November 2008

Better Safe than Sorry: Precautionary Reasoning and Implied Dominance in Risky Decisions

Michael L. DeKay
Carnegie Mellon University

Dalia Patino-Echeverri
Carnegie Mellon University

Paul S. Fischbeck
Carnegie Mellon University, pf12@andrew.cmu.edu

Better Safe than Sorry:

Precautionary Reasoning and Implied Dominance in Risky Decisions

Michael L. DeKay[1,2], Dalia Patiño-Echeverri[1], and Paul S. Fischbeck[1,3]

[1]*Department of Engineering and Public Policy, Carnegie Mellon University, USA*

[2]*H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, USA*

[3]*Department of Social and Decision Sciences, Carnegie Mellon University, USA*

Address correspondence to:

Michael L. DeKay
H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213
Phone: 412-268-1877
Email: dekay@adrew.cmu.edu

Sponsors:

Natural Hazards Center, University of Colorado (as part of National Science Foundation grant number BCS–9213479 to the Natural Hazards Center)
Center for the Study and Improvement of Regulation, Carnegie Mellon University
National Science Foundation (grant number SES-0112005)

Running head: Better Safe than Sorry

Request to revise and resubmit received from *Journal of Behavioral Decision Making*

ABSTRACT

In four studies, student and nonstudent participants evaluated the possible outcomes of binary decisions involving health, safety, and environmental risks (e.g., whether to issue a dam-failure evacuation order). Many participants indicated that false positives (e.g., evacuation, but no dam failure) were better than true negatives (e.g., no evacuation and no dam failure), thereby implying that an incorrect decision was better than a correct one, and that the more protective action dominated the less protective action. A common rationale for this response pattern was the precautionary maxim "Better safe than sorry." Participants apparently evaluated outcomes partly on the basis of the decisions that might lead to them, in conflict with consequentialist decision models. Consistent with this explanation, the prevalence of implied dominance decreased substantially when the emphasis on decisions was reduced. These results demonstrate that initial preferences for decision alternatives can seriously bias the evaluation of consequences in risky high-stakes decisions.

KEY WORDS bias; consequentialism; dominance; information distortion; outcome evaluation; precautionary principle; risk; threshold

INTRODUCTION

Decision analysis and consequentialism

Evaluating the possible outcomes of decision alternatives is a key step in any decision analysis (Clemen, 1996; Raiffa, 1968; von Winterfeldt & Edwards, 1986). According to the standard procedure, the relative desirability or utility of each outcome is weighted by the probability that the outcome will occur, the weighted utilities are summed to yield an expected utility for each decision alternative, and the alternative with the highest expected utility is selected. Even if the specific dictates of utility theory are relaxed or rejected, broader *consequentialist* theories still require that decisions be based on assessments of the desirabilities and likelihoods of the possible outcomes (Frisch & Clemen, 1994). It goes without saying that the potential consequences should be considered and evaluated before the decision is actually made.

Behavioral complications

Despite the appeal of the expected-utility approach and of consequentialism more generally, research suggests that people often make decisions that are *nonconsequentialist* in the sense that they are affected by considerations other than consequences (Baron, 1994). For example, people sometimes make a distinction between action and inaction, even when the end result is ostensibly the same (e.g., death from vaccination versus death from the disease) (Baron & Ritov, 2004; Ritov & Baron, 1990, 1992; Spranca, Minsk, & Baron, 1991). One interpretation of these results is that some people have preferences for particular decision alternatives (e.g., forgoing vaccination) in addition to preferences for possible outcomes. A slightly different interpretation is that evaluations of consequences are based at least in part on the actions or decisions that lead to those consequences (e.g., a death from vaccination is actually worse than a death from the disease). Similarly, people view environmental damage as more serious when it results from human actions rather than from natural causes (Brown, Nannini, Gorter, Bell, & Petersen, 2002; Brown, Peterson, Brodersen, Ford, & Bell, 2005; Bulte, Gerking, List, & de Zeeuw, 2005; DeKay & McClelland, 1996; Kahneman & Ritov, 1994; Kahneman, Ritov, Jacowitz, & Grant, 1993). Evaluations of acts may also contaminate evaluations of consequences when decisions are based on *protected* or *sacred values* (Baron, 1997; Baron & Spranca, 1997; Tetlock, Kristel, Elson, Green, & Lerner, 2000), although it is difficult to discern whether participants with such values evaluate consequences differently from those without them or simply ignore consequences altogether.

Several other lines of research are consistent with the notion that evaluations of decision alternatives might affect evaluations of consequences. First, research on *outcome bias* indicates that people are often unable to ignore the actual consequences of a decision when evaluating the quality of the decision that led to those consequences (Baron & Hershey, 1988b; Hershey & Baron, 1992). Although this effect works in the opposite direction (evaluations of known outcomes affect evaluations of prior decisions), it nonetheless demonstrates that people have difficulty making clear distinctions between decisions and consequences.

Second, studies of *biased predecision processing* indicate that an initial or emergent preference for one alternative over another leads people to distort attribute information or importance in ways that are consistent with that preference (Russo, Medvec, & Meloy, 1996, Russo, Meloy, & Medvec, 1998; Russo, Meloy, & Wilks, 2000; Simon, Krawczyk, & Holyoak, 2004); that evidence in legal settings is distorted in favor of the tentatively preferred verdict (Carlson & Russo, 2001; Hope, Memon, & McGeorge, 2004; Simon, Pham, Le, & Holyoak, 2001); that task-irrelevant and unjustifiable factors influence judgments in favor of the preferred alternative, particularly in the presence of ambiguity (Hsee, 1995, 1996); and that the relative weights assigned to forecasts from different sources systematically favor the initially preferred alternative (Boiney, Kennedy, & Nye, 1997). Similarly, Nagel (1979) reported that people may alter their estimate of the probability that a defendant is guilty or their threshold probability for conviction in order to rationalize a desired verdict. These and other results indicate that decision makers' motivations affect the way in which information is gathered, evaluated, and processed (Brownstein, 2003; Kunda, 1990; Larrick, 1993).

Third, research on the *affect heuristic* indicates that people's initial affective reactions to hazards, gambles, and other stimuli have important implications for related judgments and decisions (Alhakami & Slovic, 1994; Finucane, Alhakami, Slovic, & Johnson, 2000; Slovic, Finucane, Peters, & MacGregor, 2002, 2004). For example, affective reactions appear to mediate the observed negative relationship between judgments of risk and benefit (Finucane et al., 2000). Results from these studies are largely consistent with the *risk-as-feelings* model of decision making (Lowenstein, Weber, Hsee, & Welch, 2001) and with *dual-process* models of information processing (Sloman, 1996, 2002; Stanovich & West, 2000, 2002).

Finally, Svenson (1992, 1996) and Montgomery (1983) have provided theoretical frameworks that address both the formation of initial preferences and the impact of those preferences on subsequent information processing. Svenson's model holds that preliminary choices are often made on the basis of holistic affective or emotional factors (Zajonc, 1980) or on the basis of only one or a few attributes. Subsequently, the decision maker restructures the alternatives and attributes with the goal of creating an alternative that appears sufficiently superior to its competitors, with restructuring being biased in favor of the preliminary choice. Similarly, Montgomery has proposed that the decision maker selects a promising alternative by very simple decision rules or affective reactions before attempting "to create dominance by changing his or her representation of the decision situation" (p. 344). Montgomery's restructuring mechanisms are also biased toward the initially promising alternative, and allow for the elimination of obvious tradeoffs among important attributes so that dominance or near-dominance may be obtained.

In summary, a wealth of theory and evidence suggests that preferences among alternatives may be formed very early in the decision process, and that these initial preferences bias subsequent information processing. In this article, we investigate whether such biases extend to the possible outcomes of risky decisions. Specifically, do initial preferences in decisions involving health, safety, and environmental risks bias evaluations of the possible consequences of those decisions? Are such biases are strong enough to imply that one course of action dominates the other?

Issuing a dam-failure evacuation order

As a motivating example, imagine that a particular dam is at some risk of failing and that the decision maker must decide whether or not to evacuate the people who are downstream. In this simplified world, there are two possible states of nature (the dam fails or it does not) and two decision alternatives (an evacuation is ordered or it is not). There are four possible outcomes: a true positive (an evacuation is ordered and the dam subsequently fails); a false positive (an evacuation is ordered, but the dam does not fail); a true negative (an evacuation is not ordered and the dam does not fail); and a false negative (an evacuation is not ordered, but the dam fails). The desirabilities or utilities of these outcomes may be denoted U_{TP}, U_{FP}, U_{TN}, and U_{FN}, respectively. We now consider the decision-analysis and behavioral-decision-theory perspectives on this problem.

The decision-analysis perspective

If the dam fails, it is much better to have evacuated the people downstream, so $U_{TP} > U_{FN}$. If the dam does not fail, it is better to have avoided the cost and inconvenience associated with an evacuation, so $U_{TN} > U_{FP}$.

If P is the subjective probability that the dam will fail, the subjective expected utilities of the two alternatives are as follows:

$$SEU(Evacuation) = P \times U_{TP} + (1 - P) \times U_{FP}$$

$$SEU(No\ Evacuation) = P \times U_{FN} + (1 - P) \times U_{TN}$$

Equating these two expressions and solving for *P* yields an expression for the threshold probability *P**:

$$P^* = \frac{U_{TN} - U_{FP}}{(U_{TN} - U_{FP}) + (U_{TP} - U_{FN})} = \frac{1}{1 + \frac{U_{TP} - U_{FN}}{U_{TN} - U_{FP}}}$$

An evacuation should be ordered if *P* exceeds *P**, otherwise not. This situation is depicted in the left panel of Figure 1, in which *P** is very low because $U_{TP} - U_{FN}$ is much greater than $U_{TN} - U_{FP}$. Indeed, thresholds may be 0.01, 0.001, or lower for some dams in the Western U.S., because a failure may lead to many fatalities if not preceded by an adequate warning (DeKay & McClelland, 1991, 1993).

Insert Figure 1 about here.

This simple threshold model has been used to assess the proper course of action in numerous decision domains, from initiating medical treatments (Pauker & Kassirer, 1975; Sox, Blatt, Higgins, & Marton, 1988) to setting standards of proof in civil and criminal trials (Connolly, 1987; Cullison, 1969; DeKay, 1996; Kaplan, 1968). Recently, DeKay et al. (2002) proposed *P** as a measure of precaution that might be useful in linking decision analysis with approaches based on the *precautionary principle* (Commission of the European Communities, 2000; O'Riordan & Cameron, 1994; Raffensperger & Tickner, 1999; Wingspread Conference, 1998), with lower thresholds reflecting greater precaution.[1]

[1] There are several formal definitions of the precautionary principle, including this popular one from the Wingspread Conference (1998): "When an activity raises threats of harm to the environment or human health, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically."

The behavioral-decision-theory perspective

The behavioral literature cited above suggests that people may deviate substantially from the decision-analysis approach. First, they may make a preliminary choice based on a holistic assessment of the decision situation or on a quick assessment of the most salient outcome (a false negative). In the case of a possible dam failure, it seems likely that evacuation will be preferred initially, because this course of action avoids a false negative with certainty. If the possible outcomes of the decision are then evaluated, these evaluations are expected to be biased in a manner consistent with the preferred alternative. Specifically, the two outcomes that might follow from evacuating (a true positive and a false positive) are expected to be evaluated more positively than they would be otherwise, whereas the two outcomes that might follow from not evacuating (a true negative and a false negative) are expected to be evaluated more negatively than they would be otherwise. (Alternatively, the evaluation of a false negative might be considered an integral part of the initial preference.) Thus, $U_{TP} - U_{FN}$ is expected to be larger, $U_{TN} - U_{FP}$ is expected be smaller, and P^* is expected to be lower when there is an initial preference for evacuation than when there is not. In more extreme cases, a false alarm might be viewed as better than a true negative, thereby implying that the more protective action (evacuation) strictly dominates the less protective action (no evacuation), as in the right panel of Figure 1.[2] In other words, an incorrect decision might seem better than a correct decision when the adverse event does not occur, because the incorrect decision corresponds to an action that is preferred on other grounds.

Overview of Studies

We investigated aspects of the above behavioral model in a series of four studies, all of which

[2] When $U_{TN} < U_{FP}$, the expression for P^* is no longer valid.

involved assessments of the relative desirabilities of the four possible outcomes of binary decisions. The scenarios used were similar to the dam-failure evacuation decision in that a false negative seemed particularly undesirable and the more protective action seemed like the most plausible initial preference. All of the scenarios involved health, safety, or environmental risks and many were modeled after contemporary high-stakes public-policy problems.

In Studies 1 and 2, participants considered several scenarios, rated the desirabilities of the four possible outcomes of each decision, and provided explanations if their rating patterns implied that the more protective action dominated the less protective action. In Studies 3 and 4, we attempted to lower the prevalence of implied dominance by reducing the emphasis on decisions and (hence) reducing the tendency for participants to evaluate possible outcomes in a manner consistent with their initially preferred alternatives. Evidence from these studies clearly indicates that evaluations of potential consequences often follow the selection of a preferred alternative, in violation of the basic consequentialist principle that decisions should be based on the consideration of possible outcomes.

These four studies spanned many years. Studies 1 and 3 were conducted in 1992 and 1993, respectively, whereas Studies 2 and 4 were conducted in 2000 and 2002, respectively. These dates are important because some scenarios involved topical issues (e.g., AIDS, "mad cow disease," airport security) that were particularly salient at the time. Replication in different years, with different scenarios, and with different participant populations (students in the earlier studies and members of the general public in the later studies) enhances the generalizability of our key results, even if specific scenarios might be viewed differently today.

STUDIES 1 AND 2

Method

Participants

Sixty-seven undergraduates enrolled in critical-thinking or introductory psychology courses at the University of Colorado received course credit for participating in Study 1. Thirty-eight of these students (57%) were female. No other demographic information was collected.

Thirty-two members of the general public from the Pittsburgh, PA area were paid $20 for participating in Study 2. One participant was dropped for completing less than half of the questionnaire. The remaining participants were 21 to 63 years old (M = 43), 22 (71%) were female, 5 (17%) were African American, 23 (77%) were white, 25 (81%) had additional education after high school, and 17 (55%) had at least a college education.

Materials and Procedures

In Study 1, participants considered 9–15 scenarios involving a variety of risky decisions. For comparability, we focus on the following eight scenarios that were considered by every participant: (A) closing an airport because of severe thunderstorms, (B) diverting one's private plane to another airport because of severe thunderstorms, (C) issuing a dam-failure warning, (D) responding to a dam-failure warning, (E) licensing an AIDS drug that might be successful but has side effects, (F) taking an AIDS drug that might be successful but has side effects, (G) setting the threshold for detection of an accidental toxic-chemical release, and (H) responding to the faint odor that might be from a toxic chemical.

In Study 2, participants considered two of the original scenarios (D and G), plus six additional scenarios, most of which were modeled on situations in which precautionary principle has been called into play: (I) banning genetically modified crops and foods; (J) banning blood donations from people who have visited England, because of the risk of "mad cow disease"; (K) avoiding the sequestration of carbon dioxide in the deep ocean or in geological formations (as a means of slowing global climate change), because of risks to people and animals; (L) avoiding beef and milk from hormone-treated cows; (M) banning the use of cellular phones by automobile drivers; and (N) canceling the launch of a spacecraft that uses a plutonium power source for instrumentation.

In each study, participants read short vignettes describing each scenario and brief descriptions of the four possible outcomes. For example, Scenario L was described as follows:

> Companies have developed hormones that increase the growth rate of beef cattle and the milk production of dairy cattle. These increases result in lower beef and milk prices for consumers. Some consumer and environmental groups have argued that eating beef or drinking milk from hormone-treated cows may increase a person's risk of getting cancer, but the evidence is inconclusive. Imagine that the U.S. Food and Drug Administration requires that all beef and milk from hormone-treated cows be labeled "This product comes from cows treated with hormones." Also imagine that you are married and have two children, and that you are the person who does the grocery shopping for your family. You must decide whether to buy beef and milk from hormone-treated cows or from "hormone-free" cows. Beef and milk from hormone-free cows cost about 10% more.
>
> *Outcome A*. You decide to buy beef and milk from hormone-treated cows. Further studies indicate that beef and milk from hormone-treated cows do increase the risk of cancer in humans. [False negative]
>
> *Outcome B*. You decide to buy beef and milk from hormone-treated cows. Further

studies indicate that beef and milk from hormone-treated cows do *not* increase the risk of cancer in humans. [True negative]

Outcome C. You decide to avoid beef and milk from hormone-treated cows and instead buy beef and milk from hormone-free cows. Further studies indicate that beef and milk from hormone-treated cows do increase the risk of cancer in humans. [True positive]

Outcome D. You decide to avoid beef and milk from hormone-treated cows and instead buy beef and milk from hormone-free cows. Further studies indicate that beef and milk from hormone-treated cows do *not* increase the risk of cancer in humans. [False positive]

In all scenarios, true positives and false positives were the possible results of the more protective action (avoidance in Scenario L), whereas true negatives and false negatives were the possible results of the less protective action. These labels were not shown to participants.

Participants wrote down the most important aspects of each outcome and rated the desirability of each outcome on a scale where –100 was the worst of the four outcomes and 0 was an outcome that was neither good nor bad (the scale was described but not shown, because it was unbounded on the upper end). The last 32 participants in Study 1 were asked to provide explanations if they rated Outcome D (a false positive) as better than Outcome B (a true negative).[3] In Study 2, participants who rated these outcomes as equal also provided explanations. All participants then completed statements like the following to indicate how sure they would have to be before taking the more protective action: "The odds that beef and milk from hormone-free cows increase the risk of cancer in humans would have to be greater than _____ out of _____ before I would avoid beef and milk from hormone treated cows and instead

[3] Study 1 was initially designed with a slightly different purpose in mind. Rating patterns indicating dominance were noticed during preliminary analyses, and procedures were altered to elicit more detailed explanations.

buy beef and milk from hormone-free cows." Participants answered all questions for one scenario before moving on to the next.

At the end of the questionnaire, participants in Study 2 rated the extent to which the benefit-cost principle and the precautionary principle should apply to each of the eight scenarios, using separate 7-point scales for the two principles. The benefit-cost principle was described as "When the available information does not imply a clear course of action, it is better to weigh the likely benefits of taking action against the likely costs and risks of taking action." The precautionary principle was described as "When the available information does not imply a clear course of action, it is better to err on the side of caution." This definition was intended to convey the gist of the precautionary principle in very simple terms (see footnote 1). Finally, participants in both studies rated the extent to which they were risk seeking or risk averse in their personal lives, again using a 7-point scale.

Results and Discussion

Implied Dominance from Ratings of Outcomes

The desirability ratings of the four possible outcomes may be considered utilities. Of the 536 sets of ratings in Study 1 (67 participants × 8 scenarios), 530 were complete. Of these, 396 sets (75%) implied tradeoffs consistent with the threshold model depicted in the left panel of Figure 1 (i.e., $U_{TN} > U_{FP}$ and $U_{TP} > U_{FN}$) and 124 sets (23%) implied that the more protective action weakly dominated the less protective action (i.e., $U_{TN} \leq U_{FP}$ and $U_{TP} \geq U_{FN}$, with at least one inequality). The remaining 10 sets of ratings (2%) implied other decision models, usually that the less protective action was dominant (9 sets).

Of the 248 sets of ratings in Study 2 (31 participants × 8 scenarios), 243 were complete. Of these, 90 sets (37%) implied tradeoffs consistent with the threshold model and 126 sets (52%)

implied that the more protective action dominated the less protective action. The remaining 27 sets of ratings (11%) implied other decision models, usually that the less protective action was dominant (20 sets). These other patterns are ignored in the analyses below.

Remarkably, participants indicated that the more protective action dominated the less protective action about one-quarter to one-half of the time, depending on the study. In the vast majority of these cases (113 of 124 cases in Study 1 and 97 of 126 cases in Study 2), the more protective action strictly dominated the less protective action, as in the right panel of Figure 1 (i.e., $U_{TN} < U_{FP}$ and $U_{TP} > U_{FN}$). In other words, false positives were often considered to be better than (and not merely equal to) true negatives.

As might be expected, the tendency to indicate that the more protective action was dominant was greater for some scenarios than for others, as seen in Figures 2 and 3. Following procedures described by Allison (1999, chapter 8), we conducted repeated-measures logistic regressions to predict weak and strict dominance of the more protective action as functions of scenario and risk aversion. In Study 2, we also included participants' ratings of the appropriateness of the precautionary principle and the benefit-cost principle for each scenario as predictors.

Insert Figures 2 and 3 about here.

In Study 1, the frequency of weak dominance varied significantly with scenario, $p < 0.0001$ (see Table 1 for more detailed statistics). In addition, more risk-averse participants appeared more likely to exhibit weak dominance, although the effect was not quite significant, $p = 0.0570$. The results for strict dominance were essentially identical.

In Study 2, weak dominance did not vary significantly with scenario, $p = 0.1999$. Weak dominance was more likely when participants indicated that the precautionary principle was more appropriate, $p = 0.0118$, less likely when they indicated that the benefit-cost principle was more appropriate, $p = 0.0425$, and more likely for more risk-averse participants, $p = 0.0395$. The directions of these three effects are sensible. When strict dominance was considered, the effect of scenario was significant, $p = 0.0364$. The benefit-cost principle remained a significant predictor, $p = 0.0285$, but the precautionary principle only approached significance, $p = 0.0750$, and risk aversion was no longer significant, $p = 0.2357$. Separate analyses indicated that participants rated the precautionary principle as more appropriate than the benefit-cost principle for the scenarios in Study 2, $M = 4.90$ and $M = 3.08$, respectively, on 0–6 scales, $t(30) = 5.02$, $p < 0.0001$, $\eta^2 = 0.465$. This difference did not vary significantly with scenario in a repeated-measures ANOVA, $F(7, 175) = 1.30$, $p = 0.2551$, $\eta^2 = 0.05$, and was significant for each of the eight scenarios.

Insert Table 1 about here.

The design of Study 1 allows for a slightly more detailed look at the effects of scenario. In that study, there were four pairs of scenarios (AB, CD, EF, and GH), with a private and a public scenario involving the same topic in each pair. The tendency to view the more protective action as weakly dominant appeared to be greater for private scenarios than for public scenarios, although the difference was not significant, $OR = 1.35$, 95% $CI = 0.96$–1.96, Wald $\chi^2(df = 1, n = 67) = 2.62$, $p = 0.1054$. There was significant variation by scenario pair, Wald $\chi^2(df = 3, n = 67) = 46.74$, $p < 0.0001$, and the private-public difference was larger for some pairs of scenarios than for others, Wald $\chi^2(df = 3, n = 67) = 17.42$, $p = 0.0006$ for the interaction. Figure 1 shows that weak dominance was more common in the private version than the public version for three scenario pairs (AB, CD, and EF), but the reverse was true for scenario pair GH, which involved a potential chemical leak. The results for strict dominance were very similar, except that the private-public distinction was significant, *OR* = 1.52, 95% *CI* = 1.01–2.29, Wald $\chi^2(df = 1, n = 67) = 4.04$, $p = 0.0445$.

Explanations for Ratings that Implied Dominance

When asked to explain why they rated Outcome D (a false positive) as just as good or better than Outcome B (a true negative), or when asked to describe the most important aspects of those outcomes, many participants provided evaluations of the decision alternatives rather than the possible outcomes:

"You took precaution & instead of lucking out/made correct decision." [Study 1, Scenario A]

"It's better not to take this big risk." [Study 1, Scenario C]

"It's better to have tried." [Study 1, Scenario F]

"Better to evacuate than not when working with toxic chemicals." [Study 1, Scenario H]

"You have taken proper cautionary measures, although you may have wasted some time." [Study 2, Scenario D]

"Outcome D is the prudent & wise decision only a fool would stay once enformed [*sic*] of the possible danger." [Study 2, Scenario D]

"Outcome D was the correct choice because no one will ever get hurt." [Study 2, Scenario G]

"It would have been taking a chance not to ban it." [Study 2, Scenario J]

"Right decision—the risk was too great." [Study 2, Scenario L]

"Plutonium is nothing to take a chance on." [Study 2, Scenario N]

These reasons suggest that participants were unable to draw clear distinctions between outcomes and decisions, as in research on outcome bias (Baron & Hershey, 1988b; Hershey & Baron, 1992). Consistent with the behavioral model outlined earlier, many participants appeared to base their evaluations of outcomes partly on the decisions that might lead to those outcomes.

In addition, many participants made clear references to features of the worst possible outcome (a false negative):

"It's better to be safe than sorry (in this case)." [Study 1, Scenario B]

"Better to have chosen lower threshold in case of leak occurring." [Study 1, Scenario G]

"It's better to be interrupted than die." [Study 1, Scenario G]

"I would rather take the time to evacuate than to be sorry in the end." [Study 1, Scenario H]

"It's better to be safe than sorry—so D is better!" [Study 2, Scenario D]

"It's smarter to be safe when unsure than possibly get cancer." [Study 2, Scenario L]

"Ban can later be changed. Death can't." [Study 2, Scenario J]

Being sorry, a leak occurring, getting cancer, and dying are all aspects of false negatives. That participants made such references when ostensibly explaining the difference between false positives and true negatives suggests that false negatives were regarded as particularly important for the decisions in these studies. Baron and Hershey (1988a), among others, have noted that people sometimes "minimize the probability of the worst kind of error, regardless of other parameters" (p. 259).

For cases in which outcome ratings implied strict dominance in Study 1 or weak dominance (including strict dominance) in Study 2, we coded participants' explanations with respect to the two elements noted above. The first code indicated whether the participant expressed a preference for one decision alternative when evaluating the outcomes. The second indicated whether the participant mentioned features of a false negative or a true positive (because a true positive is the same as avoiding a false negative, distinguishing between references to these two outcomes was occasionally difficult). The two codes were not mutually exclusive, as is evident in some of the examples above. Although "better safe than sorry" could reflect either category, we coded this phrase as reflecting aspects of a false negative (sorry) rather than as a preference for a decision alternative. We considered participants' responses to three questions about each scenario (aspects of a true negative, aspects of a false positive, and the comparison of a true negative and a false positive) as a single unit. From Study 1, we considered only the 32 participants who had been asked to make the comparison (for these participants, there were 53 sets of ratings that implied dominance).

In Study 1, preference for a decision alternative was mentioned in 33 of 53 cases (62%) and aspects of a false negative or true positive (usually a false negative) were mentioned in 21 cases (40%); 40 cases (75%) fit one or both categories. A version of the phrase "better safe than sorry" was mentioned in 15 cases (26%). In Study 2, preference for a decision alternative was mentioned in 54 of 124 cases (44%) and aspects of a false negative or true positive (usually a false negative) were mentioned in 37 cases (30%); 72 cases (58%) fit one or both categories. Only 8 cases (6%) included a version of the phrase "better safe than sorry." The Study-2 percentages were slightly higher when only strict dominance was considered (45%, 35%, 62%, and 8% respectively).

Relationships among Outcome Ratings

If participants were basing their evaluations of outcomes on the decisions that might lead to them, as the above explanations suggest, the proposed behavioral model implies specific relationships between outcome evaluations. Prior to these analyses, we normalized each participant's ratings by dividing them by the absolute value of his or her most extreme rating, so that the possible range of the normalized ratings was –1 to 1.

Although $U_{TP} - U_{FN}$ must be positive for participants exhibiting the standard threshold pattern in the left panel of Figure 1 and positive for participants exhibiting dominance,[4] the behavioral model predicts that the difference will be larger when dominance is observed than when the threshold pattern is observed (other response patterns were excluded from this analysis to avoid inflating the difference between differences). In Study 1, $U_{TP} - U_{FN}$ was larger for the weak-dominance pattern ($M = 1.14$) than for the threshold pattern ($M = 0.87$), $t(49) = 3.10$, $p = 0.0032$, $\eta^2 = 0.16$, as expected. In Study 2, the corresponding means were 1.50 and 1.22, $t(24) = 2.76$, $p = 0.0109$, $\eta^2 = 0.24$. Similar results were obtained when strict rather than weak dominance was considered, and in repeated-measures analyses that controlled for scenario and risk aversion.

When the $U_{TN} - U_{FP}$ difference is also treated as continuous, the behavioral model implies that $U_{TP} - U_{FN}$ should increase as $U_{TN} - U_{FP}$ decreases (or becomes more negative). In other words, $U_{TP} - U_{FN}$ and $U_{TN} - U_{FP}$ should be negatively correlated. We calculated the correlation between these two differences across the eight scenarios separately for each participant. In both studies, the mean correlation was significantly less than zero, $M(r) = -0.27$, $t(65) = -3.68$, $p = 0.0005$, $\eta^2 = 0.17$ in Study 1 and $M(r) = -0.47$, $t(30) = -4.62$, $p < 0.0001$, $\eta^2 = 0.42$ in Study 2.[5]

[4] Weak dominance would also be implied if $U_{TP} - U_{FN} = 0$ and $U_{TN} - U_{FP} > 0$,
[5] Correlations were transformed using Fisher's *Z* before averaging and conducting statistical tests. The resulting means were transformed back into *r*s.

All of these results are consistent with the proposed model.

Stated Thresholds when Dominance was Implied

When ratings of outcomes implied that the more protective action dominated the less protective action, participants should have reported that the threshold probability for taking the more protective action was zero (i.e., that they would always take the more protective action, regardless of the probability of the adverse event). Despite the fact that the last 48 participants in Study 1 and all of the participants in Study 2 received explicit instructions to report thresholds of zero in such cases, only three of the 250 response patterns that implied dominance were accompanied by stated thresholds of zero. Three of the 486 response patterns consistent with nonzero thresholds were also accompanied by stated thresholds of zero. It is tempting to interpret the former result (3/250) as a preference reversal in which different response modes were associated with different utility orderings. However, it seems more likely that, when asked for a decision threshold, participants did not understand how to indicate that they would always take the more protective action. In Study 4, we attempted to address this issue using a different question format.

Regardless of whether participants really believed that the more protective action dominated the less protective action, the explanations and statistics reported in the previous two sections still provide insight into how they evaluated possible outcomes.

Summary

In both studies, participants often rated false positives as more desirable than true negatives, thereby implying that the more protective action dominated the less protective action. Such responses varied by scenario, more so in Study 1 than in Study 2. Such variation is not surprising, and almost certainly reflects differences in the situations described. Although results were not always statistically significant, the frequency of implied dominance tended to be higher when participants viewed the precautionary principle as appropriate, lower when they viewed the benefit-cost principle as appropriate, and higher for more risk-averse participants. When implied dominance was observed, participants' explanations for their ratings of false positives and true negatives were very likely to include references to the decisions themselves or to the other outcomes (particularly false negatives) that might follow from those decisions. As expected, the $U_{TP} - T_{FN}$ difference was larger when the more protective action was dominant and when $U_{TN} - T_{FP}$ was smaller (or more negative).

These results are generally consistent with the behavioral model outlined earlier. The central roles of decision alternatives and false negatives in participants' explanations for $U_{TN} \leq U_{FP}$ rating patterns suggest that many participants may have been jumping to conclusions regarding the most appropriate decision alternative on the basis of holistic impressions or after considering only the worst outcome (a false negative). These participants apparently evaluated the remaining outcomes with respect to the alternatives that might lead to them, with such biases often resulting in rating patterns that implied dominance.

STUDIES 3 AND 4

Although the results of Studies 1 and 2 are consistent with the proposed behavioral model, the data are mostly descriptive. A stronger test of the model would involve an experimental manipulation of participants' initial preferences. In many studies of predecision processing, participants are randomly assigned to conditions designed to engender a preference for one of two relatively similar alternatives. For example, an initial preference for one backpack over another might be stimulated by noting that one is produced by a manufacturer who is an alumnus of the participant's university (Russo et al., 1998), or an initial preference for one job offer over another might be stimulated by nothing that the job is located in a fun part of town (Simon et al., 2004). Subtle manipulations of this type seemed unlikely to succeed for decisions, like those in Studies 1 and 2, in which initial preferences may be quite strong. Moreover, altering our scenarios to make them more neutral would destroy the very characteristics that make them interesting, and would make it very difficult for biases—even very large ones—to produce ratings patterns that implied dominance. We therefore chose a different approach.

In his review of nonconsequentialist decisions, Baron (1994, p. 6) asked "whether subjects faced with a description of the consequences, divorced from the decisions that led to them, would judge the consequences for goal achievement in a way that was consistent with their decisions." Thus, we might be able to prevent the formation of initial preferences for decision alternatives, and the biasing effects of such preferences, by having some participants evaluate the possible outcomes without having considered the actual decisions. We predicted that reducing the amount of attention that participants devote to the decisions would reduce the number of $U_{TN} \leq U_{FP}$ response patterns by rendering irrelevant the causal link between initial preferences and outcome evaluations. In order to evaluate this hypothesis, we created a *reduced-context* version of each scenario, as described below.

Method

Participants

Sixty-three undergraduates enrolled in introductory psychology courses at the University of Colorado received course credit for participating in Study 3. Twenty-four of these students (38%) were female. No other demographic information was collected.

Fifty-seven parents, relatives, and teachers associated with a parochial school in Pittsburgh, PA participated in Study 4. A donation of $30 per participant was made to the school. One participant was dropped for completing less than half of the questionnaire. The remaining participants were 22 to 76 years old (*M* = 45), 50 (91%) were female, 53 (96%) were white, 26 (47%) had additional education after high school, and 8 (15%) had at least a college education.

Materials and Procedures

We created reduced-context versions of the scenarios by eliminating the opening vignettes, by not placing participants in the role of decision maker, and by referring to "situations" rather than "outcomes." For example, the reduced-context version of Scenario L consisted entirely of the following situations:

> *Situation A*. Your family of four regularly consumes milk from "hormone-treated" cows. Milk from hormone-treated cows costs a bit less than milk from "hormone-free" cows. Milk from hormone-treated cows causes cancer in humans. [False negative]

> *Situation B*. Your family of four does **not** consume milk from "hormone-treated" cows and instead consumes milk from "hormone-free" cows. Milk from hormone-treated cows costs a bit less than milk from hormone-free cows. Milk from hormone-treated cows causes cancer in humans. [True positive]

> *Situation C*. Your family of four does **not** consume milk from "hormone-treated" cows and instead consumes milk from "hormone-free" cows. Milk from hormone-treated cows costs a bit less than milk from hormone-free cows. Milk from hormone-treated cows does **not** cause cancer in humans. [False positive]

> *Situation D*. Your family of four regularly consumes milk from "hormone-treated" cows. Milk from hormone-treated cows costs a bit less than milk from "hormone-free" cows.

Milk from hormone-treated cows does **not** cause cancer in humans. [True negative]

The salience of the decisions was substantially reduced in all scenarios, but it was difficult to completely eliminate all references to decisions. For example, in the chemical-plant scenario (G), removal of references to evacuation led some pilot participants to conclude that the workers were not in the plant because it was a holiday. The outcomes described in the full-context condition were always identical to the situations described in the reduced-context condition.

Participants considered versions of eight scenarios (A–H) in Study 3 and versions of six scenarios (D, I, J, L, M, and O) in Study 4. Scenarios K (sequestering carbon dioxide in the deep ocean or in geological formations) and N (canceling the launch of a spacecraft that uses a plutonium power source) were omitted from Study 4 because it was difficult to describe the outcomes without the full scenario context. Scenarios I and L were simplified to refer to genetically modified potatoes (rather than crops and foods) and milk (rather than beef and milk) from hormone-treated cows. Scenario O was a new scenario that involved evacuating an airport because a screening machine had failed a reliability check. In both studies, participants were randomly assigned to the full-context and reduced-context conditions.

In Study 3, outcomes (or situations) were evaluated in the same way as in Studies 1 and 2, with a separate desirability rating for each of the four outcomes. In Study 4, outcomes (or situations) were presented in a different order (A = false negative, B = true positive, C = false positive, and D = true negative, as in Scenario L above) and participants compared them more directly. Specifically, participants indicated the relative desirability of outcomes C and D on an 11-point scale with endpoints labeled *Outcome C is much more desirable than outcome D* and *Outcome D is much more desirable than outcome C* and with the midpoint labeled *Outcomes C and D are equally desirable*. Subsequent questions asked about the relative desirability of outcomes A and B and about the relative magnitude of the difference between outcomes A and B and the difference between outcomes C and D.

In Study 3, stated thresholds for taking the more protective action were assessed as in Studies 1 and 2. In Study 4, we used a different question format that made it easier for participants to indicate that one alternative dominated the other by checking (in Scenario D, for example) "I would evacuate, regardless of the chance of dam failure" or "I would **not** evacuate, regardless of the chance of dam failure." Arranged vertically between these two endpoints were 17 specific thresholds ranging from "1 out of 10,000 (0.01%)" to "9,999 out of 10,000 (99.99%)," centered around "50 out of 100 (50%)." In the full-context conditions of both studies, participants provided stated thresholds immediately after answering questions about possible outcomes, as in Studies 1 and 2. In the reduced-context conditions, stated thresholds were not obtained until after ratings of situations had been completed for all scenarios, because the threshold questions required the same decision vignettes as the full-context conditions.

Results and Discussion

Effects of Decision Context on Implied Dominance

In the full-context condition of Study 3, 38 of 256 sets of ratings (15%) implied that the more protective action weakly dominated the less protective action. In the reduced-context condition, 24 of 247 sets (10%) implied weak dominance. When only strict dominance was considered, the percentages were 10% and 4% in two conditions, respectively.

The pattern was very similar in Study 4, although the overall rate of dominance was higher. In that study, 89 of 163 sets of ratings (55%) implied that the more protective action weakly dominated the less protective action in the full-context condition, compared to 60 of 165 sets (36%) in the reduced context-condition. For strict dominance, these percentages were 45% and

25%, respectively.

We assessed the significance of these decreases with repeated-measures logistic regressions to predict weak and strict dominance of the more protective action as functions of context and scenario. Additional predictors included risk aversion and (in Study 4) participants' ratings of the acceptability of the precautionary and benefit-cost principles.

In Study 3, the effect of context on weak dominance approached significance, $p = 0.0837$ (see Table 2 for more detailed statistics). The effect of scenario was significant, $p = 0.0335$, as was the effect of risk aversion, $p = 0.0437$. However, when only strict dominance was considered, the effect of context was significant, $p = 0.0050$, whereas the effects of scenario and risk aversion were not, $p = 0.3442$ and $p = 0.2116$, respectively.

Figure 4 suggests that the effect of context depended on scenario. Dominance was less likely in the reduced-context condition for only five of the eight scenarios, with Scenario A (closing an airport because of severe thunderstorms) and Scenarios E and F (licensing or taking an AIDS drug that might be successful but has side effects) exhibiting the opposite pattern. We were not able to assess the significance of the context × scenario interaction when all eight scenarios were treated as separate, because there were no cases of weak dominance in the reduced-context condition for scenario B (diverting one's private plane to another airport because of severe thunderstorms). When scenarios were grouped into similar pairs (AB, CD, EF, and GH) to eliminate this zero, the effect of scenario pair was significant, Wald $\chi^2(df = 3, n = 63) = 14.96$, $p = 0.0018$, as was the context × scenario-pair interaction, Wald $\chi^2(df = 3, n = 63) = 8.03$, $p = 0.0455$. When scenarios were grouped as private (B, D, F, and H) versus public (A, C, E, and G), neither effect was significant, $OR = 1.41$, 95% $CI = 0.88$–2.26, Wald $\chi^2(df = 1, n = 63) = 2.07$, $p = 0.1502$ for the private-public distinction and $OR = 1.38$, 95% $CI = 0.54$–3.54, Wald

$\chi^2(df = 1, n = 63) = 0.45$, $p = 0.5018$ for the interaction with context. When only strict dominance was considered, there were more zeros (see Figure 4) and the private-public approach was the only feasible option. In that analysis, context remained significant, $OR = 2.82$, 95% CI = 1.19–6.68, Wald $\chi^2(df = 1, n = 63) = 5.55$, $p = 0.0185$, the private-public distinction was not significant, $OR = 1.23$, 95% CI = 0.65–2.31, Wald $\chi^2(df = 1, n = 63) = 0.40$, $p = 0.5274$, and the context × private-public interaction was significant, $OR = 3.55$, 95% CI = 1.00–12.60, Wald $\chi^2(df = 1, n = 63) = 3.88$, $p = 0.0495$.

Insert Figure 4 about here.

The results for Study 4 were much simpler (see Figure 5). For weak dominance, the effect of context was significant, $p = 0.0057$, as was the effect of scenario, $p < 0.0001$, but the context × scenario interaction was not, $p = 0.1204$ (see Table 2 for more detailed statistics). Dominance was less likely in the reduced-context condition for all scenarios except D (banning the use of cellular phones by automobile drivers), for which the rate was equal in the two conditions. The results for strict dominance were very similar: context and scenario were both significant, p = 0.0054 and $p < 0.0001$, respectively, but the context × scenario interaction was not, $p = 0.1174$. Strict dominance was less likely in the reduced-context condition for all scenarios except I (banning genetically modified potatoes), for which the rate was equal in the two conditions. Ratings of the precautionary principle, the benefit-cost principle, and participants' risk aversion were not significant predictors of dominance in either analysis. As in Study 2, participants rated the precautionary principle as being more appropriate than the benefit-cost principle, $M = 4.53$ and $M = 2.45$, respectively, $t(52) = 7.48$, $p < 0.0001$, $\eta^2 = 0.52$. This difference was significant for each of the six scenarios.

Insert Figure 5 about here.

In summary, reducing the decision context lowered the frequency of response patterns that implied that the more protective action dominated the less protective action, as expected. This effect of context was significant or nearly significant in all four of the analyses reported in Table 2.

Effects of Decision Context on Continuous Outcome Ratings

The behavioral model posited above suggests that the effects of decision context should be evident for continuous measures of outcome desirability, not just for dichotomous measures of dominance. We conducted repeated-measures ANOVAs to assess these hypothesized effects.

In Study 3, we first normalized each participant's ratings by dividing them by the absolute value of his or her most extreme rating, so that the possible range of the normalized ratings was –1 to 1. As predicted, normalized ratings of true positives and false positives were both significantly higher in the full-context condition than in the reduced-context condition, $F(1, 60) = 11.36$, $p = 0.0013$, $\eta^2 = 0.16$ for true positives and $F(1, 60) = 4.26$, $p = 0.0435$, $\eta^2 = 0.07$ for false positives (see Table 3 for means; for brevity, we do not report the effects of scenario and the context × scenario interaction). Contrary to expectations, ratings of true negatives were higher in the full-context condition than in the reduced-context condition, although not significantly so, $F(1, 60) = 2.11$, $p = 0.1519$, $\eta^2 = 0.03$. One explanation is that the worst outcome (almost always a false negative) was assigned a rating of –100 by design. If participants viewed false negatives as worse in the full-context condition, they may have expressed this by giving higher ratings to all of the other outcomes, including true negatives. The difference between true negatives and false positives, $U_{TN} - U_{FP}$, was smaller in the full-context condition than in the reduced-context condition, but the effect of context did not approach significance, $F < 1$. The difference between true positives and false negatives, $U_{TP} - U_{FN}$, was greater in the full-context condition, as expected, $F(1, 60) = 12.11$, $p = 0.0009$, $\eta^2 = 0.17$.

Insert Table 3 about here.

In Study 4, the outcomes of interest were compared directly. Ratings of the difference between true negatives and false positives, $U_{TN} - U_{FP}$, were significantly lower in the full-context condition than in the reduced-context condition, $F(1, 51) = 8.29$, $p = 0.0058$, $\eta^2 = 0.140$. Ratings of the difference between true positives and false negatives, $U_{TP} - U_{FN}$, were greater in the full-context condition, although the effect was not quite significant, $F(1, 51) = 3.90$, $p = 0.0536$, $\eta^2 = 0.07$. Both results were in the expected directions.

As in Studies 1 and 2, we assessed whether $U_{TP} - U_{FN}$ was larger when weak dominance was observed than when the threshold pattern was observed. In Study 3, $U_{TP} - U_{FN}$ was larger for the weak-dominance pattern ($M = 1.19$) than for the threshold pattern ($M = 1.02$), $t(32) = 2.38$, $p = 0.0236$, $\eta^2 = 0.15$, as expected. In Study 4, the means for $U_{TP} - U_{FN}$ were not significantly different, $t < 1$, in part because the difference between means could be calculated for fewer than half of the participants (many participants never exhibited the threshold pattern). As before, we calculated the correlation between $U_{TP} - U_{FN}$ and $U_{TN} - U_{FP}$ across scenarios separately for each participant. As predicted, the mean correlation was significantly less than zero in both studies,

$M(r) = -0.15$, $t(62) = -2.30$, $p = 0.0247$, $\eta^2 = 0.08$ in Study 3 and $M(r) = -0.37$, $t(47) = -4.31$, $p < 0.0001$, $\eta^2 = 0.28$ in Study 4.[6]

Thus, the patterns of outcome ratings in these two studies were generally consistent with the proposed model, although there were a few exceptions.

Effects of Decision Context on Decision Thresholds when Dominance Was Not Implied

The effects of decision context on the evaluation of outcomes (or situations) were sometimes sufficiently strong to lead to rating patterns that implied dominance, as noted above. When dominance was not implied, the effect of context may have still have lowered decision thresholds for taking the more protective action. We evaluated this prediction by calculating a threshold probability for each set of ratings that was consistent with the threshold model depicted in the left panel of Figure 1 (i.e., $U_{TN} > U_{FP}$ and $U_{TP} > U_{FN}$). Analyses were conducted after transforming the threshold probabilities using the logit transformation.

In Study 3, we conducted a repeated-measures ANOVA to assess the effect of context on the logit-transformed threshold probabilities. The effect of context was in the predicted direction, but was not significant, $F(1, 26) = 1.47$, $p = 0.2369$, $\eta^2 = 0.05$ (as before, we omit the effects of scenario and the context $\times$ scenario interaction for brevity). However, this analysis included only those 28 participants who had valid threshold probabilities for all eight scenarios. To get around this problem, we averaged the valid transformed threshold probabilities within participants, and regressed the resulting means onto context. As expected, the mean transformed thresholds were lower in the full-context condition than in the reduced-context condition, $t(62) = 2.28$, $p = 0.0028$, $\eta^2 = 0.14$. Transformed back into probabilities, the mean thresholds were 0.39 in the

[6] As before, we used Fisher's Z transformation. In study 4, one correlation of +1.0 and one correlation of –1.0 were excluded from the analysis, because the transformation is not defined for perfect correlations. Similar results were obtained without the transformation, and when these correlations were changed to ±0.98 prior to the transformation.

full-context condition and 0.52 in the reduced-context condition (see Table 3). When the logit transformation was not used, the mean calculated threshold probabilities were 0.42 and 0.51 in the two conditions, respectively.

In Study 4, the high prevalence of ratings that implied dominance precluded the repeated-measures approach. In the simpler analysis, the mean logit-transformed thresholds were not significantly different in the two conditions, $F < 1$. Transformed back into probabilities, the mean thresholds were 0.06 in the full-context condition and 0.13 in the reduced-context condition (see Table 3). When the logit transformation was not used, the mean calculated threshold probabilities were 0.29 and 0.39 in the two conditions, respectively. We suspect that the lack of a context effect for calculated thresholds in Study 4 reflects the smaller number of scenarios with valid thresholds for each participant (17 participants had no valid thresholds at all) and participants' difficulty in quantifying the magnitude of the difference between U_{TP} and U_{FN} relative to the magnitude of the difference between U_{TN} and U_{FP}.

Stated Thresholds when Dominance was Implied

In Study 3, only five of the 62 response patterns that implied that the more protective action dominated the less protective action (8%) were accompanied by stated thresholds of zero. Thirteen of the 435 response patterns consistent with nonzero thresholds (3%) were also accompanied by stated thresholds of zero. These frequencies are only slightly higher than the 1% rates reported earlier for Studies 1 and 2 combined.

The results of Study 4 were very different. Of the 149 response patterns that implied dominance, 117 (79%) were accompanied by answers indicating that the participant would always take the more protective action, regardless of the chance of the adverse outcome. Thus, when participants were asked in a more straightforward way to state their preferences, most of those whose outcome ratings implied dominance agreed with that implication.

Perhaps surprisingly, 64 of the 104 response patterns in Study 4 that were consistent with nonzero thresholds (62%) were also accompanied by answers indicating dominance. This high percentage is not as troubling as it initially appears, however, because 47 of those 64 cases (73%) occurred in the reduced-context condition. In Study 3, 12 of 13 such cases (92%) occurred in the reduced-context condition. Recall that the decision context was reduced only in that portion of the study in which participants rated situations (i.e., possible outcomes), and that the decision context was described fully just before participants indicated their preferred course of action. Thus, the apparent inconsistency seems to have resulted from participants in the reduced-context condition reporting dominance more often after learning the details of the scenarios.

We assessed this explanation statistically. In the reduced-context condition of Study 4, the frequency of weak dominance was 60/165 = 36% when inferred from ratings of situations, but 126/167 = 75% when assessed directly. In the full-context condition, these frequencies were 89/163 = 55% and 102/153 = 67%, respectively. Repeated-measures logistic regressions indicated that the increase was significant for the reduced-context condition, $OR = 5.32$, 95% $CI = 3.14$–9.01, Wald $\chi^2(df = 1, n = 56) = 38.77$, $p < 0.0001$, but not quite significant for the full-context condition, $OR = 1.63$, 95% $CI = 0.99$–2.67, Wald $\chi^2(df = 1, n = 56) = 3.74$, $p = 0.0531$. The interaction between context and dominance assessment (inferred from outcome ratings versus assessed directly) was significant, $OR = 0.31$, 95% $CI = 0.15$–0.63, Wald $\chi^2(df = 1, n = 56) = 10.21$, $p = 0.0014$, and remained so when scenario and interactions with scenario were included as predictors. These results provide additional evidence that the decision context was an important factor in determining whether participants viewed the more protective action as dominating the less protective action.

Summary

To recap, the results of Studies 3 and 4 were generally consistent with the proposed behavioral model. Except for ratings of true negatives in Study 3, ratings of outcomes and differences between outcomes were affected by decision context in the expected manner, with most tests being statistically significant (see Table 3). In both studies, these effects resulted in dominance being more likely in the full-context condition than in the reduced-context condition (see Table 2). When dominance was not observed, thresholds for taking the more protective action were lower in the full-context condition than in the reduced-context condition, although the difference was not significant in Study 4 (see Table 3). Taken together, these findings indicate that many participants based their evaluations of possible outcomes at least in part on the decision alternatives that might lead to those outcomes.

Participants in Study 3 (like those in Studies 1 and 2) appeared unable to express dominance when asked for a decision threshold. However, participants in Study 4, who were asked more directly, often indicated that the more protective action dominated the less protective action. Participants in the reduced-context condition of Study 4 were more likely to report dominance after the full decision context had been provided than before it had been provided.

GENERAL DISCUSSION

Evaluations of the possible outcomes of risky decisions are often influenced by preferences for decision alternatives, in conflict with consequentialist models of decision making. Three types of evidence support this conclusion. First, participants' explanations for rating a false positive as being as good or better than a true negative often referred to the preferred decision alternative or to the outcome that participants most wanted to avoid (a false negative). Second, the desirability difference between true positives and false negatives, $U_{TP} - U_{FN}$, was greater when the more protective action dominated the less protective action and was negatively correlated with the desirability difference between true negatives and false positives, $U_{TN} - U_{FP}$. Third, participants' evaluations of possible outcomes responded as expected when the emphasis on decisions was reduced, with fewer response patterns implying that the more protective action dominated the less protective action.

These results are consistent with those of previous studies indicating that an initial preference for a decision alternative biases the evaluation of information that should ideally be used as input to the decision process. The present studies extend that literature by demonstrating that such biases apply to the outcomes of risky decisions. These results have been corroborated in separate studies using methods more typical of research on biased predecision processing. In one study, for example, DeKay, Fischbeck, Patiño-Echeverri, and Hartman (2003) presented participants with scenarios like those used in Study 4 and assessed participants' initial strength of preference for a decision alternative. Participants then viewed one of five pieces of information (descriptions of the four possible outcomes, or the probability of the adverse event), evaluated the extent to which that information favored one alternative or the other, updated their preferred course of action, and moved on to the next piece of information. In 29 of 30 tests (6 scenarios × 5 pieces of information), the evaluation of information was significantly biased in the direction of participants' currently preferred decision alternatives.

In the studies reported here, participants' "reversals" of false positives and true negatives were particularly remarkable. Given that the adverse event did not in fact occur, participants indicated that it was better to have made an incorrect decision than a correct one! Although it might seem reasonable for participants to incorporate feelings of safety or risk into their evaluations of a false positive or a true negative, respectively, such feelings should become negligible as the probability of the adverse event approaches zero, thereby allowing the costs of a false positive to take precedence. One possible explanation for the observed ratings is that participants inferred the probability of the adverse event from the description provided in the scenario and based their evaluations of the possible outcomes on the risk that they would or would not bear prior to the resolution of uncertainty. If participants interpreted the task in this manner, rating a false positive as better than a true negative would make sense if the inferred probability of the adverse event was greater than the (nonzero) threshold probability for taking the more protective action. Evidence against this explanation comes from two sources. First, none of the participants' explanations for $U_{TN} \leq U_{FP}$ ratings mentioned the likelihood of the adverse event. Second, almost 80% of Study-4 participants with such rating patterns indicated that they would take the more protective action regardless of the probability of the adverse event. Thus, it does not appear that participants' outcome orderings were specific to inferred probabilities.

It is worth noting that the implied dominance observed in these studies cannot be explained by *regret theory* (Bell, 1982; Loomes & Sugden, 1982), *disappointment theory* (Bell, 1985; Loomes & Sugden 1986), or *decision-affect theory* (Mellers, 2000; Mellers, Schwartz, Ho, & Ritov, 1997; Mellers, Schwartz, & Ritov, 1999). According to regret theory, a person who chose the less protective action might feel regret if the adverse event came to pass. Incorporation of anticipated regret would increase the difference between the desirability of a true positive and a false negative ($U_{TP} - U_{FN}$ in our notation). Similarly, a person who chose the more protective action might feel regret if the adverse event did not occur, and anticipation of this regret would increase the difference between the desirability of a true negative and a false positive, $U_{TN} - U_{FP}$. If anticipated regret increased $U_{TP} - U_{FN}$ more than it increased $U_{TN} - U_{FP}$, as seems likely, this would lower the threshold P^* for taking the more protective action. However, there is no mechanism in regret theory for reducing or reversing the difference between the desirabilities of two outcomes that result from different actions (e.g., evacuating or not) when a given state of the world occurs (e.g., when the dam does not fail), and therefore no mechanism by which anticipated regret can lead to implied dominance.

According to disappointment theory, a person will feel disappointed if the outcome obtained is worse than the expectation for the chosen alternative, and elated if the outcome obtained is better than that expectation. The amount of disappointment or elation experienced is assumed to be a positive function of the difference between the obtained outcome and the expectation. Only elation is relevant to the evaluation of true negatives and false positives. If the expectations for the more protective action and the less protective actions are those given in the introduction and in the left panel of Figure 1, elation should be greater for true negatives than for false positives regardless of the probability of the adverse event, because the expectation of the less protective action decreases more steeply as a function of probability than does the expectation of the more protective action. Thus, consideration of elation serves only to increase $U_{TN} - U_{FP}$, not to reduce or reverse this difference. If disappointment increases $U_{TP} - U_{FN}$ more than elation increases $U_{TN} - U_{FP}$, then P^* will be lower, as in regret theory, but dominance will not be implied.

Decision-affect theory incorporates versions of regret theory and disappointment theory, and thus has similar implications for the evaluations of possible outcomes. Specifically, there is no mechanism for reducing or reversing the $U_{TN} - U_{FP}$ difference. Note that we do not claim that anticipated regret and disappointment play no role in the assessment of consequences, only that such effects cannot by themselves lead to implied dominance.

Other outcome comparisons might be more successful in explaining our results. In particular, true negatives might be considered "guilty by association" with false negatives, whereas true positives and false positives might be contrasted with false negatives and therefore rated more positively. However, such comparisons are fundamentally different from those involving regret and disappointment, because the directions of the adjustments rely on an existing preference for the more protective action. Indeed, this is just another way of saying that the evaluation of consequences depends in part on the choices that might lead to them.

An alternative explanation for the results of Studies 3 and 4 is that the reduced-context condition did not lower participants' awareness of decisions, but simply removed participants from the role of decision maker. If evaluations of consequences are allowed to depend on perspective, as has been suggested by Sen (1981, 1983) and others (e.g., Portmore, 2003), then one might expect these evaluations to be different in the full-context and reduced-context conditions. However, the source of such differences is crucial: it is assumed that outcome evaluations may be decomposed into the goodness of actions and the goodness of the "other" consequences that result from those actions, with only the former component varying with perspective (Sen, 1981, pp. 31–32). Thus, even if our manipulation affected only the role of the participant relative to the decision, the observed results still imply that evaluations of consequences follow in part from the evaluations of acts. Most decision theorists would call such reasoning nonconsequentialst, despite arguments to the contrary.

A final interpretation of our results is that many participants thought that consequentialist reasoning was simply not very appropriate for the scenarios at hand. In Studies 2 and 4, participants indicated that it was more appropriate to err on the side of caution than to weigh the benefits of taking action against the costs and risks of taking action. Such ratings were predictive of dominance in Study 2, but not in Study 4. In addition, explanations for dominance often indicated that participants wanted to "do the right thing" in the situations described. Thus, many participants appeared to espouse nonconsequentialist or *deontological* decision rules for choosing the preferred course of action. Such rules are based on duties, obligations, and the rights of others, and may be distinguished from those based on habits, customs, or imitation. Some evidence for rule following may be found in the literature on judgment and decision making (Baron, 1994; Baron & Spranca, 1997; March, 1994; Mellers, Schwartz, & Cooke, 1998; Tetlock et al., 2000). For example, Hunt and Vasquez-Parraga (1993) reported that managers' decisions to reward or punish salespeople's behavior are based more on deontological than consequentialist considerations, and Böhm and Pfister (2000) reported that the perceived causal structure of environmental risks determines whether action tendencies are based primarily on deontological or consequentialist judgments. Even so, Hastie's (2001, p. 664) conclusion that "clear models of nonconsequential decision processes are still needed" remains true. Miner and Petocz (2003) have recently provided useful guidance for mapping moral theories onto ethical decision problems.

In the present studies, however, rule following can "explain" only participants' initial preferences for the more protective action; it cannot by itself explain why such preferences bias outcome evaluations. Some other motivational mechanism, such a the desire for a clear or defensible decision, must underlie the resolution of the discrepancy between a simple rule (e.g., "better to evacuate than not when working with toxic chemicals") and the cost and inconvenience that often accompany precautionary action. In an odd way, the fact that some participants who espoused precautionary decision rules biased their evaluations of outcomes to make them consistent with those rules suggests that the participants believed consequences to be relevant after all. It remains to be seen whether "better safe than sorry" represents a strong commitment to precautionary action despite the consequences, or whether is it merely a strong preference that is weakly held and therefore amenable to challenge (Baron & Leshner, 2000).

ACKNOWLEDGEMENTS

This research was supported by a grant from the Natural Hazards Center at the University of Colorado (as part of National Science Foundation grant BCS–9213479 to the Natural Hazards Center), a grant from the Center for the Study and Improvement of Regulation at Carnegie Mellon University, and National Science Foundation grant SES-0112005. We are grateful to Gary McClelland and Brian Knudsen for assistance with data collection.

REFERENCES

Alhakami, A. L., & Slovic, P. (1994). A psychological study of the inverse relationships between perceived risk and perceived benefit. *Risk Analysis*, *14*, 1085–1096.

Allison, P. D., (1999). *Logistic regression using the SAS system: Theory and application*. Cary, NC: SAS Institute.

Baron, J. (1994). Nonconsequentialist decisions [with commentaries]. *Behavioral and Brain Sciences*, *17*, 1–42.

Baron, J. (1997). Biases in the quantitative measurement of values for public decisions. *Psychological Bulletin*, *122*, 72–88

Baron, J., & Hershey, J. C. (1988a). Heuristics and biases in diagnostic reasoning: I. Priors, error costs, and test accuracy. *Organizational Behavior and Human Decision Processes*, *41*, 259–279.

Baron, J., & Hershey, J. C. (1988b). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*, 569–579.

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.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*, 961–981.

Bell, D. E. (1985) Disappointment in decision making under uncertainty. *Operations Research*, *33*, 1–27.

Böhm, G., & Pfister, H. R. (2000). Action tendencies and characteristics of environmental risks. *Acta Psychologica*, *104*, 317–337.

Boiney, L. G., Kennedy, J., & Nye, P. (1997). Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes*, *72*, 1–24.

Brown, T. C., Nannini, D., Gorter, R. B., Bell, P. A., & Peterson, G. L. (2002). Judged seriousness of environmental losses: Reliability and cause of loss. *Ecological Economics*, *42*, 479–491.

Brown, T. C., Peterson, G. L., Brodersen, R. M., Ford, V., & Bell, P. A. (2005). The judged seriousness of an environmental loss is a matter of what caused it. *Journal of Environmental Psychology*, *25*, 13–21.

Brownstein, A. (2003). Biased predecision processing. *Psychological Bulletin*, *129*, 545–568.

Bulte, E., Gerking, S., List, J. A., & de Zeeuw, A. (2005). The effect of varying the causes of environmental problems on stated WTP values: Evidence from a field study. *Journal of Environmental Economics and Management*, *49*, 330–342.

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *7*, 91–103.

Clemen, R. T. (1996). *Making hard decisions: An introduction to decision analysis*. Belmont, CA: Duxbury.

Commission of the European Communities. (2000). *Communication from the Commission on the precautionary principle*, Brussels: Author. Retrieved June 23, 2005, from the European Union Web site: http://europa.eu.int/comm/dgs/health_consumer/library/pub/pub07_en.pdf

Connolly, T. (1987). Decision theory, reasonable doubt, and the utility of erroneous acquittals. *Law and Human Behavior*, *11*, 101–112.

Cullison, A. D. (1969). Probability analysis of judicial fact-finding: A preliminary outline of the subjective approach. *Toledo Law Review*, *1969*, 538–598.

DeKay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law & Social Inquiry*, *21*, 95–132.

DeKay, M. L., Fischbeck, P. S., Patiño-Echeverri, D., & Hartman, J. A. (2003, November). *Information distortion and utility reversals in precautionary reasoning*. Poster presentation at the annual meeting of the Society for Judgment and Decision Making, Vancouver, BC Canada.

DeKay, M. L., & McClelland, G. H. (1991). *Setting decision thresholds for dam failure warnings: A practical theory-based approach* (Tech. Rep. No. 328). Boulder: University of Colorado, Center for Research on Judgment and Policy.

DeKay, M. L., & McClelland, G. H. (1993). Predicting loss of life in cases of dam failure and flash flood. *Risk Analysis*, *13*, 193–205.

DeKay, M. L., & McClelland, G. H. (1996). Probability and utility components of endangered species preservation programs. *Journal of Experimental Psychology: Applied*, *2*, 60–83.

DeKay, M. L., Small, M. J., Fischbeck, P. S., Farrow, R. S., Cullen, A., Kadane, J. B., Lave, L. B., Morgan, M. G., & Takemura, K. (2002). Risk-based decision analysis in support of precautionary policies. *Journal of Risk Research*, *5*, 391–417.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risk and benefit. *Journal of Behavioral Decision Making*, *13*, 1–17.

Frisch, D., & Clemen, R. T. (1994). Beyond expected utility theory: Rethinking behavioral decision research. *Psychological Bulletin*, *116*, 46–54.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, *52*, 653–683.

Hershey, J. C., & Baron, J. (1992). Judgment by outcomes: When is it justified? *Organizatinal Behavior and Human Decision Processes*, *53*, 89–93.

Hope, L., Memon, A., & McGeorge, P. (2004). Understanding pretrial publicity: Predecisional distortion of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *10*, 111–119.

Hsee, C. K. (1995). Elastic justification: How tempting but task-irrelevant factors influence decisions. *Organizational Behavior and Human Decision Processes*, *62*, 330–337.

Hsee, C. K. (1996). Elastic justification: How unjustifiable factors influence judgments. *Organizational Behavior and Human Decision Processes*, *66*, 122–129.

Hunt, S. D., & Vasquez-Parraga, A. Z. (1993). Organizational consequences, marketing ethics, and salesforce supervision. *Journal of Marketing Research*, *30*, 78–90.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study of the headline method. *Journal of Risk and Uncertainty*, *9*, 5–38.

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science*, *4*, 310–315.

Kaplan, J. (1968). Decision theory and the factfinding process. *Stanford Law Review*, *20*, 1065–1092.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

Larrick, R. P. (1993). Motivational factors in decision theories: The role of self-protection. *Psychological Bulletin*, *113*, 440–450.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, E. S. (2001). Risk as feelings. *Psychological Bulletin*, *127*, 267–286.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–824.

Loomes, G., & Sugden, R. (1986). Disappointment and dynamic consistency in choice under uncertainty. *Review of Economic Studies*, *53*, 271–282.

March, J. G. (with Heath, C.). (1994). *A primer on decision making: How decision happen*. New York: Free Press.

Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin*, *126*, 910–924.

Mellers, B. A., Schwartz, A., & Cooke, A. D. J. (1998). Judgment and decision making. *Annual Review of Psychology*, *49*, 447–477.

Mellers, B. A., Schwartz, A., Ho, K., & Ritov, I. (1997). Elation and disappointment: Emotional responses to risky options. *Psychological Science*, *8*, 423–429.

Mellers, B. A., Schwartz, A., & Ritov, I. (1999). Emotion-based choice. *Journal of Experimental Psychology: General*, *128*, 332–345.

Miner, M., & Petocz, A. (2003). Moral theories in ethical decision making: Problems, clarifications and recommendations from a psychological perspective. *Journal of Business Ethics*, *42*, 11–25.

Montgomery, H. (1983). Decision rules and the search for a dominance structure: Towards a process model of decision making. In P. Humphreys, O. Svenson, & A. Vari (Eds.), *Analysing and aiding decision processes* (pp. 343–369). Amsterdam: North-Holland.

Nagel, S. (1979). Bringing the values of jurors in line with the law. *Judicature*, *63*, 189–195.

O'Riordan, T., & Cameron, J. (Eds.). (1994). *Interpreting the precautionary principle*. London: Earthscan.

Pauker, S. G., & Kassirer, J. P. (1975). Thereapeutic decision making: A cost-benefit analysis. *New England Journal of Medicine*, *293*, 229–234.

Portmore, D. W. (2003). Position-relative consequentialism, agent-centered options, and supererogation. *Ethics*, *113*, 303–332.

Raffensperger, C., & Tickner, J. (Eds.). (1999). *Protecting public health and the environment: Implementing the precautionary principle*. Washington, DC: Island Press.

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, MA: Addison-Wesley.

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making*, *3*, 263–277.

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

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*, 102–110.

Russo, J. E., Meloy, M. G., & Medvec, V. H. (1998). Predecisional distortion of product information. *Journal of Marketing Research*, *35*, 438–452.

Russo, J. E., Meloy, M. G., & Wilks, T. J. (2000). Predcisional distortion of information by auditors and salespersons. *Management Science*, *46*, 13–27.

Sen, A. (1982). Rights and agency. *Philosophy & Public Affairs*, *11*, 3–39.

Sen, A. (1983). Evaluator relativity and consequential evaluation. *Philosophy & Public Affairs*, *12*, 113–132.

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*, 331–336.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1250–1260.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovick, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379–396). New York: Cambridge University Press.

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

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*, 311–322.

Sox, H. C., Blatt, M. A., Higgins, M. C., & Marton, K. I. (1998). *Medical decision making*. Newton, MA: Butterworth-Heinemann.

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

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? [with commentaries]. *Behavioral and Brain Sciences*, *23*, 645–726.

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases* (pp. 421–440). New York: Cambridge University Press.

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica*, *80*, 143–168.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational Behavior and Human Decision Processes*, *65*, 252–267.

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.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. London: Cambridge University Press.

Wingspread Conference (1998, January). *Wingspread statement on the precautionary principle*. Racine, WI. Retrieved June 23, 2005, from the Science & Environmental Health Network Web site: http://www.sehn.org/wing.html

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175.

Authors' biographies:

Michael L. DeKay is an Associate Professor of Engineering, Public Policy, and Decision Science at Carnegie Mellon University. His research interests involve judgment and decision making in environmental and medical contexts, with particular emphasis on risk perception, risk ranking, precautionary reasoning, and perceived fungibility.

Dalia Patiño-Echeverri is a graduate student in the Department of Engineering and Public Policy at Carnegie Mellon University. Her research interests concern decision making, risk management, and regulation, with specific attention to the electricity industry.

Paul S. Fischbeck is a Professor of Social and Decision Sciences and a Professor of Engineering and Public Policy at Carnegie Mellon University. His research involves descriptive and prescriptive studies into decisions involving risk, with applications that include space-shuttle and oil-tanker design and cultural differences in decisions about purchasing insurance.

Authors' addresses:

Michael L. DeKay, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

Dalia Patiño-Echeverri, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

Paul S. Fischbeck, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

Table 1. Repeated-measures logistic regressions for predicting dominance in Studies 1 and 2

	Study 1 (n = 67)		Study 2 (n = 31)	
Predictor	Weak dominance	Strict dominance	Weak dominance	Strict dominance
Scenario	$\chi^2(7) = 56.15$****	$\chi^2(7) = 63.41$****	$\chi^2(7) = 9.80$	$\chi^2(7) = 14.97$*
Precautionary principle applies			$\chi^2(1) = 6.34$* *OR* = 1.40 (1.08, 1.81)	$\chi^2(1) = 3.17$† *OR* = 1.33 (0.97, 1.81)
Benefit-cost principle applies			$\chi^2(1) = 4.11$* *OR* = 0.82 (0.68, 0.99)	$\chi^2(1) = 4.79$* *OR* = 0.81 (0.67, 0.98)
Risk aversion	$\chi^2(1) = 3.62$† *OR* = 1.30 (0.99, 1.70)	$\chi^2(1) = 3.60$† *OR* = 1.28 (0.99, 1.64)	$\chi^2(1) = 4.24$* *OR* = 1.34 (1.01, 1.77)	$\chi^2(1) = 1.41$ *OR* = 1.18 (0.90, 1.55)

Note: All results are Wald statistics from generalized estimating equations, fit using PROC GENMOD in SAS (Allison, 1999, chapter 8). Degrees of freedom for χ^2 tests appear in parentheses. Odds ratios are for 1-unit increases on 7-point scales. 95% confidence intervals appear in parentheses below the corresponding odds ratios.

†$p < 0.1$. *$p < 0.05$. ****$p < 0.0001$.

Table 2. Repeated-measures logistic regressions for predicting dominance in Studies 3 and 4

	Study 3 (n = 63)		Study 4 (n = 56)	
Predictor	Weak dominance	Strict dominance	Weak dominance	Strict dominance
Context	$\chi^2(1) = 2.99$† *OR* = 1.83 (0.92, 3.63)	$\chi^2(1) = 7.88$** *OR* = 3.41 (1.45, 8.03)	$\chi^2(1) = 7.47$** *OR* = 2.57 (1.32, 5.03)	$\chi^2(1) = 7.74$** *OR* = 2.75 (1.35, 5.61)
Scenario	$\chi^2(7) = 15.20$*	$\chi^2(7) = 7.87$	$\chi^2(5) = 40.05$****	$\chi^2(5) = 52.94$****
Context × Scenario	See text	See text	$\chi^2(5) = 8.73$	$\chi^2(5) = 8.80$
Precautionary principle applies			$\chi^2(1) = 0.93$ *OR* = 1.09 (0.92, 1.28)	$\chi^2(1) = 0.55$ *OR* = 1.08 (0.88, 1.33)
Benefit-cost principle applies			$\chi^2(1) = 0.29$ *OR* = 0.96 (0.82, 1.12)	$\chi^2(1) = 0.03$ *OR* = 1.01 (0.87, 1.18)
Risk aversion	$\chi^2(1) = 4.07$* *OR* = 1.29 (1.01, 1.64)	$\chi^2(1) = 1.56$ *OR* = 1.23 (0.89, 1.69)	$\chi^2(1) = 0.84$ *OR* = 0.89 (0.68, 1.15)	$\chi^2(1) = 0.03$ *OR* = 0.98 (0.75, 1.27)

Note: All results are Wald statistics from generalized estimating equations, fit using PROC GENMOD in SAS (Allison, 1999, chapter 8). Degrees of freedom for χ^2 tests appear in parentheses. Odds ratios for context compare the full-context condition to the reduced-context condition. Other odds ratios are for 1-unit increases on 7-point scales. 95% confidence intervals appear in parentheses below the corresponding odds ratios.

†$p < 0.1$. *$p < 0.05$. **$p < 0.01$. ****$p < 0.0001$.

Table 3. Effects of decision context on continuous measures of outcome desirability (for all cases) and on probability thresholds (when dominance was not implied).

	Study 3			Study 4		
Dependent variable	Full context[a]	Reduced context[a]	Effect in predicted direction[b]	Full context[a]	Reduced context[a]	Effect in predicted direction[b]
U_{TP}	0.34 (0.22, 0.46)	0.07 (–0.06, 0.19)	Yes** ($n = 62$)			
U_{FP}	–0.21 (–0.32, –0.10)	–0.37 (–0.49, –0.26)	Yes* ($n = 62$)			
U_{TN}	0.50 (0.39, 0.61)	0.40 (0.28, 0.51)	No ($n = 62$)			
U_{FN}	–0.84 (–0.92, –0.75)	–0.80 (–0.91, –0.70)	N/A[c]			
$U_{TN} - U_{FP}$	0.71 (0.57, 0.85)	0.77 (–0.93, –0.61)	Yes ($n = 62$)	–0.54 (–1.33, 0.25)	1.10 (0.24, 1.97)	Yes** ($n = 53$)
$U_{TP} - U_{FN}$	1.17 (1.04, 1.31)	0.87 (0.74, 1.00)	Yes*** ($n = 62$)	3.19 (2.55, 3.83)	2.61 (2.09, 3.14)	Yes† ($n = 53$)
$P^* = \dfrac{1}{1+\dfrac{U_{TP}-U_{FN}}{U_{TN}-U_{FP}}}$	0.39 (0.34, 0.44)	0.52 (0.45, 0.58)	Yes** ($n = 63$)	0.06 (0.01, 0.24)	0.13 (0.03, 0.45)	Yes ($n = 39$)

[a]Values were averaged over scenarios within participants. Cell entries are the means of those within-participant means, with 95% confidence intervals in parentheses. Values in the two studies were measured on different scales and are not directly comparable. For P^*, the means and confidence limits were calculated by converting logit values back into probabilities.

[b]All analyses except those for P^* were repeated-measures ANOVAs. Statistical significance is reported for the effect of context, but not for the effects of scenario or the context × scenario interaction. Analyses for P^* were 2-sample t tests comparing mean logit-transformed probabilities in the two context conditions.

[c]No prediction was made for U_{FN}. A floor effect was anticipated because participants were instructed to give the worst outcome (usually a false negative) a rating of –100. The observed difference was not significant.

†$p < 0.1$. *$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

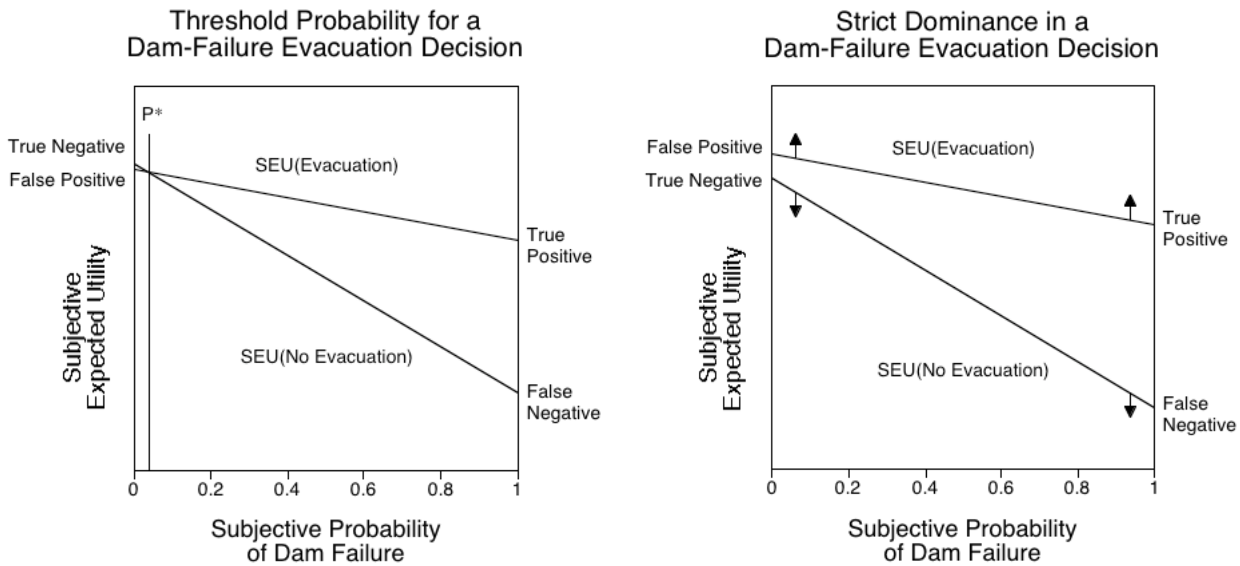

Figure 1. Subjective expected utilities of decision alternatives as functions of the probability of the adverse event, with a decision about issuing a dam-failure evacuation order used as an example. In the left panel, the difference between a true positive and a false negative, $U_{TP} - U_{FN}$, is much larger than the difference between a true negative and a false positive, $U_{TN} - U_{FP}$, so the threshold probability *P** for taking the more protective action (evacuation) is very low. In the right panel, arrows indicate likely biases in the evaluation of outcomes, consistent with an initial preference for evacuation. If such biases are strong enough, the more protective action strictly dominates the less protective action. Weak dominance would be obtained if $U_{TN} = U_{FP}$.

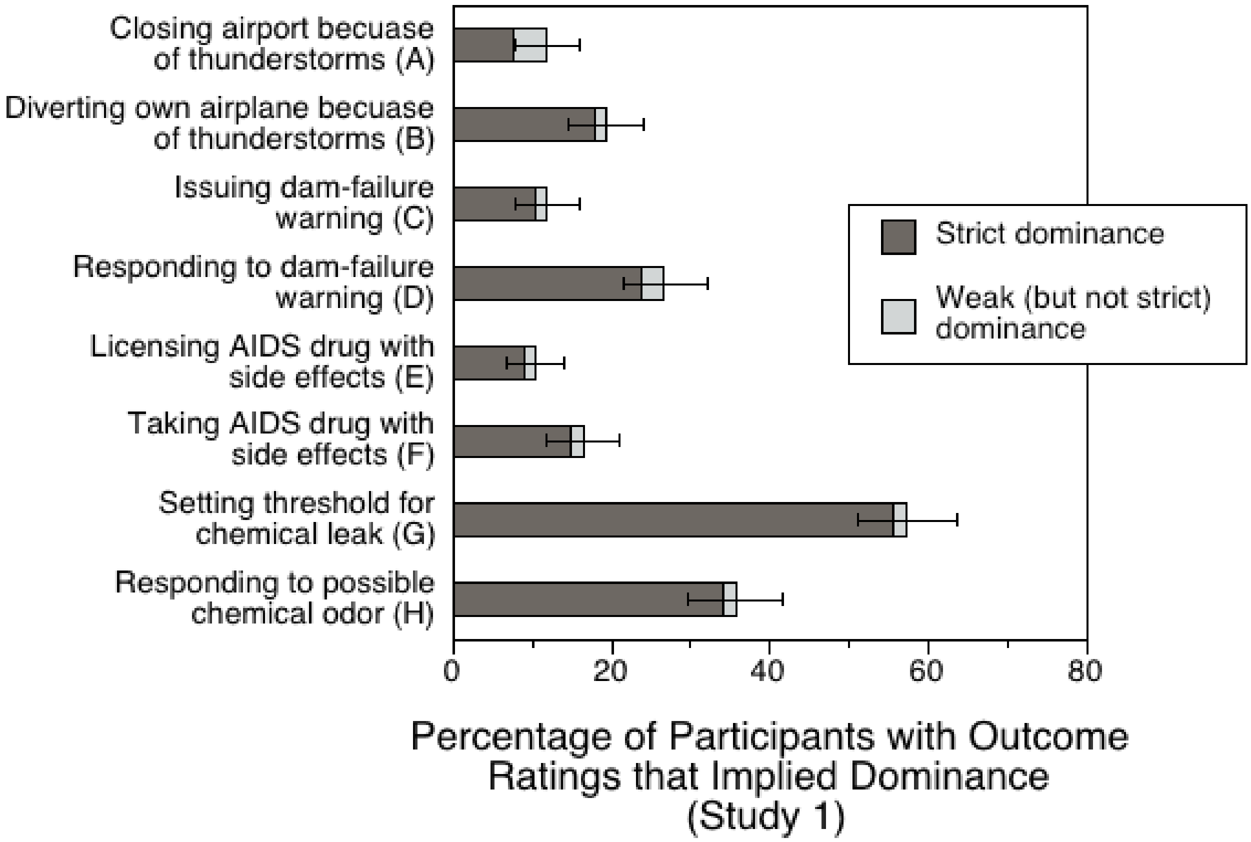

Figure 2. Many Study-1 participants reported outcome ratings that implied that the more protective option dominated the less protective option. In these scenarios, strict dominance means that $U_{TN} < U_{FP}$ and $U_{TP} > U_{FN}$, as in the right-hand panel of Figure 1. Weak dominance means that $U_{TN} \leq U_{FP}$ and $U_{TP} \geq U_{FN}$, with at least one inequality. Error bars depict standard errors for percentages of participants who exhibited weak dominance (including those who exhibited strict dominance).

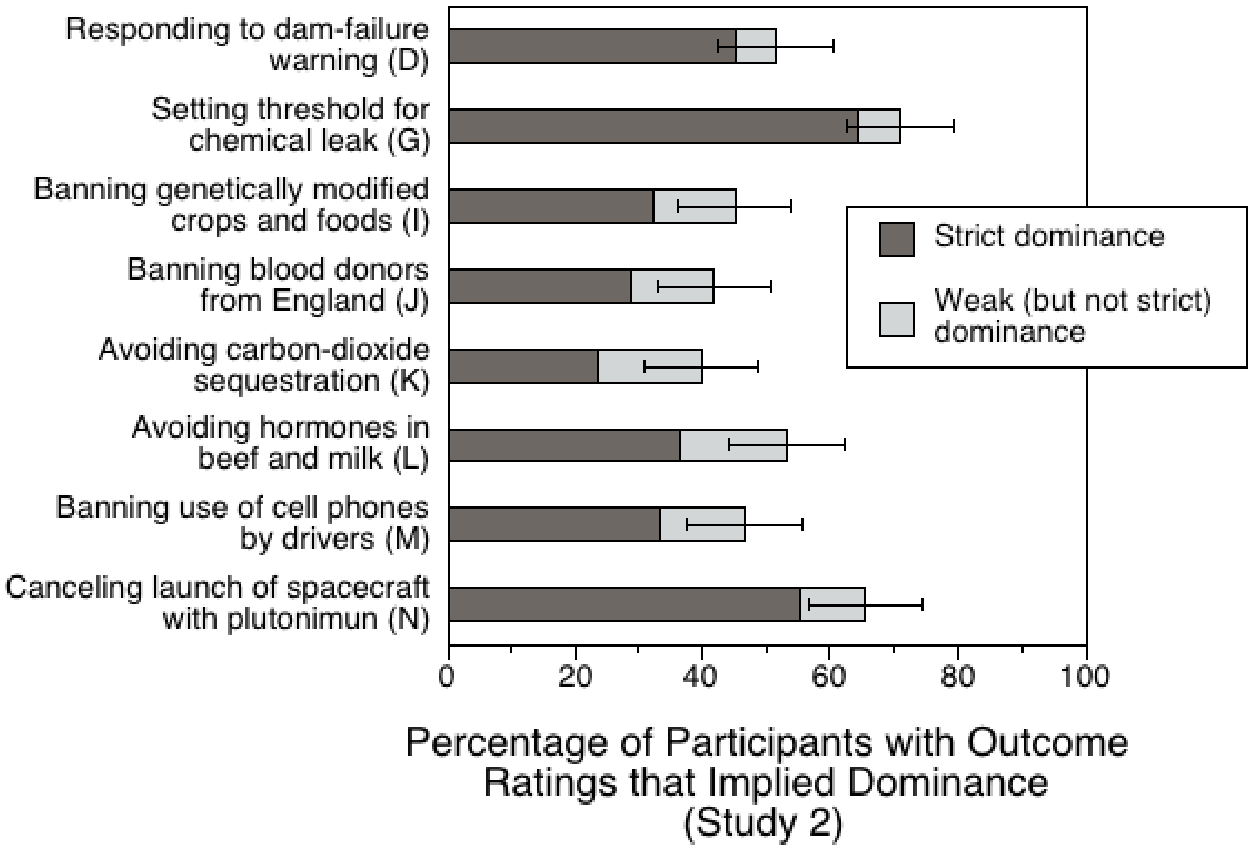

Figure 3. Many Study-2 participants reported outcome ratings that implied that the more protective option dominated the less protective option. In these scenarios, strict dominance means that $U_{TN} < U_{FP}$ and $U_{TP} > U_{FN}$, as in the right-hand panel of Figure 1. Weak dominance means that $U_{TN} \leq U_{FP}$ and $U_{TP} \geq U_{FN}$, with at least one inequality. Error bars depict standard errors for percentages of participants who exhibited weak dominance (including those who exhibited strict dominance).

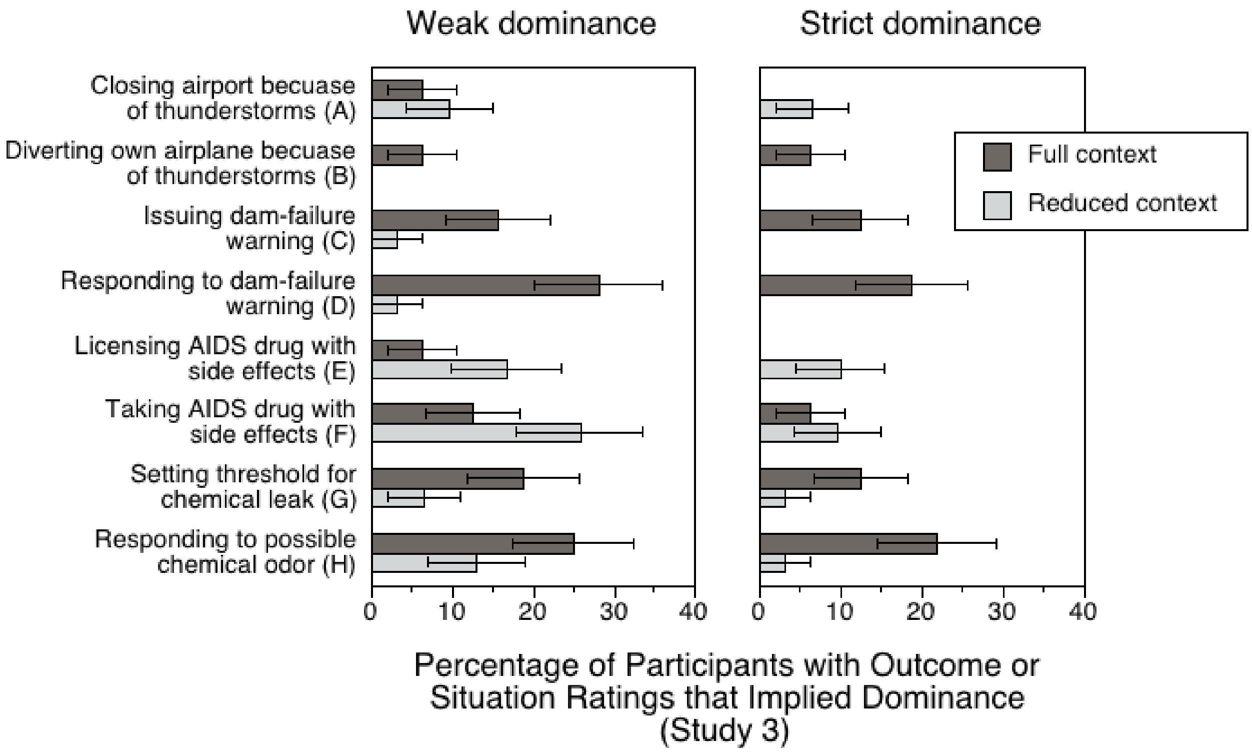

Figure 4. The percentage of Study-3 participants with outcome or situation ratings that implied that the more protective option dominated the less protective option was greater in the full-context condition than in the reduced-context condition for five of the eight scenarios, regardless of whether weak or strict dominance was considered. The effect of context was significant only for strict dominance. Error bars depict standard errors.

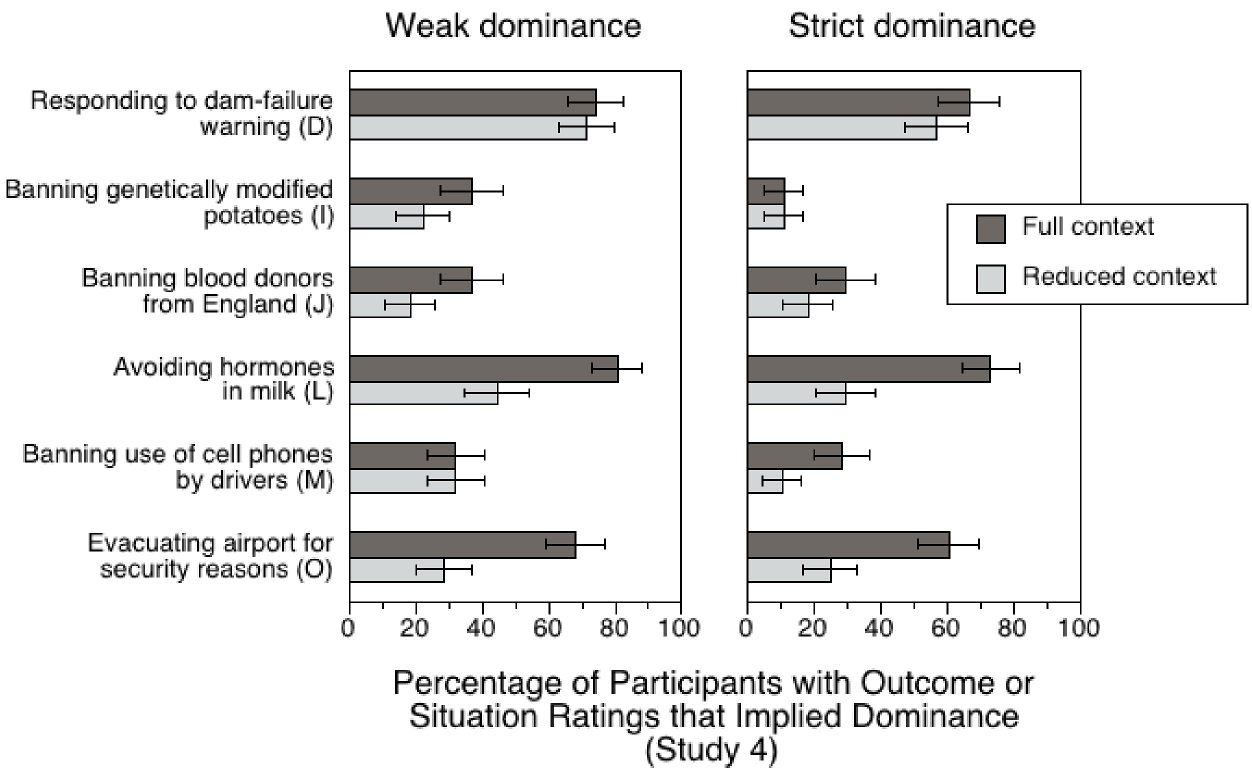

Figure 5. The percentage of Study-4 participants with outcome or situation ratings that implied that the more protective option dominated the less protective option was greater in the full-context condition than in the reduced-context condition for five of the six scenarios, regardless of whether weak or strict dominance was considered. The effect of context was significant in both cases. Error bars depict standard errors.