference. 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 3) and PVs (Study 4a) and the test of the relationship between judgments

of rule violation and disapproval (in Study 4a) as stronger support for their framework than they adduce.

6.3. Revisiting the protected values framework

The protected values framework tests for the presence or absence of moral rules but has decidedly less to say about psychological process than some of the accounts intended to capture people's responses to trolley problems. This thesis argues that a better understanding of the role of protected values in decision making is achieved by studying the processes that PVs motivate and the processes that make use of PVs. And so, one purpose of the studies reported here was to test whether some of the competing processes identified in the trolley problem literature are at play in PV-driven judgment and decision making. Study 4 develops 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 framework. Indeed, Studies 1 and 4a offer a great deal of support for predictions motivated by previous empirical work on the role of protected values in decision making. One part of Study 1 was a replication of previous work: in one condition, PVs were associated with quantity insensitivity, which is taken as evidence that PVs motivate nonconsequentialist, deontological decision strategies. Similarly, in Study 4a, 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 has its best chance for descriptive adequacy. However, when consequences are highlighted, either by the preference elicitation procedure (Study 1) or the comparative nature of the choice context (Study 4b), 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 4b) or even *more utilitarian* (in Study 1) 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 presented in this thesis are the first to date (to my knowledge) suggesting 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.

6.4. On combining experimental and individual differences-based approaches

The approach embodied in these studies placed a greater emphasis on individual differences than some other approaches have. Cognitive psychologists, social psychologists, and experimental philosophers who investigate processes involved in moral judgment tend to neglect the variance attributable to individual differences. Some research has focused on variability in moral judgment attributable to socioeconomic factors (e.g., a study comparing two high and two low SES groups in Philadelphia and Brazil found that high SES Penn undergraduates were unique in *not* judging harmless,[24] but disgusting acts — like eating one's dog after it is killed in an accident — as "wrong"; Haidt, Koller, & Dias, 1993).

Other research has focused on cultural differences, suggesting that social conventions strongly influence moral cognition (e.g., Shweder, Mahapatra, & Miller, 1987). In these cases, groups are compared, but variance within a group is not treated as meaningful for theory development as it could be. Other research programs aim to identify universal, or at least pan-cultural, 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. This thesis demonstrates how an individual-differences based approach that accounts for variance *within a sample* can complement experimental tests of psychological process.

[24] "Harmless" in this context means that none of the actions resulted in loss of life or other physical harm, loss of rightful property, loss of privacy or other threats to autonomy.

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.)

Each of the studies in this thesis treats individual differences as instructive, using differences between people to inform a process-based explanation of morally-motivated judgment and preference. 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 able to account for some variability of morally-motivated cognition across contexts and across different groups of people. These studies offer a more enriched and informative view of moral cognition only because of the fruitful combination of experimental and individual differences-based approaches they present.

6.5. Conclusion

The studies reported here find that morally-motivated judgment and preference (a) makes use of intuitive and deliberative process, (b) is reliant on the way in which the problem is mentally represented, and (c) recruits representations of both deontological constraints and utilitarian considerations. This thesis implicates 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.

References

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. (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. (1997). Confusion of relative and absolute risk in valuation. *Journal of Risk and Uncertainty, 14*, 301–309.

Baron, J. (2000). Moral thinking. In *Thinking and deciding (3rd ed.)* (pp. 381-407). New York: Cambridge University Press.

Baron, J., & Leshner, S. (2000). How serious are expressions of protected values? *Journal of Experimental Psychology-Applied, 6*, 183-194.

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes, 94*, 74-85.

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

Bartels, D. M. (2006). Proportion dominance: The generality and variability of favoring relative savings over absolute savings. *Organizational Behavior and Human Decision Processes, 100*, 76–95.

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.

Bloom, P. (1996). Possible individuals in language and cognition. *Current Directions in Psychological Science, 5,* 90–94.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition, 56,* 1–30.

Bloom, P., & Veres, C. (1999). The perceived intentionality of groups. *Cognition, 71,* B1–B9.

Bloomfield, A. N., Sager, J. A., Bartels, D. M., & Medin, D. L. (2006). Caring about framing effects. *Mind & Society, 5*, 123-128

Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology, 69,* 29–40.

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

Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3,* 14–25.

Connolly, T., & Reb, J. (2003). Omission bias in vaccination decisions: Where's the "omission"? Where's the "bias"? *Organizational Behavior and Human Decision Processes, 91*, 186-202.

Connolly, T., & Reb, J. (2004). Omission bias in vaccination decisions: Still searching for evidence: Working Paper, University of Arizona.

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. (1998). *Philosophical ethics*. Boulder, CO: Westview Press.

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

Ehrich, K. R., & Irwin, J. R. (2005). Willful Ignorance in the Request for Product Attribute Information. *Journal of Marketing Research, 42*, 266-277.

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

Fetherstonhaugh, D., Slovic, P., Johnson, S. M., & Friedrich, J. (1997). Insensitivity to the value of human life: A study of psychophysical numbing. *Journal of Risk and Uncertainty, 14*, 283-300.

Fischhoff, B. (1991). Value elicitation: Is there anything in there? *American Psychologist, 46*, 835-847.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1980). Knowing what you want: Measuring labile values. In T. Wallsten (Ed.), *Cognitive processes in choice and decision behavior* (pp. 117–142). Hillsdale, NJ: Erlbaum.

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

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

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

Friedrich, J., Barnes, P., Chapin, K., Dawson, I., Garst, V., & Kerr, D. (1999). Psychophysical numbing: When lives are valued less as the lives at risk increase. *Journal of Consumer Psychology, 8*, 277-299.

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

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press.

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.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814-834.

Haidt, J. & Baron, J. (1996). Social roles and the moral judgment of acts and omissions. *European Journal of Social Psychology, 26*, 201-218.

Haidt, J., & Joseph, C. (2004). Inuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*, 55-66.

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

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review, 103,* 336–355.

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

Harman, G. (1977). *The nature of morality: An introduction to ethics*. New York: Oxford University Press.

Hastie, R. & Dawes, R. M. (2001). *Rational choice in an uncertain world.* Thousand Oaks, CA: Sage Publications.

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

Hofstede, G. (1980). *Culture's consequences: International differences in work-related values.* Beverly Hills, CA: Sage.

Hooker, B. (1999). Sacrificing for the good of strangers—Repeatedly. *Philosophy and Phenomenological Research, 59*, 177-181.

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

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin, 125*, 576-590.

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

Jenni, K. E., & Loewenstein, G. F. (1997). Explaining the "identifiable victim effect". *Journal of Risk and Uncertainty, 14*, 235-257.

Joyce, R. (2001). *The myth of morality*. Cambridge, UK: Cambridge University Press.

Joyce, R. (2006). Metaethics and the empirical sciences. *Philosophical Explorations, 9*, 133-148.

Kagan, S. (1994). Normative Ethics. Boulder, CO: Westview Press.

Kahneman, D. (2000). A psychological point of view: Violations of rational rules as diagnostic of mental processes (Reply to Stanovich & West). *Behavioral and Brain Sciences, 23*, 681-683.

Kahneman, D., & Tversky, A. (1979). Prospect theory: Analysis of decision under risk. *Econometrica, 47*, 263-291.

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

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research*. Chicago: Rand McNally.

La Puma, J., & Lawler, E. F. (1990). Quality-adjusted life years: Ethical implications for physicians and policymakers. *Journal of the American Medical Association, 263*, 2917-2921.

Mackie, J. L. (1977). *Ethics: Inventing right and wrong*. New York: Penguin.

Markus, H. R., & Kitayama, S. (1991). Culture and self: Implications for cognition, emotion and motivation. *Psychological Review, 98,* 224–253.

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. (2005). Moral heuristics or moral competence? (Reply to Sunstein) *Behavioral and Brain Sciences, 28*, 557-558.

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. (2004). After objectivity: An empirical study of moral judgment. *Philosophical Psychology, 17*, 5-28.

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

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108,* 291–310.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know - verbal reports on mental processes. *Psychological Review, 84*, 231-259.

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.

Pidgen, C. R. (1993). Naturalism. In P. Singer (Ed.), *A companion to ethics* (pp. 421-431). Malden, MA: Blackwell Publishing.

Pizarro, D. A., & Uhlmann, E. L. (2005). Do normative standards advance our understanding of moral judgment? (Reply to Sunstein) *Behavioral and Brain Sciences, 28*, 558-559.

Prinz, J. (in press). *The emotional consruction of morals*. Oxford: Oxford University Press.

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

Raz, J. (1975). Reasons for action, decisions and norms. *Mind, 84*, 481-499.

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. Shanteau (Eds.), *Emerging perspectives on judgment and decision making* (pp. 165-200). Cambridge, UK: Cambridge University Press.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making, 54*, 263-277.

Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty, 5*, 49-61.

Ritov, I., & Baron, J. (1995). Outcome knowledge, regret, and omission bias. *Organizational Behavior and Human Decision Processes, 64*, 119-127.

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

Rohan, M. J. (2000). A rose by any name? The values construct. *Personality and Social Psychology Review, 4*, 255-277.

Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley.

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.

Singer, P. (1973). Triviality of debate over is-ought and definition of moral. *American Philosophical Quarterly, 10*, 51-56.

Singer, P. (1999). Review of 'Living high and letting die'. *Philosophy and Phenomenological Research, 59*, 177-181.

Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). Cambridge University Press.

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

Skitka, L. J. (2002). Do the means always justify the ends, or do the ends sometimes justify the means? A value protection model of justice reasoning. *Personality and Social Psychology Bulletin, 28*, 588-597.

Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: Another contributor to attitude strength or something more? *Journal of Personality and Social Psychology, 88*(6), 895-917.

Smith, M. (1994). *The moral problem*. Oxford: Blackwell.

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

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences, 28*, 531-542.

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. (2005). Gauging the heuristic value of heuristics. (Reply to Sunstein) *Behavioral and Brain Sciences, 28*, 562-563.

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. (1976). Killing, letting die, and trolley problem. *Monist, 59*, 205-217.

Thomson, J. J. (1990). *The realm of rights*. Cambridge, MA: Harvard University Press.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*, 297-323.

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.

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior (2nd ed)*. Princeton, NJ: Princeton University Press.

Wakker, P. P. (1996). A criticism of healthy-years equivalents. *Medical Decision Making 16*, 207-214

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

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology, 4th ed.* (pp. 446-496). New York: McGraw-Hill.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe. *Psychological Science, 16*, 780-784.

Wilder, D. A. (1981). Perceiving persons as a group: Categorization and intergroup relations. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 213–257). Hillsdale, NJ: Erlbaum.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much - introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology, 60*, 181-192.

Tables

Table 1—Study 1: Proportion PVs Endorsed, Quantity Sensitivity as Expressed in Two Procedures by Presence/Absence of PV

	Ritov-Baron				Connolly-Reb			
Item	**%PV**	**Thresholds**		***t*-value (η_p^2)**	**%PV**	**Thresholds**		***t*** **(η_p^2)**
		No PV	**PVs**			**No PV**	**PVs**	
Starvation	0.78	0.75	0.57	1.14 (.04)	0.78	0.58	0.65	<1 (.01)
Cutting Forests	0.41	0.58	0.32	2.31* (.13)	0.46	0.55	0.75	2.08* (.11)
River Diversion	0.35	0.60	0.34	2.22* (.12)	0.41	0.47	0.66	2.04* (.11)

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

Table 2—Study 2: Effects of Condition, rated Moral Conviction, and their Interaction on Proportion Dominance.

	Group-Individual		**Interaction**
	***F*-value (η_p^2)**	**Conviction**	***F*-value (η_p^2)**
Anthrax	6.07* (.11)	< 1	3.92* (.07)
Otter	1.65 (.03)	< 1	4.46* (.09)
Paper	7.84** (.13)	< 1	5.02* (.09)
Tuna	5.31* (.09)	< 1	5.92* (.11)
Zaire	< 1	< 1	3.92† (.09)
Ps' Avg's	5.14* (.08)	< 1	11.02**(.17)

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

Table 3—Study 3: 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)

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

Table 4—Study 3: 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

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

Table 5—Study 4a: 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.

		Action		Hyp 4.1	Omission		Decision	PV	Interaction
Item	**%PV**	**NoPV**	**PV**	***t*-value(η_p^2)**	**NoPV**	**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
ThePoor	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)

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

Table 6—Study 4b: Study 4: 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.

		Action		Omission		Decision	PV		Interaction	
Item	**%PV**	**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)

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

Figures

Figure 1—Displays presented to participants in Study 2

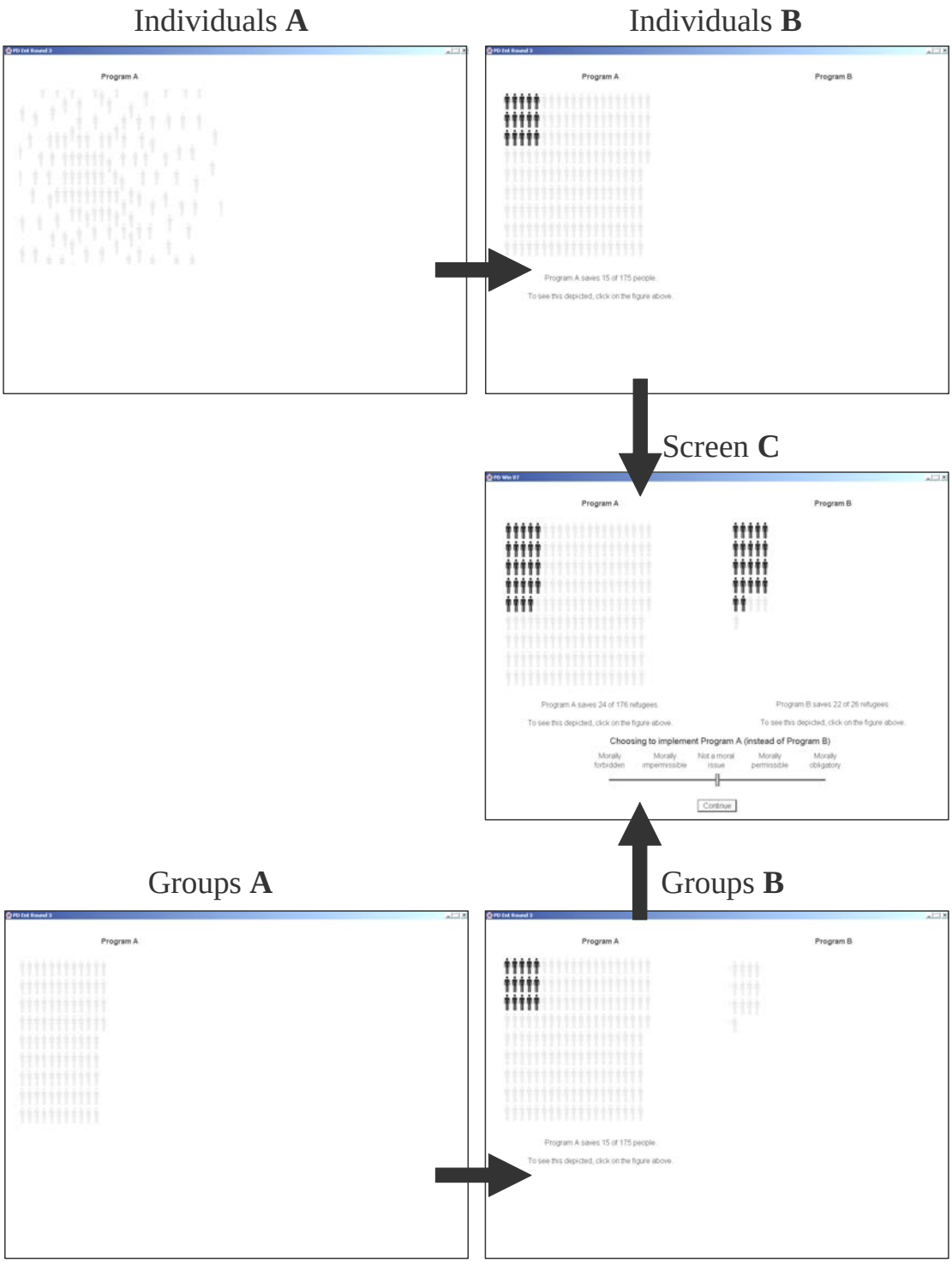


Figure 2— Study 2: Proportion Dominance by Condition (Group or Individual Construal) and rated Moral Conviction (median split)

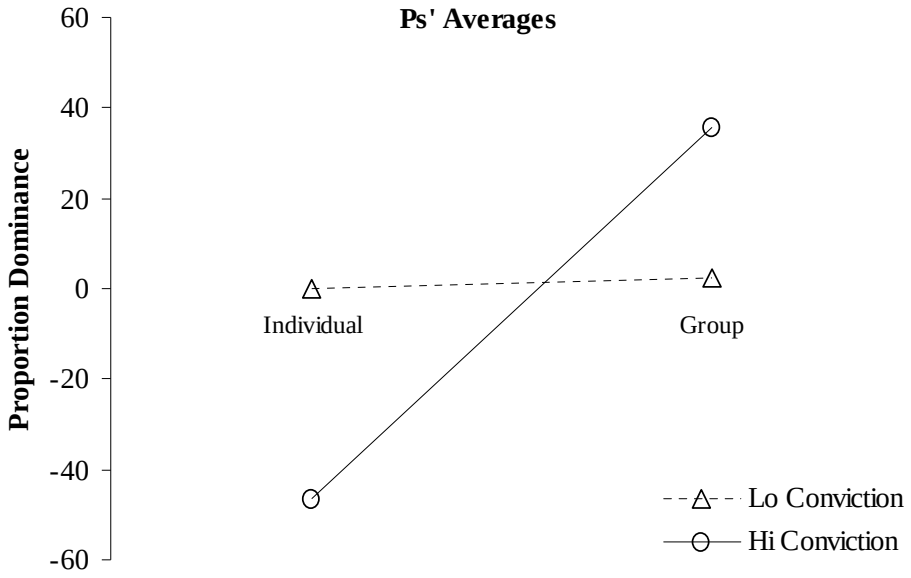


Figure 3—Study 3: Utilitarianism in Morally-Motivated Preference by Condition (Vivid, Standard, Catastrophe), Thinking Styles (Intuitive, Deliberative), and Idealism

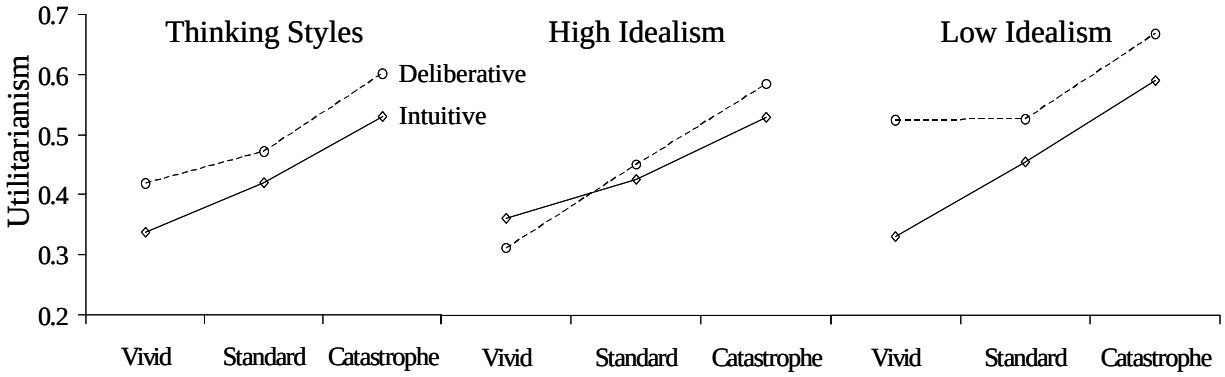
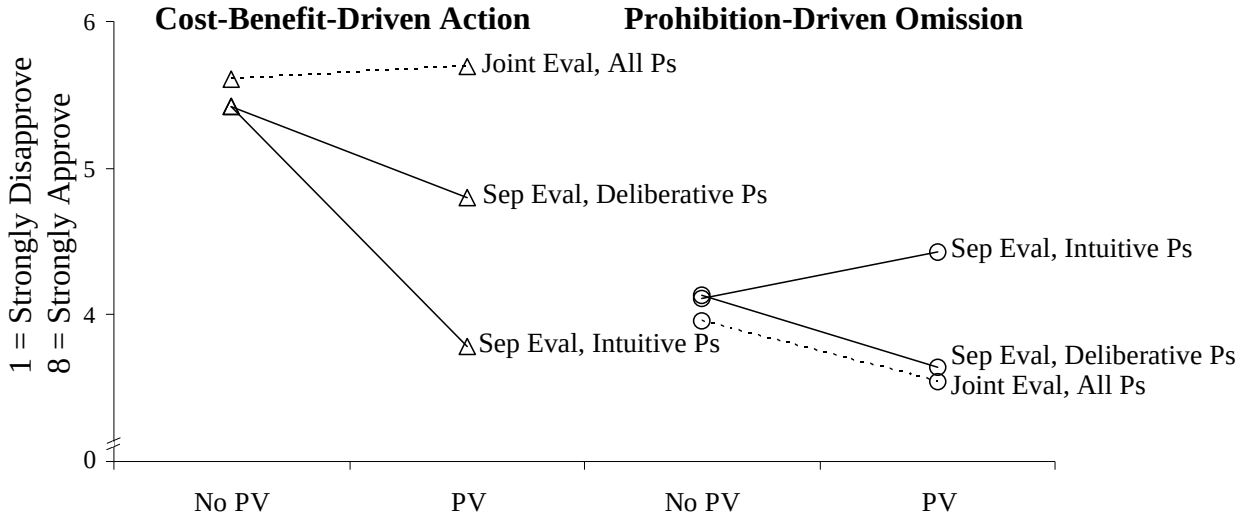


Figure 4—Study 4: Approval ratings as a function of Decision (Action vs. Omission) and presence or absence of PVs. Results are presented separately as a function of Thinking Styles (Intuitive vs. Deliberative) and Separate Evaluation (Study 4a) versus Joint Evaluation (Study 4b)



Appendix A: Study 2 Stimuli

Anthrax. Anthrax powder has been weaponized and released into the air above two mid-sized cities. In each city, a number of people are expected to die as a result of anthrax inhalation. There exists a powerful antibiotic that will successfully treat some victims, but there is a limited amount of this treatment. Program A would delegate the treatment to City A, and 15 of the 175 at risk of death will be saved. Program B would delegate the treatment to City B, and 14 of the 17 people at risk of death will be saved. These programs are mutually exclusive and the only two options available.

Otters. An oil spill around Puget Sound is threatening the sea otter populations in two areas of the bay. Two cleanup plans are proposed, but there is only enough money to support one plan. So, there are only enough resources to save otters in one of these areas of the bay. Program A will save 24 of the 171 otters near the north end of the bay. Program B will save 22 of the 26 otters near the south end of the bay. These programs are mutually exclusive and the only two options available.

Paper. You are on a committee at a major paper company with two factories on a mid-sized river. These factories use water from the river to cool their machines. Once used, the water is exhausted back into the stream. This polluted water causes a number of fish to die every year near the factory from which it is exhausted. Filters can be installed that will save a number of fish, but filter installation is expensive, and there is only enough money in the budget to install filters at one factory. Program A filters the water exhausted from Factory A, resulting in the prevention of 25 of the annual 182 fish deaths due to pollution. Program B filters the water exhausted from Factory B, preventing 24 of the annual 39 fish deaths due to pollution. These programs are mutually exclusive and the only two options available.

Tuna. Two areas off the southeast coast of Florida are heavily populated with dolphins and tuna. Tuna fishermen accidentally catch a number of dolphins in these areas every year. Dolphins that get caught in the tuna nets drown, because they cannot surface to breathe. To combat this

problem, new nets have been designed that will save a number of dolphins. The tuna fishing industry has agreed to fish with the new nets in only one of these two areas. Program A would require boats in Area A to use a different type of net, which would save 15 of the 28 dolphins that die in that area each year. Program B would require boats in Area B to use a different type of net, which would save 13 of the 17 dolphins that die in that area each year. These programs are mutually exclusive and the only two options available.

Zaire. Recent political developments in Zaire have severely marginalized some of the population. These refugees are clustered about in two camps, struggling to survive, because very little clean water is available. A plane with water treatment capabilities will be sent. There is only enough fuel, supplies, and time to visit one camp. Program A would treat enough water to save 24 refugees in the camp of 176. Program B would treat enough water to save 22 refugees in the camp of 26. These programs are mutually exclusive and the only two options available.

Appendix B: Individual differences items used in Study 3

(Note: Rational-Experiential Inventory also used in Study 4a)

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.

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 C: Ethical Dilemmas from Study 3

(Note: Catastrophe manipulation in parentheses, vividness manipulation appears in bold at end of item.)

Footbridge. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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. 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 cellmate's body writhing violently as he hangs.**

In this situation, would you remove the chair?

Soldiers. 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 Study 4

(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 southeast 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.