Running Head: PROPORTION DOMINANCE

Here's the link to the published version: http://dx.doi.org/10.1016/j.obhdp.2005.10.004

Please cite this article as:

Bartels, Daniel M. (2006), "**Proportion Dominance: The Generality and Variability of Favoring Relative Savings over Absolute Savings**," *Organizational Behavior and Human Decision Processes, 100,* 76-95.

Abstract

Four studies probe Ps' sensitivity to absolute and relative savings. In three studies, Ps read scenarios forcing a tradeoff of saving more lives (230 vs. 225) versus saving a larger proportion of a population (225 ÷ 300 = 75% vs. 230 ÷ 920 = 25%). Ps' preferences were driven by both absolute and relative savings. Maximizing relative savings, called "proportion dominance" (PD), at the expense of absolute savings is non-normative, and most participants concur with this argument upon reflection (Studies 2 & 3). PD is related to individual differences, such that people scored as "rational" thinkers exhibited less PD than people scored as "experiential" thinkers (Studies 1 & 3). Finally, a fourth study extends these results, finding proportion dominance in other domains using a different paradigm. These four studies demonstrate both the generality (across domains and paradigms) and the variability (inter- and intra-individual) of proportion dominance.

Introduction

Consider the following caricature of a debate about social policy X that typifies the type of punditry aired on cable news channels:

> Opponent: It seems entirely unreasonable to me that we would be willing to spend 1.2 billion dollars each year on this program that benefits only 4% of the U.S. population!
>
> Proponent: Come on, you cannot object to the average taxpayer spending about 9 dollars a year on this program when it benefits 12 million Americans, can you?

These two are talking about the same numbers; they just construe the information differently. The disagreement between these two is reflected in how they choose to frame the program's cost and benefits. The opponent discusses the program's total (i.e., absolute) cost and frames its benefits in relative terms. The proponent discusses the program's per capita (i.e., relative) cost and frames its benefits in absolute terms. So, who's "right"? The four studies reported here explore the approaches people take in problems like these—problems in which the benefits can be framed in terms of absolute or relative savings[1].

Many studies find that people are sensitive to relative savings (e.g., relative risk reduction). For example, courses of action that completely eradicate some problem are preferred over courses of action that provide incomplete eradication (see Ritov & Kahneman, 1997, Slovic, Fischhoff, & Lichtenstein, 1980). Importantly, these preferences have demonstrable real-world impact. For example, in 1993, many people fell ill due to *Cryptosporidium* contamination in Milwaukee's municipal water supply. At that time, the water treatment facilities removed approximately 99% of *Cryptosporidium* from the water supply (although this estimate was developed later). In 1994, Milwaukee allocated $89 million to improve the infrastructure and thereby increase removal rate to

[1] In complex issues like debates about social policy, it is difficult to define what is normative because these problems reference political attitudes, perceptions of fairness, etc. The problems I presented to participants were intended to be less complex than the social policy example.

99.999%, among other improvements (Milwaukee Water Works Superintendent, personal communication, September 8, 2004). Apparently, the opportunity to completely address the problem was considered very important despite the fact that the absolute improvement was quite small.

A number of studies have examined people's preferences for risk-reducing policy interventions like the one described above. For example, Jenni and Loewenstein (1997) examined a phenomenon known as the "identifiable victim effect", which refers to people's willingness to exert greater effort to save identified victims (e.g., baby Jessica) rather than unidentified victims (e.g., statistical victims). They investigated a number of potential causes of this phenomenon and settled on what they call the "reference group effect"—identifiable victims are perceived as comprising the entire at-risk population for whatever misfortune affects them. Jenni and Loewenstein found that participants' preferences were strongly influenced by the proportion of the relevant risk group that was saved; sensitivity to relative savings drove preferences. Similarly, Baron (1997) independently manipulated both the number of people saved by some intervention (absolute savings) and the size of the at-risk population from which those people were saved, and showed that both the absolute number and the proportion of the total population saved influenced people's preferences. In addition, Fetherstonhaugh, Slovic, Johnson, and Friedrich (1997) and Friedrich et al. (1999) showed that people's value for a statistical life decreases as the size of the reference group increases. They call this "psychophysical numbing". These effects (e.g., the reference group effect, psychophysical numbing, etc.) are largely attributable to people's acute sensitivity to relative savings. I will use the term "proportion dominance" (PD) (Finucane, Peters, & Slovic, 2003) to refer to this sensitivity.

Should one support social policy X because it benefits 12 million Americans, or should one not approve because it only benefits 4% of Americans? The four studies reported here explore the way in which people approach problems like these and find that people are sensitive to relative

savings—that they exhibit proportion dominance. In addition, these studies address the relationship between the putative normative model for problems like these (i.e., what people *should* do) and what people (*actually*) do. Specifically, the normative account suggests that people should be insensitive to relative savings when choosing on that basis undermines absolute savings. These studies show that people exhibit PD even in the contexts in which the normative approach should be most apparent—contexts affording direct comparisons of absolute and relative savings.

Beyond the simple demonstration of PD, these studies explore a number of factors contributing to the generality and variability of sensitivity to relative savings. Proportion dominance is shown to be a robust phenomenon, generalizing across paradigms and domains. In addition, the studies show that PD varies predictably across people (mirroring the findings of studies relating individual differences to the incidence of framing effects, see LeBoeuf & Shafir, 2003; McElroy & Seta, 2003) and that the perceptions underlying PD vary across domains (consistent with a recent line of research investigating content effects on decision making, see Rettinger & Hastie, 2001; 2003; Goldstein & Weber, 1995)

Study 1

Much of the research in judgment and decision making investigates people's performance on some task relative to an external norm. Judgments about whether or not a behavior should be viewed as (non)normative are better informed by careful consideration of the task in which the behavior is exhibited. Study 1 examines the influence of two different experimental paradigms used to elicit PD. Further, Study 1 explores the relationship between individual differences in thinking styles and PD. This study examines whether participants who have shown greater susceptibility to other reasoning and decision making fallacies in previous research will exhibit greater PD, which in

turn, informs us on the relationship between the task and the putative normative approach to the task.

The studies that report proportion dominance in policy preferences tend to use between-participants designs, or when repeated-measures designs are used, options are presented sequentially, rather than simultaneously. This design characteristic is of note because in many cases, people are better at making comparative judgments than absolute judgments—having two comparison objects engenders an evaluative context not present in the sparser, absolute judgment context[2].

A number of studies have shown preference reversals across these two types of contexts (see Hsee, Loewenstein, Blount, & Bazerman, 1999). Consider the example in Figure 1 from Hsee and Leclerc (1998). In one set of conditions (separate evaluation), participants are shown either the product on the left or the product on the right and are asked for a buying price. In another condition (joint evaluation), participants are shown both products and are asked for a buying price for each.

Please insert Figure 1 about here

Buying prices are higher for the 7 oz. serving of ice cream than for the 8 oz. serving in the separate evaluation conditions, but the reverse is true for the joint evaluation condition. Since most people do not have an accessible metric for converting the size of an ice cream serving to its appropriate buying price, people instead tend to use whatever other cues are salient. In separate evaluation, what is salient is the overfilled/underfilled aspect of the choice options. The joint evaluation condition affords participants with a richer evaluative context, where participants are able to pick out the most important attribute for setting a buying price.

[2] See Moore (1999) for a comparison of the evaluative contexts engendered in separate evaluation versus sequential evaluation.

Applying this rationale to the study of proportion dominance, it seems that the use of separate or sequential evaluation designs might lead to different inferences about what drives people's policy preferences than the use of a joint evaluation design might suggest. Whether people choose to maximize relative savings (and people's related normative perceptions) might vary with task used to elicit and measure proportion dominance. Specifically, one might expect PD to disappear in joint evaluation.

Separate Evaluation versus Joint Evaluation and Proportion Dominance

Some results from Fetherstonhaugh et al. (1997) suggest that PD varies across separate and joint evaluation. In one study, they asked participants for the minimum number of lives a new medical treatment would have to save to merit $10 million in funding. The only information participants were given was the number of annual deaths attributable to a given disease—the size of the reference group. So, for this task, participants were given only the information in the left and center columns of Table 1 and were asked for three such threshold sequentially. Most participants (65%) required more lives to be saved for diseases that were expected to kill more people. That is, thresholds for Disease Z were larger than those for Disease Y, and thresholds for Disease Y were larger than thresholds for Disease X.

Please insert Table 1 about here

Dividing $10 million by a participant's threshold response yields an index of the value of a single statistical life for that participant. Given that these judgments were about meriting a fixed amount of funding, these results might imply that the value of one statistical life does not equal the value of one statistical life, but rather that the value of a statistical life is a decreasing function of the size of the reference group to which the life is compared.

After making these judgments, participants were then asked to rank-order the treatment programs by the importance of funding them. Participants were again given the number of people expected to die from each disease, but for this ranking task, participants were also given the number of lives each treatment was expected to save. The numbers were set up such that absolute and relative savings were inversely related (see Table 1). When faced with this task, the majority of participants (60%) rank-ordered the importance of funding these treatments in terms of absolute savings. That is, they assigned higher rankings to treatments that saved more people, ranking the treatment for Disease Z as being more important than the treatment for Disease Y and the treatment for Disease Y as more important than the treatment for Disease X, even though Disease X's treatment saves the largest proportion of its at-risk group. Only a small minority (16%) ranked the options in the inverse order—in terms of maximizing the proportion of lives saved for a given disease.

When options were presented separately, participants' exhibited a high level of PD-type responding, but when the options were presented jointly in the ranking task, proportion dominance was attenuated. Importantly, however, proportion dominance did not completely disappear. Some people continued to exhibit proportion dominance in joint evaluation while the majority did not.

The difference across tasks—that 60% exhibit PD in separate evaluation, while only 16% exhibit PD in joint evaluation—is suggestive of an important way in which task characteristics interact with the evaluation process. One contribution and major focus of the studies reported here is the demonstration of proportion dominance even in a joint evaluation task and in a preference matching task. What makes these demonstrations of some theoretical interest is that these two tasks present participants with all of the cues relevant to the evaluation process. Under these conditions, we might expect PD to be nearly eliminated.

The Normative Account

Some have suggested (Baron, 1997) that choosing based on relative savings is non-normative. The argument is that the value of a single life should be the same regardless of the size of the reference group. That is, the value of saving any given number of resources should be independent of the size or proportion of the population left unsaved.

To demonstrate why this argument is compelling, consider a finding from Jenni & Loewenstein (1997). Participants' willingness to support an intervention preventing 2 traffic fatalities in Pittsburgh annually was greater when they were told that these deaths account for 2 of the 4 annual auto-related fatalities at that intersection than when told that these deaths account for 2 of the 112 annual auto-related fatalities in Southwestern Pennsylvania (these options were not presented in joint evaluation).

Of course, 2 deaths are 2 deaths and 4 deaths at an intersection in Pittsburgh are a subset of 112 fatalities in the greater geographical region. Nevertheless, when these 2 deaths were compared against smaller reference group, participants were more willing to support this intervention. So, demonstration of PD in separate evaluation could be thought of as a violation of descriptive invariance—a framing effect. While framing effects are common in the judgment and decision making literature, they are sometimes reduced or eliminated in repeated measures designs (Frisch, 1993; LeBeouf & Shafir, 2003; Levin, Gaeth, Schreiber, & Lauriola, 2002; Stanovich & West, 1998).

While some have argued that "much of life resembles a between-subjects experiment" (Kahneman, 2000, p. 682)—that many of life's experiences take place in separate evaluation contexts (e.g., Kahneman, Knetsch, & Thaler, 1986; Shafir, 2002)—separate evaluation designs may lack some of the crucial reference points and evaluative context that our cognitive systems rely on. So, conclusions drawn about how people's judgments and preferences are shaped using *only*

separate evaluation designs may be underinformed, because in some separate evaluation contexts, people are not presented with cues they would otherwise use to reach an optimal solution. While proportion dominance in separate evaluation may be suboptimal, it may prove difficult to define the appropriate prescriptive approach for problems presented in these sparse contexts. This point is best demonstrated by a paper entitled "How to show that 9 > 221: Collect judgments in a between-subjects design" (Birnbaum, 1999). (See Hsee & Zhang, 2004 for contexts in which joint evaluation leads to suboptimal choices.)

It seems more difficult to argue that proportion dominance in joint evaluation is normative. When given two options to evaluate, one option that saves a greater proportion of its at-risk group at the expense of saving fewer people and another option that saves more people while saving a smaller proportion of its at-risk group, selecting the former causes needless deaths. It can be argued that such a preference is akin to selecting the overfilled cup with the smaller serving of ice cream in the Hsee and Leclerc (1998) example. For this reason, one might not expect to find proportion dominance in joint evaluation.

Dual-Process and Individual Differences in Decision Making

Study 1 also examined individual differences in proportion dominance. Individual differences were examined partly to determine whether a comparison between participants' responses and the normative account constitutes a reasonable contrast. While "[normative models] are justified by analysis, not consensus" (Baron, 2000, p. 668), Stanovich and West (2000) suggest that whether or not a normative model is applicable to a given task can be investigated by examining whether- and which participants adopt the normative approach. Relating participants' responses to individual differences in cognitive capacity or thinking styles can provide evidence

pertaining to the fairness of the task—whether one is justified in calling some responses normative and others non-normative.

Standardized individual differences indices have been related to a variety of reasoning and decision making fallacies (Chatterjee et al., 2000; Lauriola & Levin, 2001; LeBoeuf & Shafir, 2003; Levin et al., 2002; McElroy & Seta, 2003; Pacini & Epstein, 1999; Simon et al., 2004; Stanovich & West, 1998). Some theories of the processes underlying individual differences suggest a dual-process view of thinking. Dual-process models abound in psychology (e.g., Chaiken & Trope, 1999; Sloman, 1996; Smith & DeCoster, 2000; Strack & Deutch, in press), and differences exist between these models. However, one commonality shared by many is the idea that one system operates relatively automatically while the other operates more deliberatively an analytically. Some of the focus in both theory and research has been on the putative biases generated by the automatic-intuitive system and the putative error-correcting functions the analytical-deliberative system performs (see Kahneman & Frederick, 2002; for an alternative view, see Smith & Decoster, 2000).

Recall that individuals differ in their sensitivity to relative savings; PD was attenuated but not eliminated in Fetherstonhaugh et al. (1997)—some people continued to exhibit proportion dominance in joint evaluation. If it is reasonable to call proportion dominance in joint evaluation non-normative, then we might expect participants' preference for relative savings *at the expense of absolute savings* to relate to individual differences in thinking styles in the same way that susceptibility to framing effects and other fallacies have been related to individual differences in thinking styles.

In the studies reported in this paper, differences in thinking styles were measured with an adapted version of Epstein's (1996; Pacini & Epstein, 1999) Rational versus Experiential Inventory (REI) because the REI reflects a dual-process approach. This measure is actually two scales: the rational scale, which is adapted from Need for Cognition scale (Cacioppo et al., 1984) measures

enjoyment of and participation in deliberative, rational-analytical thinking; and the faith-in-intuition scale measures enjoyment of and participation in intuitive, nondeliberative thinking. Given the deliberative system's putative function as an error-correction tool (among other functions), the difference between scores on the two scales of the REI was used to predict proportion dominance. The resultant score indexes participants' relative reliance on deliberative-analytical vs. intuitive-experiential thought. Participants who are willing to engage their cognitive capacities in this task should be indexed by relatively high scores and vice versa. Thus, participants who are more rational-analytic than intuitive will show a reduced susceptibility to proportion dominance in joint evaluation when compared to more intuitive-nondeliberative thinkers.

Following the logic of Stanovich and West (2000), this analysis also provides a test of the relationship between how people should do the task and how people actually do the task. If the predicted relationship obtains between PD and thinking styles—where the normative approach is more likely to be adopted by participants who are relatively more willing to engage their computational resources on the task—this pattern of data can be interpreted as indirect evidence that the normative model is applicable to this task. In other words, if participants who are more likely to rely on intuition than careful analysis exhibit more proportion dominance in joint evaluation, this pattern of data may suggest that this task is as fair as other tasks (e.g., framing effects experiments) in eliciting responses that can be reasonably contrasted with the normative model.

Method

Participants

Forty-four Northwestern University undergraduates (27 women; 17 men) participated for partial course credit. Each participant filled out a questionnaire at their own pace. They were tested individually but in a small group setting (typically, 1 to 4 Ps per session). Usually, other

participants were filling out the questionnaire at their own pace in the same room. No gender differences were observed in any of the studies collecting gender information.

Materials & Design

Participants were asked to read and respond to ten different decision items in a packet. Five of these items were separate evaluation items and five were joint evaluation items. For each of the ten stimulus items, three versions were constructed: a separate evaluation-small reference group version, a separate evaluation-large reference group version, and a joint evaluation version. An example of each version is provided below for the Anthrax item. Eight versions of a basic questionnaire counterbalanced the order and version for each of the ten items. All of the joint evaluation versions of the items used in Studies 1 through 3 are included in Appendix A.

Each participant read and responded to five separate evaluation items (either 3 large- and 2 small-reference group or 2 large- and 3 small-reference group). In these items, a single intervention was described and participants were asked for a willingness-to-support rating for this intervention. For a given item, the number of lives saved was described against either a relatively small or a relatively large reference group between-participants, yielding one item where absolute savings is comparatively large (small reference group item) and one where absolute savings is comparatively small (large reference group item). For example, this is the separate evaluation-small reference group version of the anthrax item:

> Anthrax powder has been weaponized and released into the air above a mid-sized city. In this city, a number of people are expected to die as a result of anthrax inhalation. There exists a powerful antibiotic that will successfully treat some victims, but there is a limited amount of this treatment. If you choose to delegate the treatment to the city, 225 of the 300 at risk of death will be saved.

The separate evaluation-large reference group item read as follows:

> Anthrax powder has been weaponized and released into the air above a mid-sized city. In this city, a number of people are expected to die as a result of anthrax inhalation. There exists a powerful antibiotic that will successfully treat some victims, but there is a limited amount of this treatment. If you choose to delegate the treatment to the city, 230 of the 920 at risk of death will be saved.

After reading each of these items, participants were asked to rate their willingness to support this intervention by circling one of seven partitioning marks on a scale bounded by "Would Not Support At All" and "Would Give Strongest Possible Support" (scored as 0-1).

Each participant also read and responded to five joint evaluation items, where two interventions are described as treating identical problems in two different populations. The joint evaluation items were constructed by including both the small and large reference group separate evaluation items in one item. (The assignment of the small/large reference group separate evaluation item to the label "Program A" or "Program B" was randomized across different versions of the questionnaire.) For these items, participants responded on a scale measuring strength of preference for one intervention over the other. For example, the joint evaluation version of the anthrax item read as follows:

> Anthrax powder has been weaponized and released into the air above two mid-sized cities. In each city, a number of people are expected to die as a result of anthrax inhalation. There exists a powerful antibiotic that will successfully treat some victims, but there is a limited amount of this treatment. Program A would delegate the treatment to City A, and 225 of the 300 at risk of death will be saved. Program B would delegate the treatment to City B, and 230 of the 920 people at risk of death will be saved. These programs are mutually exclusive and the only two options available.

After reading each of these items, participants were asked to rate their relative strength of preference for one intervention over the other by circling one of seven partitioning marks along a scale bounded by "Strong Preference for Program A" and "Strong Preference for Program B". Responses on this scale were coded with respect to which of the two interventions described

maximized absolute savings. Responses at the end of the scale corresponding to a strong preference for the intervention that maximized absolute savings were coded as zeros; responses at the end of the scale corresponding to a strong preference for the intervention maximizing relative savings were coded as ones. This yielded a measure of proportion dominance in joint evaluation, with responses coded by their difference (from zero to one) from the absolute savings end of the scale.

Participants were then asked to write a short explanation of their ratings for each item. An initial, cursory analysis of these justifications motivated one design change implemented in Studies 2 and 3, but in general, these justifications proved to be "noisy" data. So, outside of a short discussion in the introduction and discussion sections of Study 2, I will not discuss the justifications collected in Study 1.

Finally, Ps responded to an adapted, 20-item version of the REI (Epstein et al, 1996; Pacini & Epstein, 1999). For each item, Ps rated their level of agreement with statements like "I prefer to do something that challenges my thinking abilities rather than something that requires little thought" (where greater agreement suggests greater reliance on analytic-deliberative thought) and "Using my gut feelings usually works well for me in figuring out problems in my life" (where greater agreement suggests greater reliance on intuition).

Results

Separate Evaluation

While participants' willingness-to-support ratings were expected to be high for each of the separate evaluation items, the size of the reference group to which an intervention was compared was expected to influence ratings. One might expect a given intervention that entails low relative savings to be rated lower due to concerns over the intervention's efficiency, for example.

As expected, participants showed marked sensitivity to relative savings for the separate evaluation items. Participants' average willingness-to-support ratings (converted to a 0 to 1 scale) were higher for interventions where the number of lives saved was described against a small reference group (M = .88, SD = .12) than for interventions described against a large reference group (M = .76, SD = .15; *paired* $t(43) = 4.85$, $p < .001$, $\eta^2 = .353$)[3]. Thirty-four of the 44 Ps in this experiment showed the effect. This result replicates similar findings by Baron (1997), Fetherstonhaugh et al. (1997), Jenni & Loewenstein (1997), and Friedrich et al. (1999). For these items offering rather sparse evaluative context, the size of the reference group acted as a salient reference point for judgment. Since proportion dominance in separate evaluation appears to be replicable, I will not explore this effect any further.

Joint Evaluation

Interestingly, proportion dominance was not eliminated under conditions that afforded comparisons of both absolute and relative savings. In the way PD is measured here, each rating *not* at the end of the scale that maximizes absolute savings suggests at least some sensitivity to relative savings. With respect to a preference for sensitivity to relative savings *over* absolute savings, 48% of participant's responses were greater than 0.5 (i.e., on the half of the scale demonstrating a preference for the intervention that maximized relative savings)[4]. The majority of analyses in this

[3] I report eta-squared (η^2) for within-sample- and non-independent sample comparisons. Eta-squared approximates the proportion of variance explained by the factor of interest. Cohen (1977) interprets η^2's of .02, .13, and .26 as small, medium, and large effects, respectively.

[4] In Studies 1 through 3, the distribution of PD in joint evaluation appears to be centered around 0.5. There are at least two potential ways to produce a distribution with this central tendency. The first possibility is that averaging across items for which Ps exhibit more PD and items for which Ps exhibit less PD produces this distribution. Alternatively, as one reviewer suggested, if Ps used one half of the scale approximately one half of the time, this might suggest a general reluctance to choose.

To address this question, I examined the relationship between variability in Ps' responses and Ps' reluctance to choose. For each participant in Studies 1 through 3, I first calculated the standard deviation of their PD responses (in joint evaluation). I then used these SD's to predict the absolute difference between Ps' mean level of PD exhibited and 0.5 (i.e., no preference) for each participant. For expository purposes, I will refer to this index as the "preference index"—high scores reflect the presence of preference; low scores reflect the absence of preference. This preference

paper do *not* focus on the demonstration that people show at least some sensitivity to relative savings. Instead, these analyses focus on magnitude of PD as it relates to other factors.

Proportion dominance was related to individual differences in thinking styles. Scores on the REI[5] reliably predicted proportion dominance ($r(43) = -.372$, $p = .013$, $\eta^2 = .118$); the greater the relative reliance on deliberative thought, the less proportion dominance Ps exhibited in joint evaluation[6].

Proportion dominance in separate evaluation was not related to individual differences. Across-Ps, REI is not related to the magnitude of each P's difference in willingness to support ratings across small/large reference groups ($r(43) = .050$, $p = .746$). Further, each participant's difference between small/large is not related to PD in joint evaluation, either ($r(43) = .015$, $p = .926$), suggesting that PD in separate and joint evaluation are dissociable effects.

Discussion

That proportion dominance was exhibited in a separate evaluation design is not surprising, but that Ps' preferences showed marked sensitivity to relative savings in joint evaluation is somewhat unexpected—recall that Fetherstonhaugh et al. (1997) found that only 16% of Ps showed

---

index is blind with respect to magnitude of PD, but it differentiates Ps with strong preferences for absolute (PD = 0) or relative savings (PD = 1.0) at the maximum ($|0 – 0.5| = |1.0 – 0.5| = .05$) from Ps who exhibit no preference (PD = 0.5) at the minimum ($|0.5 – 0.5| = 0$ at the minimum).

If the central tendency of PD is due to Ps' general reluctance to choose, SD's should be positively realated to this preference index. That is, the distribution might be centered around 0.5 because many Ps tend not to stray too far from the midpoint of the scale. If the Ps who demonstrate a consistent lack of preference across items predominate, then should expect a positive correlation between SD's and preference indices.

The results of these analyses suggest the reverse: Ps were not exhibiting a general reluctance to choose. The distributions' centers, instead, are produced by averaging across items for which Ps show varied preferences—preferences going in one direction or the other. The correlations between SD's and preference indices in each of the first three studies were strongly negative; Study 1: $r(43) = -.444$, $p = .003$, $\eta^2 = .178$; Study 2: $r(69) = -.465$, $p < .001$, $\eta^2 = .205$; Study 3: $r(53) = -.492$, $p < .001$, $\eta^2 = .228$. Ps whose responses varied greatly tended to exhibit average levels of PD that were closer to 0.5 than Ps whose responses did not vary greatly.

[5] Consistent with Epstein et al. (1996) and Pacini & Epstein (1999), the Rational versus Experiential Inventory consisted of two independent scales ($r(43) = .107$, *n.s.*), with each scale reaching a reasonable level of reliability for only 10 items (Chronbach's $\alpha = .89$ and .80 for the rational and experiential indices, respectively).

[6] Scores on each subscale of the REI were not as predictive as the difference score. While the experiential score predicted PD ($r(42) = .328$, $p = .030$, $\eta^2 = .086$), it did not explain as much variance as the difference score. The rational score did not predict PD ($r(42) = -.146$, n.s.).

PD in the joint evaluation rank-ordering task. In Study 1, Ps favored relative savings *over* absolute savings for 48% of all items. This demonstration is of theoretical interest because some would argue that people should not show proportion dominance when all the information (affording both the relative and the absolute comparisons) is displayed as it is in joint evaluation. This argument seems intuitive for this set of stimuli, where such choices that maximize relative savings lead to needless deaths.

Certain people appear to be more susceptible to judgment and reasoning fallacies than others, but predicting this susceptibility with individual difference indices is not always possible. In Study 1, proportion dominance was related to individual differences in joint evaluation but not in separate evaluation. These findings are not inconsistent with those found in studies investigating individual differences in susceptibility to framing effects. One general finding is that Ps who prefer analytical thought to intuition show reduced framing effects in repeated-measures designs (Stanovich & West, 1998; Levin et al., 2002; LeBeouf & Shafir, 2003). However, many of these studies find significant gain/loss risky choice framing effects in between-Ps designs, even among people who enjoy and engage in analytical-deliberative thought (LeBeouf & Shafir, 2003, McElroy & Seta, 2003; Simon, Fagley, & Halleran, 2004). In many cases, it appears that Ps who prefer analytical thought to intuition are affected by the original framing of the problem, but when they see the second wording of the problem, they remain consistent with their earlier risk preference. Here, consistency reflects coherent preferences (satisfying one criterion for rationality); inconsistency is viewed as non-normative. The advantage enjoyed by analytical Ps is in successfully avoiding non-normative behavior.

Between-Ps separate evaluation designs offer little evaluative context. Perhaps the framing of the problem or the size of the reference group constitute a comparatively larger proportion of cues present in the problem, making weighting the cues for relevance comparatively more difficult

than in a repeated-measures or joint evaluation design. It is possible that these between-Ps separate evaluation manipulations are so powerful that they leave relatively little "wiggle room" for individual differences in thinking styles (for further discussion on this point, see Appendix B).

Rips (personal communication) and a reviewer suggested that the relationship between PD and thinking styles could be driven by something else; they suggested that the REI could be proxying for intelligence or for a type of intelligence. The data reported here cannot rule out the possibility that differences in intelligence are driving this effect. I agree with this general sentiment; intelligence and thinking styles are probably not orthogonal—they are probably inextricably linked—but they are at least stochastically independent. Indeed, Stanovich and West (1997; 1998; Sa, West, & Stanovich, 1999) consistently find correlations between cognitive ability and thinking styles, but they also find that performance on reasoning tasks is predicted by differences in thinking styles even after controlling for variability in cognitive ability. So, thinking styles can sometimes provide a unique contribution in the explanation of an effect. However, intelligence and thinking styles relate to different questions and to different levels of analysis. The REI does not index individual differences in computational power (e.g., working memory), but rather individual differences in the willingness to engage one's computational power.

The question in which I was interested was whether participants who choose to engage their computational power (i.e., those who engage in thoughtful analysis) would be more likely to adopt the normative approach than participants who tend to rely on their intuition[7]. Holding intelligence constant, one could predict the opposite relationship between thinking styles and proportion dominance. For example, Wilson and colleagues (Wilson, Lisle, & Kraft, 1990; Wilson et al.,

[7] Holding thinking styles constant, one should expect a strong positive correlation between intelligence (i.e., computational power) and normative responding in tasks where the normative strategy is demonstrably more complex (i.e., taxing on working memory) than other plausible strategies. In the proportion dominance tasks reported here, one could argue assessing which of two numbers (e.g., 225 and 230) is larger is a less demanding task than approximating proportions and comparing them.

1993; Wilson & Schooler, 1991) have shown that introspection—excessive deliberation on one's own choice process—can reduce the quality of one's choices. In a different line of work, Kahneman and Frederick (2002) suggest that many of the biased responses in the judgment literature may be attributable to "attribute substitution", whereby participants answer an easier question than the one they were asked or heavily weight more accessible, less diagnostic attributes in their judgments. For instance, participants may be answering the question of how similar the predicate "bank teller who is active in the feminist movement" is to Linda's description rather than invoking set theory to assess how likely this conjunctive predicate applies to Linda. Kahneman and Frederick propose a two-system framework in which the intuitive system sometimes generates errors like these and where one of the functions of the more deliberative, analytical system is to correct these errors. In Study 1 and in all the studies reported here, the normative response may entail answering the easier of two questions (i.e., which number is larger?) rather than the more difficult of two questions (i.e., which proportion is larger?). Participants can plausibly "overthink" in proportion dominance tasks. On this view, the observed relationship between thinking styles and PD might be counter-intuitive.

I expected the normative (and potentially more efficient) strategy would be more likely to be recognized and adopted by the "rational" participants than by the "experiential" participants, because a thoughtful analysis of this task should entail distinguishing important from unimportant information and choosing based on the relative merit of absolute savings versus relative savings. The predicted relationship obtained, and is interpreted as at least weak diagnostic evidence that the normative account applies to this task and, subsequently, that this task fairly discriminated normative from non-normative responses.

When arguing that some pattern of responses is non-normative, one bears a hefty burden of proof and has to address a number of oft-levied criticisms (see Shafir & LeBoeuf, 2002; Stanovich

& West, 2000). I reserve a short analysis of some of these issues for the general discussion, but Study 2 addresses some of these criticisms. In particular, one criticism is that researchers and participants sometimes have different interpretations of the problems' structure and appropriate solutions. Study 1 leaves open the possibility that participants engender a different set of norms for these problems than the ones suggested by researchers. Study 2 addresses the question of whether or not PD in joint evaluation is non-normative by more directly addressing participants' construal of the task.

Study 2

The individual differences findings in Study 1 provide one kind of test suggesting that PD in joint evaluation is non-normative. Study 2 seeks out an additional and qualitatively different type of evidence addressing this issue, expanding Study 1's findings in two ways. First, Study 2 directly assesses participants' understanding of the decision problems, including measures probing what participants perceive to be a normative approach for the task. Second, Study 2 examines the potential differences in these perceptions across domains.

The general motivation for Study 2 was that judging whether or not PD is non-normative could be better informed by asking participants about their normative perceptions for these decision problems. It could be that tasks like these are more difficult than their designers intended them to be. Naïve participants in this- and other decision making experiments face the dual tasks of screening the information to decide what is relevant and weighting the information that describes alternatives. In some cases, some of the information that the task designers (who are focused on the "deep structure" of their scenarios) treat as irrelevant may not be treated as irrelevant by participants. Surface features have long been known to influence participants' preferences, even for stimuli as semantically-sparse as monetary gambles (Payne & Braunstein, 1971; Slovic & Lichtenstein, 1968).

A number of studies have shown that embedding a structurally-identical decision problem in different contexts can produce different responses. For instance, Goldstein and Weber (1995) showed that people use different strategies for selecting consumer goods versus a long-term mate. Rettinger and Hastie (2001; 2003) embedded structurally-identical decision problems within a legal context, a financial investment context, a gambling context, and an academic context. Participants faced with different contexts chose differently and used different strategies in arriving at those decisions. Similarly, Fagley and Miller (1997) presented participants with Asian disease-type problems embedded in a human lives domain versus a financial domain and found different risk preferences in these two domains.

Proportion dominance may vary across domains. Recall that participants' post-decision justifications were collected in Study 1. Though not amenable to a formal analysis, they did point to a potential content- or domain-specific effect in proportion dominance. Some participants wrote that since every human life is highly valued, one should strive to maximize the number saved, while some participants' justifications for the natural resource-related items stressed the importance of saving the largest proportion of a specific population. Content/domain effects were not anticipated, and unfortunately, the small number of responses per item did not facilitate formal tests in Study 1. To probe for domain effects more systematically in Study 2, three of the six items concerned human life, three concerned natural resources, and each participant provided a response for each of these six items.

The normative account suggests that participants should maximize the absolute savings, regardless of whether the resources under consideration are human lives or natural resources. Study 2 includes measures adapted from Tordesillas and Chaiken (1999) that more directly examine participants' explicit opinions about which dimension (i.e., absolute or relative) is more important and their opinions about which dimension(s) drove their evaluation of choice options.

Tordesillas and Chaiken (1999) investigated the finding that introspection sometimes leads to poorer choices (Wilson, Lisle, & Kraft, 1990; Wilson et al., 1993; Wilson & Schooler, 1991). They asked participants to assess whether or not they would enroll in college courses that were described on a number of attributes. Some participants were asked to introspect on their choice processes as they performed the task; participants in the control condition were given no such instruction. After participants had completed the assessment portion of the task, they were asked (among other things) to rate the degree to which they had been influenced by various attributes and to assess how much weight they thought should be given to each attribute in choosing courses. Tordesillas and Chaiken found that participants in the introspection condition were influenced by attributes that they themselves rated as relatively unimportant, that participants in the control condition better discriminated important from unimportant attributes when choosing, and importantly, that consensus was achieved across groups with respect to which attributes were rated as important. In other words, the groups did not differ in their beliefs about the optimal use of information in their decisions, what differed was the ability to discriminate between important and unimportant information when assessing courses in the first part of the task. Participants in the introspection condition recognized a discrepancy between the attributes that should have influenced their judgments versus the attributes that actually influenced their judgments.

In a similar vein, I used measures adapted from Tordesillas and Chaiken to assess whether or not participants felt they were behaving normatively in the task. These measures asked participants both how much they felt they had been influenced by absolute and relative savings and how much they think they should have been influenced by these attributes. I included these measures to address two questions: 1) What are participants normative perceptions of the task? 2) What do participants' retrospective evaluations of their own preferences tell us about their behavior

in this task—specifically, do participants who exhibit high(low) levels of proportion dominance feel that they weighted the attributes as they should have in the task?

The answer to the first question provides a more direct test of the normative model's applicability to this task. Participants' normative perceptions were measured by having them rate how much weight each attribute *should* be given in choosing between interventions. These measures were included at the end of the task because participants have more knowledge of the task after having completed the task. More specifically, it is likely that participants will know more about the deep structure of the task after having completed it. Research in similarity, analogy, and problem-solving has shown that exposure to- or comparison of exemplars sharing deep structure (e.g., problems with the same solution) can facilitate recognition of that deep structure, even when the exemplars themselves share relatively little surface similarity (e.g., Gentner & Medina, 1998; Gick & Holyoak, 1980; Medin & Ortony, 1989). This research suggests that after having encountered a number of scenarios sharing the same analogical structure (i.e., a tradeoff between absolute and relative savings), participants are more likely to have generated normative rules that apply across the scenarios (that vary with respect to the specific dilemma involved, the resources under consideration, etc.). In light of the more enriched knowledge state that hindsight provides, if participants do not feel the normative model is applicable to this task, then this should be reflected in the relative importance they assign to absolute (versus relative) savings. In addition, these measures address the second question, which helps us to interpret (non)normative responses through participants' perception of their own behavior as it relates to the normative approach.

Method

Participants

Seventy Northwestern University undergraduates (45 women; 25 men) participated for partial course credit. Participants were run under the same conditions as the previous study.

Materials & Design

The materials and design of Study 2 were the same as the joint evaluation condition in Study 1, with all changes noted subsequently. First, six of the ten joint evaluation scenarios used in Study 1 were used as the materials for Study 2. As noted previously, three of these scenarios entailed losses of human lives, and three entailed losses of natural resources. Second, importance and influence ratings, adapted from Tordesillas and Chaiken (1999) were added in Study 2. Third, Ps in Study 2 did not respond to REI items. Finally, Study 2 used eight partitioning marks on the scale running from "Strong Preference for Program A" to "Strong Preference for Program B".

After participants had read all six scenarios and indicated a strength of preference for each, participants were given a summary of the numerical information conveyed in each of the six scenarios, followed by two rating scales. One scale measured the degree to which participants felt differences along the absolute savings dimension were important for guiding choice; the other scale measured the degree to which participants felt differences along each dimension actually influenced their choices. Henceforth, I refer to the former as the importance measure and the latter as the influence measure. An example of these measures and the information display appears below:

Anthrax & Antibiotics Scenario:

- Program A saves 225 of 300 people
- Program B saves 230 of 920 people

1. Program A saves 75% of those at risk; Program B saves 25% of those at risk

2. Program B saves 5 more people than Program A

Given this information, please indicate how much weight you think SHOULD be given to attributes 1 and 2 above in choosing which program to implement:

Attribute 1:

+_____+_____+_____+_____+_____+_____+_____+
Not Important Very Important

Attribute 2:
+_____+_____+_____+_____+_____+_____+_____+
Not Important Very Important

Please take a look back at the decision you made in the Anthrax & Antibiotics scenario. Please indicate the degree to which each attribute above influenced your decision.

Attribute 1:
+_____+_____+_____+_____+_____+_____+_____+
Not Influential Very Influential

Attribute 2:
+_____+_____+_____+_____+_____+_____+_____+
Not Influential Very Influential

Results

Most participants agree with the normative approach upon reflection. Participants rated the differences in absolute savings as being more important (M = .74, SD = .19) than differences in relative savings (M = .52, SD = .25; *paired* $t(69) = 4.39$, $p < .001$, $\eta^2 = .218$). Forty-nine of the 70 Ps in this study rated the differences in absolute savings as more important than differences in relative savings. I also addressed this issue in a similar test by first subtracting the average rated importance for the relative dimension from the average rated importance for the absolute dimension for each participant, yielding an index that ranges from -1 (where only the relative dimension is important) to 1 (where only the absolute dimension is important). A value of zero indicates that both dimensions are equally important. I compared the mean of these difference scores (M = .22) to zero, (*one sample* $t(69) = 4.40$, $p < .001$). I repeated this analysis for each item individually. These results are reported in Table 2 under the column labeled "NP" for "normative perceptions".

Please insert Table 2 about here

Participants' agreement with the normative approach differed as a function of content domain. Participants' average differences scores (absolute – relative) were larger for human life scenarios (M = .26, SD = .42) than their differences scores for natural resource related scenarios (M = .17, SD = .44; *paired* $t(69) = 2.67$, $p = .010$, $\eta^2 = .094$). So, Ps in this study viewed absolute savings as especially important in the context of saving human lives.

Participants' differing perceptions of the relative importance of absolute savings (over relative savings) across domains did not result in differences in PD across domains. The average level of proportion dominance exhibited in human life-related scenarios (M = .37, SD = .22) did not differ from the level of proportion dominance exhibited in natural resource-related scenarios (M = .42, SD = .22).

To address the relationship between participants' normative perceptions and their interpretation of their own behavior, I created composite indices intended to reflect the degree to which Ps' ratings suggested that they themselves should have adopted a more normative approach to the task they just completed. These indices reflected Ps' own retrospective judgments concerning whether absolute savings should have influenced them more than it did (where *absolute importance – influence* is a positive number) and that relative savings should have influenced them less than it did (where *relative savings importance – influence* is a negative number). For each item, I subtracted the latter (relative savings) difference score from the former (absolute savings) difference score to arrive at measures I will refer to as "hindsight indices"[8] because these indices approximate Ps' retrospective judgments, informed by the fuller knowledge of having completed the task, about how they could have responded more normatively.

[8] Subtracting the relative savings difference score (which should be a negative number if Ps feel they should have paid less attention to relative savings) from the absolute savings difference score (which should be a positive number if Ps feel they should have paid more attention to absolute savings) in effect produces a summed approximation of Ps' feelings that their responses should have better approximated the normative approach (because subtracting a negative number is equivalent to adding a positive number).

Participants recognized a discrepancy between what should have influenced their decisions and what actually influenced their decisions. Hindsight indices predicted proportion dominance, both for an overall analysis across participants and when analyzed when analyzed separately for each item (see Table 2).

Another way to address the relationship between hindsight indices and PD is to compare the within-participant correlations. For each participant, the correlation between hindsight indices and levels of PD was computed for each item. The mean of the distribution of within-participant correlations is significantly greater than zero ($\bar{r}$ = .284, *one sample* t = 3.57, $p < .001$). This result does not change if the r's are first subjected to a Fisher transformation to correct for the skew in r's distribution. The results of both sets of analyses suggest that for items where Ps exhibited high levels of PD, they tended to indicate either that absolute savings should have influenced them more than it did, that relative savings should have influenced them less than it did or both.

Discussion

Study 2 dealt with both normative and descriptive issues. In terms of normative issues, participants generally agree with the normative approach for dealing with decision problems that pose absolute versus relative savings tradeoffs. Moreover, Study 2's results suggest that participants are able to recognize discrepancies between the normative and descriptive.

In terms of descriptive issues—how people *actually* choose—the points about normative issues are qualified by the finding that participants' normative perceptions differed across domains. What underlies this domain difference? It is possible that this difference may be rooted in participants' construal of the resource populations. Some of the justifications collected in Study 1 stressed the importance of saving every human life. Also, some justifications suggested the importance of saving the largest "part", "percentage", "portion", or "proportion" of a natural

resource "population". It could be that participants in these studies tend to view human populations as groups of individuals and tend to view natural resource populations as a collective, mass entity. Future research should address the influence of individuation on PD.

The combined findings of Studies 1 and 2 suggest both that proportion dominance varies across people as a function of thinking style and that strategies for dealing with these types of decision problems varies within individuals as a function of domain. However, there are at least two potential problems. First, the individual differences finding is correlational; second, all of the findings reported in Studies 1 and 2 concern stated policy preferences using only a strength of preference scale. The first issue is addressed in Study 3; the second issue is addressed in Study 4.

Study 3

Study 3 was conducted to replicate the major findings of Studies 1 and 2. Specifically, Study 3 was designed to test the reliability of the finding that people scored as rational thinkers exhibited less proportion dominance than people scored as intuitive thinkers. In an effort to get more statistical power through increased leverage, Study 3 examined this relationship through the use of a quasi-experimental design. Specifically, groups of participants were selected who were scored as highly rational or highly experiential based on responses given in a pre-testing session earlier in the quarter. Study 3 also used the importance and influence measures from Study 2 to assess the reliability of Study 2's findings.

Method

Five of the six scenarios used in Study 2 were used as the materials for Study 3. The method was identical to Study 2's method except for the three following changes: First, Study 3 was computerized. Second, participants now indicated their preference by moving a slider along an

unpartitioned continuum book ended by "Strong Preference for Program A" and "Strong Preference for Program B". Finally, participants were not asked to explain their preferences.

Participants

Fifty-four Northwestern University undergraduates (29 women; 25 men) participated for partial course credit. Participants were run under the same conditions as the previous studies. Participants were pre-selected for having exhibited either very high or very low REI difference scores in a mass pretesting session at the beginning of the academic quarter. Very high REI scores reflect greater reliance on deliberative thought than on intuition, while very low REI scores reflect greater reliance on intuition than on deliberative thought[9].

Results and Discussion

Please insert Table 3 about here

Participants who were scored as rational thinkers (N = 29) showed markedly less PD (M = .34, SD = .23) than participants who were scored as experiential thinkers (N = 25; M = .52, SD = .34; $t(52) = 2.72$, $p = .008$, $d = 1.0$)[10]. This difference holds for the overall comparison across participants and for some of the items; the effects sizes range from medium to large for all but one item (see Table 3). If PD is non-normative, we should expect more PD to be exhibited by those Ps who tend to reason more intuitively than analytically; those Ps show greater susceptibility to

[9] Consistent with previous research and the results of Study 1, the (identical) REI used as a pretesting instrument in Study 3 consisted of two independent scales ($r(53) = -.110$, *n.s.*), with each scale reaching a reasonable level of reliability for only 10 items (Chronbach's α = .84 and .85 for the rational and experiential indices, respectively).

[10] I report Cohen's d for 2 independent-samples comparisons. This index represents the mean difference score divided by the standard deviation of differences scores. Cohen (1977) interprets d's of .20, .50, and .80 as small, medium, and large effects, respectively.

fallacious responding elsewhere (e.g., in the framing effects literature). This finding lends indirect support to the argument that PD is non-normative. Additional evidence is garnered by the replication of Study 2's findings concerning the importance and influence measures.

Once again, participants rated differences along the absolute dimension as being more important for preference (M = .64, SD = .22) than differences along the relative dimension (M = .52, SD = .20; *paired* $t(53) = 2.059$, $p = .044$, $\eta^2 = .074$), although other analyses, conditioned on the normative perception indices as calculated in Study 2, suggest the effects observed in Study 3 are weaker than the effect observed in Study 2 (see Table 2). These differences could be attributable to procedural differences between the studies (e.g., paper- vs. computer-based task, changes in the ratings scales, etc.). Despite these differences; however, the difference in normative perceptions across domains first identified in Study 2 replicated. Participants' average differences scores (absolute – relative) were larger for human life scenarios (M = .18, SD = .47) than their differences scores for natural resource related scenarios (M = .06, SD = .45; *paired* $t(53) = 2.77$, $p = .008$, $\eta^2 = .127$). In other words, Ps' ratings were more in agreement with the normative account for scenarios entailing losses of human lives.

In addition, the predicted relationship between hindsight indices and PD obtained. As is shown in Table 2, hindsight indices predicted proportion dominance for an overall analysis across participants and within each item when analyzed separately. Again, I compared the mean of the distribution of within-participant correlations between hindsight indices and PD to zero and found that this effect replicated as well ($\bar{r}$ = .380, *one sample* $t = 5.99$, $p < .001$). Again, this result does not change if the r's are first Fisher-transformed.

Study 3 replicates the individual differences finding in a quasi-experimental design, suggesting that observed differences in PD between rational and experiential thinkers are robust.

Also, additional evidence in support of the normative-descriptive distinction was garnered from participants' ratings.

PD appears to vary across people and the underpinnings of PD vary with domain. To examine the contributions of both individual differences and domain simultaneously, a 2 x 2 mixed-model ANOVA, fitting condition (Experiential/Rational) as a between-participants factor and domain as a repeated-measures factor found an effect of condition ($F(1, 51) = 7.169$, $p = .010$, $\eta^2 = .121$), but no effect of domain ($F(1, 51) = 0.588$, $p = .447$, $\eta^2 = .011$) and no interaction ($F(1, 51) = 0.015$, $p = .901$, $\eta^2 = .000$).

While PD did not differ as a function of domain, normative perceptions varied across domains in both Study 2 and Study 3. Considering the domain differences in a realm as narrow as policy preference, one might wonder whether or not proportion dominance extends to other contexts or whether PD would be exhibited in other experimental designs. Study 4 addresses these issues.

Study 4

The materials and design of the first three studies were chosen to examine proportion dominance in joint evaluation for intervention policies with the more specific purpose of examining variability in PD across individuals and domains. I attempted to minimize variability in design to prevent confounding factors of interest with changes in design. One unintended effect of minimizing cross-study variability is that the scope of these studies was narrowed.

Study 4 introduces a number of changes with the purpose of determining more precisely how general proportion dominance is as a phenomenon. Some important differences in the design between Studies 1 - 3 and Study 4 are worth noting because they could plausibly affect whether or not PD would be observed.

First, the consequences of choice in the first three studies were grave (human deaths or natural resource depletion). It remains unclear whether or not participants' sensitivity to relative savings in and/or their choices for life and death decisions are comparable to more mundane choices. On one hand, these preferences could differ widely. Recall that Fagley and Miller (1997) found different risk preferences for scenarios involving human life than for scenarios involving major financial decisions. On the other hand, Tversky and Kahneman (1981) offer suggestive evidence that relative savings is a motivating factor for more pedestrian choices. They found that people were more willing to drive 20 minutes to save $10 if a discount was applied to a $15 calculator/jacket than when it was applied to a $125 jacket/calculator, even though Ps were told that they would be purchasing both items.

Second, the consequences of the first three studies were distributed across sizeable and nearly anonymous populations. The choices people made in these studies did not entail personal consequences (real or imagined). It remains unclear whether or not participants' choices for others or other populations are comparable to choices they would make for themselves. Hsee and Weber (1997), for instance, found that people had different risk preferences for others than they had for themselves. Self versus other discrepancies abound in the psychological literature (e.g., Camerer, Loewenstein, & Weber, 1989; Gilovich, Savitsky, & Medvec, 1998; Heath, 1999; Karniol, 2003; Nickerson, 1999; Robinson, Keltner, Ward, & Ross, 1995; Schkade & Kahneman, 1998).

Third, undergraduates are not accustomed to making life-and-death policy choices. Indeed, experts and novices differ greatly in how they reason about, solve problems, and make choices in a domain (e.g., Bailenson et al., 2002; Wiley, 1998).

Study 4 explores the generality of proportion dominance. The choice situations used in Study 4 are more likely to be familiar, relevant, and/or easily imaginable to undergraduates, and these choices entail consequences that are far less grave and are only relevant to the decision maker.

Also, Study 4 uses a matching paradigm, rather than a strength-of-preference paradigm to assess whether or not PD generalizes across tasks.

The predictions are that proportion dominance will persist even in these contexts and for this dependent measure. How people show PD in this study is by failing to match absolute savings—by providing responses that are higher in Condition A than in Condition B. The prediction and this test are more intuitively conveyed in the methods section.

Method

Participants

One hundred sixty-eight Northwestern University undergraduates participated for partial course credit. Participants were run individually. No information concerning gender was collected.

Materials & Design

Three new scenarios were developed and used as materials for Study 4. For each scenario, there were two conditions. Participants in Condition A read about two options and were asked to match Option A's attractiveness to Option B's attractiveness by supplying a value for "X". So, the conditions were defined in terms of where the "X" appeared in the scenario. Each participant read and responded to only one item. Each of the three items is presented below, with Condition A appearing in parentheses "()" and Condition B appearing in brackets "[]".

Participants read "For this question, you will be asked to make "***X***" large enough to "equalize" the attractiveness of the two options." They then read and responded to the item.

> *Department Store Item*:
> You walk into a department store and pick out a jacket and a calculator. When you reach the cash register, you notice immediate rebate coupons for both of the items you intend to purchase, but you can only redeem one instant rebate.
>
> Rebate A offers [$10] (***X***) savings on the $100 jacket.

Rebate B offers ($10) [***X***] savings on the $20 calculator.

What is the minimum amount of savings at which you'd be indifferent between rebates A and B?_____

Lawsuit Item:
You are being sued for a recent indiscretion, and your personal assets are at stake. You have two accounts, but the law only affords you protection for one account (of your choosing).

Option A saves [$1000] (***X***) of your $8000 in assets in Account A
Option B saves ($1000) [***X***] of your $2000 in assets in Account B

What is the minimum amount of savings at which you'd be indifferent between options A and B?_____

Paper Item:
It's rapidly approaching the end of the quarter and you've got 2 term papers to write. You have only got a few hours to study anything in-depth. So, you have to choose to put your efforts into studying only one topic. Assume that studying each topic would result in exactly these savings.

Studying Topic A would save [4] (***X***) hours on a paper that would otherwise take 10 hours to write.
Studying Topic B would save (4) [***X***] hours on a paper that would otherwise take 5 hours to write.

What is the minimum amount of savings at which you'd be indifferent between studying topics A and B?_____

The findings reported here focus on whether people choose to match absolute savings or relative savings. If participants choose to match absolute savings, no differences between conditions should exist, because people will just be matching the one absolute number (that is common across conditions). If participants choose to match relative savings, responses should be higher for each item in Condition A than in Condition B.

Results and Discussion

Results were analyzed two ways. First, because I did not expect Ps' responses to be normally-distributed within a given condition, a set of non-parametric (Mann-Whitney Wilcoxon) tests were performed that compared these distributions across conditions for each item. For a second set of analyses, participants' matching responses were categorized as responses that were smaller than the absolute savings, matching the absolute savings, and larger than the absolute

savings. The frequencies of responses of each type were compared within-item, across conditions in a set of chi-square analyses.

If participants only paid attention to the absolute number, the distributions of matching responses would be independent of condition—participants would not exhibit proportion dominance. However, proportion dominance was observed for each of these items (Store: Mann-Whitney $U = 206$, Wilcoxon $W = 557$, $p < .001$; Lawsuit: Mann-Whitney $U = 184$, Wilcoxon $W = 619$, $p < .001$; Paper: Mann-Whitney $U = 161$, Wilcoxon $W = 539$, $p < .001$).

Please insert Table 4 about here

The results of these analyses are supported by the chi-square analyses. The absolute and relative frequencies for responses of each type by condition, as well as the corresponding chi-square statistics and observed p-values appear in Table 4. The frequencies of absolute number matches for each item are presented in the center column in Table 4. The frequencies of smaller and larger responses are presented in the left and right columns, respectively. Furthermore, it should be noted that the modal responses in each of the highlighted cells were proportion matches. The distributions of responses for the store, lawsuit, and paper items, conditioned on condition (A versus B) appear in Figures 2, 3, and 4, respectively.

Please insert Figures 2, 3, and 4 about here

One could argue that Study 4 is the most conservative test of proportion dominance of the studies reported here both because participants should be familiar with and potentially possess some expertise in these domains and because indicating that one is indifferent between saving $10 and

$50 (in the Store Item, Condition A) or saying one is indifferent between saving $10 and $2 (in the Store Item, Condition B) seems indefensible. Nevertheless, PD was exhibited for each of these items.

Nowhere is PD more apparent than in the paper item, where the majority of responses are proportion matches (i.e., PD-type responses); relatively few people matched absolute numbers, unlike the results of the other two items. A reviewer suggested that one major difference between the paper item and the other items is that in the paper item, the alternatives to which Ps match their preferences offer relatively large relative savings.

It does not seem unreasonable to expect a positive relationship between the number of PD-type responses and the size of the relative savings offered by the alternative; large relative savings might serve as a more salient cue than small relative savings. Indeed, 62% of Ps who participated in condition A of the paper item provided PD-type responses. In this condition, the alternative offered 80% savings. However, an even greater incidence of PD-type responding (70%) was observed in condition B, where the alternative only offered 40% savings. Contrast the high incidence of PD-type responding in this condition with the lower incidence of PD-type responses in the other condition B's—where the alternatives offer greater relative savings, and the relationship between relative savings offered by the alternative and proportion matches seems less clear. Despite these nonmonotonicities, without more data that afford cleaner comparisons, I cannot rule out this (intuitively plausible) explanation for this difference.

Another possible explanation for the high incidence of proportion matches to the paper item may be that PD varies across domains. Both the lawsuit and department store items involve money (losses and gains, respectively), while the paper item involves time. Recall that many studies have shown that the processes underlying financial decisions differ from other types of decisions (e.g., Goldstein & Weber, 1995; Rettinger & Hastie, 2001; 2003). Soman (2001) contrasted financial

decisions versus choices involving time and found that although people exhibited sunk cost effects for money, they did not show sunk cost effects for time. Soman's results suggest that the mental accounting of time may differ from the mental accounting of money. These findings may suggest that people might be more likely to construct comprehensive, rather than nominal, accounts for time than for money, and thus avoid showing sunk cost effects.

In Study 4, Ps who responded to the paper item were more likely to match relative savings (66% of responses) than Ps who responded to either the department store item or the lawsuit item (32% and 36%, respectively; $p < .01$ for both comparisons). These results of Study 4 seem inconsistent with Soman's (2001) contrast in the construction of accounts for time versus money. In Study 4, Ps appear to be constructing nominal accounts (one for each paper) and matching relative savings. If Ps constructed comprehensive accounts for time, they would likely match absolute numbers. That is, in answering the question, "How much time would I have to save to be indifferent between saving 4 hours and X hours of my 15-hour total?", many Ps would be indifferent between saving 4 of 15 hours and 4 of 15 hours. However, due to the differences in motivations and methods between Soman's work and Study 4, comparisons of these projects are muddied. Future research could more directly examine people's mental accounting of time and its sensitivity to differing task demands.

General Discussion

Proportion dominance was observed in a number of different contexts in separate evaluation, joint evaluation, and matching tasks. In Study 1, Ps exhibited proportion dominance in separate evaluation, showing marked sensitivity to relative savings, replicating previous research (Baron, 1997; Fetherstonhaugh et al., 1997; Friedrich et al., 1999; Jenni & Loewenstein, 1997). In addition, these studies found PD in joint evaluation (Studies 1, 2, and 3)—despite indications that

participants perceived PD in joint evaluation to be non-normative (Studies 2 and 3)—and evidence that these effects generalize across domains and paradigms (Study 4).

Normative Status of PD in Joint Evaluation

It seems that proportion dominance in joint evaluation is non-normative, at least for the choices participants faced in these studies. In the stimuli used in the first three studies, absolute and relative savings were always inversely related, so that choosing to maximize relative savings meant saving fewer people. The individual differences effects reported in Studies 1 and 3 offer one type of support for the argument that PD in joint evaluation is non-normative. In those studies, proportion dominance was related to differences in participants' reliance on analytical thought versus intuition. These findings count as evidence because similar measures have been related to other fallacies in reasoning and decision making, and because this relationship provides weak diagnostic evidence that the normative model is the prescriptive model for this task (following the logic of Stanovich and West, 2000). However, neither the argument about the right way to approach these problems, nor the individual differences effects can conclusively argue for proportion dominance's non-normative status.

When one decides to label a phenomenon as non-normative, one needs to address a number of potential criticisms. While a thorough discussion on rationality is beyond the scope of this paper, (see Stanovich and West, 2000 and Shafir & LeBoeuf, 2002 for excellent reviews on the topic). In what follows, I address some anticipated criticisms (largely borrowed from these reviews).

Some criticisms focus on the interpretation of task performance relative to some normative model. For example, (1) some argue that some responses that appear non-normative may be simple performance errors—temporary mishaps, not systematic errors. In the case where systematic errors

are observed, (2) some argue that researchers sometimes hold participants to performance standards that are unachievable because of limitations in cognitive capacity.

Other criticisms are levied at the tasks researchers use. For instance, even if researchers are using the appropriate norm for the task they envisioned, (3) researchers sometimes present problems in an especially odd format or in an especially odd context, which can lead to participants providing especially odd responses. Also, (4) researchers sometimes violate Gricean maxims by providing participants with irrelevant information, which participants, in turn, use in arriving at the "wrong" response. On a related note, (5) researchers sometimes provide sparse descriptions of their tasks, which can invite alternative interpretations of what the task is "about". In this last case, participants may be providing the normatively appropriate response for their interpretation of the task, but the researcher would never know because the participants' interpretations were never solicited.

Gathering and analyzing participants' assessments of (non)normative responses addresses each of these concerns. The argument that proportion dominance in joint evaluation is non-normative receives direct support from Study 2's and Study 3's findings that participants endorse the normative account upon reflection. What is most striking is that participants are able to recognize and indicate a difference between their stated preference and the normative response. The fact that participants endorse the normative model addresses criticisms 2 through 5; that they indicate a normative-descriptive discrepancy addresses criticisms 1 and 2.

Generality of Proportion Dominance

Study 4 explored the generality of proportion dominance, extending the results of the first three studies to a new paradigm and to new domains. Study 4 used more mundane choices that were both more personally relevant and more likely to be familiar to undergraduates than the choices faced in the first three studies. In this study, participants were asked to provide a numerical

response that matched the attractiveness of a fully described choice option that saved X of Y of some resource. This offered participants the option of matching either absolute savings (X) or relative savings (X ÷ Y).

Since matching relative savings involved computation (in many cases, Ps worked out their solutions through cross multiplication and division—writing out all their work on the experimental materials), whereas matching absolute savings involved a simpler comparison, one might argue that Study 4 was a conservative test. Nevertheless, many participants chose to match relative savings, thus exhibiting proportion dominance. PD was observed for each of the three items, but this tendency was especially pronounced for one item offering a choice between two options that would save them some time. It is difficult to say why more, as opposed to less, PD was exhibited for this item. The exact source of variation across domains remains an important topic for future study.

A number of other issues related to the generality of proportion dominance remain open for study as well. First, it is important to explore boundary conditions for proportion dominance based on people's perceptions of the magnitude of differences in absolute and relative savings. In Studies 1 through 3, most of the populations were fairly large, most of the differences in relative savings were probably perceived to be large, and most of the differences in absolute savings were probably perceived to be small. The concentration of these studies was not on the relationship between proportion dominance and the numerical information conveyed in these stimuli.

Finally, while the generality of proportion dominance extended over a number of domains and across different ways of assessing proportion dominance in these studies, an additional meaningful step would be to examine the external and ecological validity of these findings. In addition to paper-and-pencil or computer-based lab tests run on undergraduates, an important step will be to investigate whether proportion dominance occurs "in the wild" (Hutchins, 1996), such as in the marketplace or in real-life policy writing.

Summary

These studies show that people are sensitive to relative savings, even in contexts where choice options posing absolute versus relative savings tradeoffs are presented simultaneously. Proportion dominance in these contexts is argued to be non-normative, and participants generally agree with the argument upon reflection, although predictable domain differences in normative perceptions were identified. Further, participants prone to other fallacies show the effect to a greater extent, and in general, participants are able to recognize a discrepancy between the normative and the descriptive—how they *should* versus how they do decide. Finally, the effect was shown to generalize across a number of domains and experimental paradigms.

Many kinds of decisions entail tradeoffs, and whether people choose to maximize absolute or relative savings is a very general but important question for applications ranging from the grocery store to the voting booth. These four studies demonstrate both the generality (across domains and paradigms) and the variability (inter- and intra-individual) of proportion dominance, taking small steps towards answering the questions of whether, when, and why people show sensitivity to relative savings.

References

Bailenson, J. N., Shum, M. S., Atran, S., Medin, D. L., & Coley, J. D. (2002). A bird's eye view: Biological categorization and reasoning within and across cultures. *Cognition, 84*, 1-53.

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

Baron, J. (2000). Normative and prescriptive implications of individual differences (Reply to Stanovich & West). *Behavioral and Brain Sciences, 23*, 668-669.

Birnbaum, M. H. (1999). How to show that 9 > 221: Collect judgments in a between-subjects design. *Psychological Methods, 4*, 243-249.

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

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97*, 1232-1254.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford.

Chatterjee, S., Heath, T. B., Milberg, S. J., & France, K. R. (2000). The differential processing of price in gains and losses: The effects of frame and need for cognition. *Journal of Behavioral Decision Making, 13*, 61-75.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences (revised edition)*. New York: Academic Press.

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

Fagley, N. S., & Miller, P. M. (1997). Framing effects and arenas of choice: Your money or your life? *Organizational Behavior and Human Decision Processes, 71*, 355-373.

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

Finucane, M. L., Peters, E., & Slovic, P. (2003). Judgment and Decision Making: The Dance of Affect and Reason. In S. L. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 327-364). Cambridge, UK: Cambridge University Press.

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

Frisch, D. (1993). Reasons for Framing Effects. *Organizational Behavior and Human Decision Processes, 54*, 399-429.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition, 65*, 263-297.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology 12*, 306--355.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332-346.

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

Heath, C. (1999). On the social psychology of agency relationships: Lay theories of motivation overemphasize extrinsic incentives. *Organizational Behavior and Human Decision Processes, 78*, 25-62.

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

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

Hsee, C. K., & Weber, E. U. (1997). A fundamental prediction error: Self-others discrepancies in risk preference. *Journal of Experimental Psychology: General, 126*, 45-53.

Hsee, C.K. & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86*, 680-695.

Hutchins, E. (1996). *Cognition in the Wild*. Cambridge, MA: MIT Press.

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

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

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York: Cambridge University Press.

Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1986). Fairness and the assumptions of economics. *Journal of Business, 59*, 285-300.

Karniol, R. (2003). Egocentrism versus protocentrism: The status of self in social prediction. *Psychological Review, 110*, 564-580.

Lauriola, M., & Levin, I. P. (2001). Personality traits and risky decision-making in a controlled experimental task: An exploratory study. *Personality and Individual Differences, 31*, 215-226.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16*, 77-92.

Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes, 88*, 411-429.

McElroy, T., & Seta, J. J. (2003). Framing effects: An analytic-holistic perspective. *Journal of Experimental Social Psychology, 39*, 610-617.

Medin, D. L., & Ortony, A. (1989). Psychological Essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179-195). Cambridge, England: Cambridge University Press.

Moore, D. A. (1999). Order effects in preference judgments: Evidence for context dependencies in the generation of preferences. *Organizational Behavior and Human Decision Processes, 78*, 146-165.

Nickerson, R. S. (1999). How we know - and sometimes misjudge - what others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*, 737-759.

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

Payne, J. W., & Braunstein, M. L. (1971). Preferences among gambles with equal underlying distributions. *Journal of Experimental Psychology, 87*, 13-18.

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

Rettinger, D. A., & Hastie, R. (2003). Comprehension and Decision Making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging Perspectives on Judgment and Decision Research* (pp. 165-200). Cambridge, UK: Cambridge University Press.

Ritov, I., & Kahneman, D. (1997). How people value the environment: Attitudes vs. economic values. In M. H. Bazerman, D. M. Messick, A. E. Tenbrunsel & K. A. Wade-Benzoni (Eds.), *Psychological Approaches to Environmental and Ethical Issues in Management* (pp. 33-51): New Lexington Press.

Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: Naive realism in intergroup perception and conflict. *Journal of Personality and Social Psychology, 68*, 404-417.

Sa, W., West, R. F., & Stanovich, K. E. (1999). The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. *Journal of Educational Psychology, 91*, 497-510.

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9*, 340-346.

Shafir, E. (2002). Cognition, intuition and policy guidelines. In R. Gowda & J. Fox (Eds.), *Judgments, choices, and public policy* (pp. 71-88). Cambridge, England: Cambridge University Press.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53*, 491-517.

Simon, A. F., Fagley, N. S., & Halleran, J. G. (2004). Decision framing: Moderating effects of individual differences and cognitive processing. *Journal of Behavioral Decision Making, 17*, 77-93.

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

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1980). "Facts and Fears: Understanding Perceived Risk." In R. C. Schwing & W. A. Alberts, Jr. (Eds.), *Societal Risk Assessment: How Safe is Safe Enough?* New York: Plenum Press

Slovic, P., & Lichtenstein, S. (1968). Importance of variance preferences in gambling decisions. Journal of Experimental Psychology, 78, 646-654.

Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4*, 108-131.

Soman, D. (2001). The mental accounting of sunk time costs: Why time is not like money. *Journal of Behavioral Decision Making, 14*, 169-185.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*, 342-357.

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*, 161-188.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*, 645-665.

Strack, F., & Deutch, R. (in press). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*.

Tordesillas, R. S., & Chaiken, S. (1999). Thinking too much or too little? The effects of introspection on the decision-making process. *Personality and Social Psychology Bulletin, 25*, 623-629.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Wiley, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*, 716-730.

Wilson, T. D., Lisle, D. J., & Kraft, D. (1990). Effects of self-reflection on attitudes and consumer decisions. *Advances in Consumer Research, 17*, 79-85.

Wilson, T. D., Lisle, D. J., Schooler, J. W., Hodges, S. D., Klaaren, K. J., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality and Social Psychology Bulletin, 19*, 331-339.

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

Figures

Figure 1 – Stimuli from Hsee and Leclerc (1998)

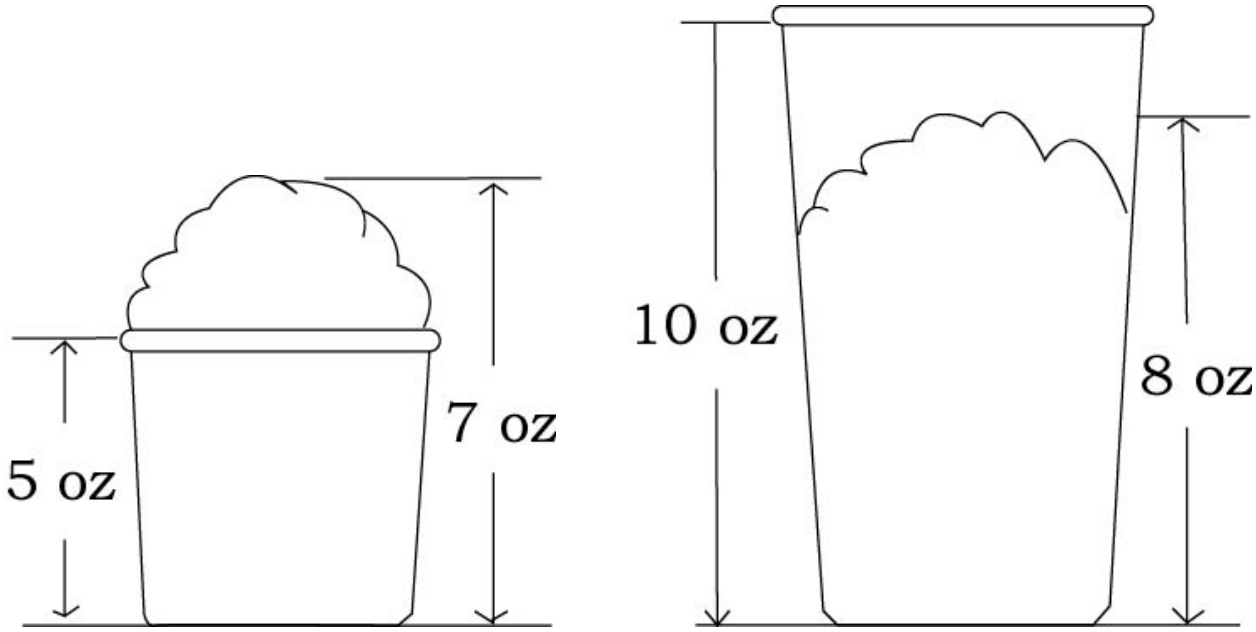


Figure 2 – Frequencies of Matching Responses by Condition for the Department Store Item (the absolute number match is 10)

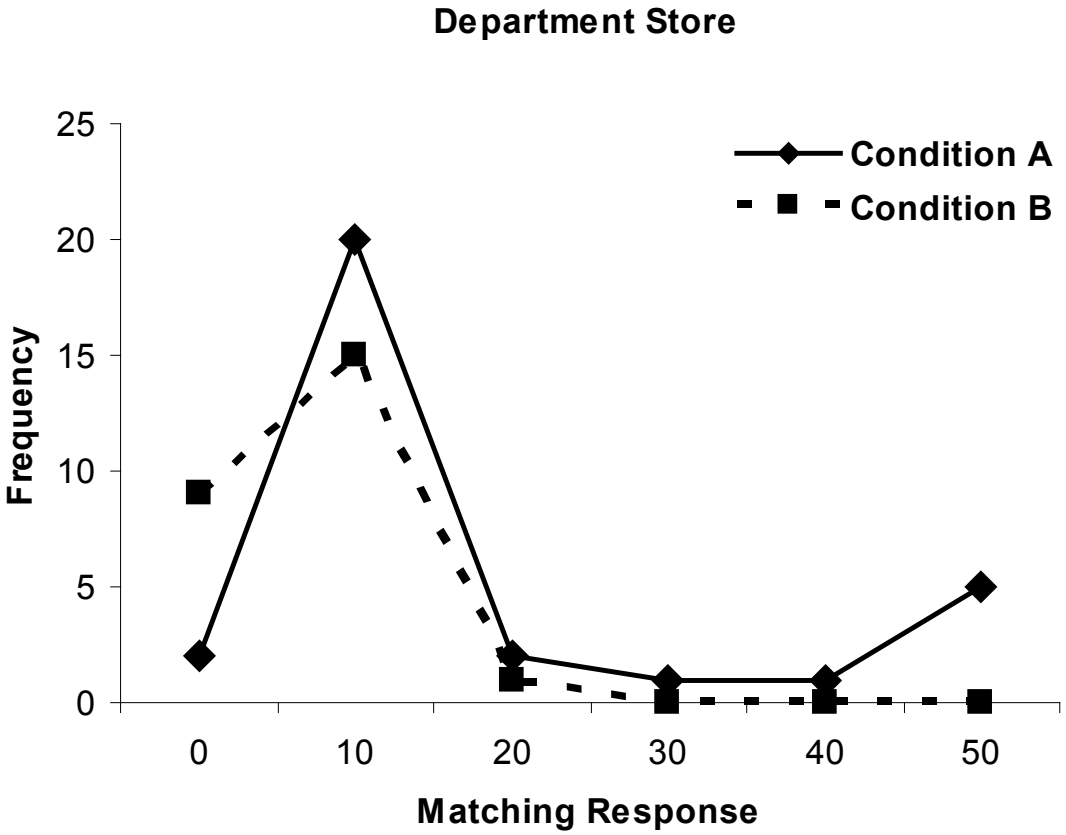


Figure 3 – Frequencies of Matching Responses by Condition for the Lawsuit Item (the absolute number match is 1000)

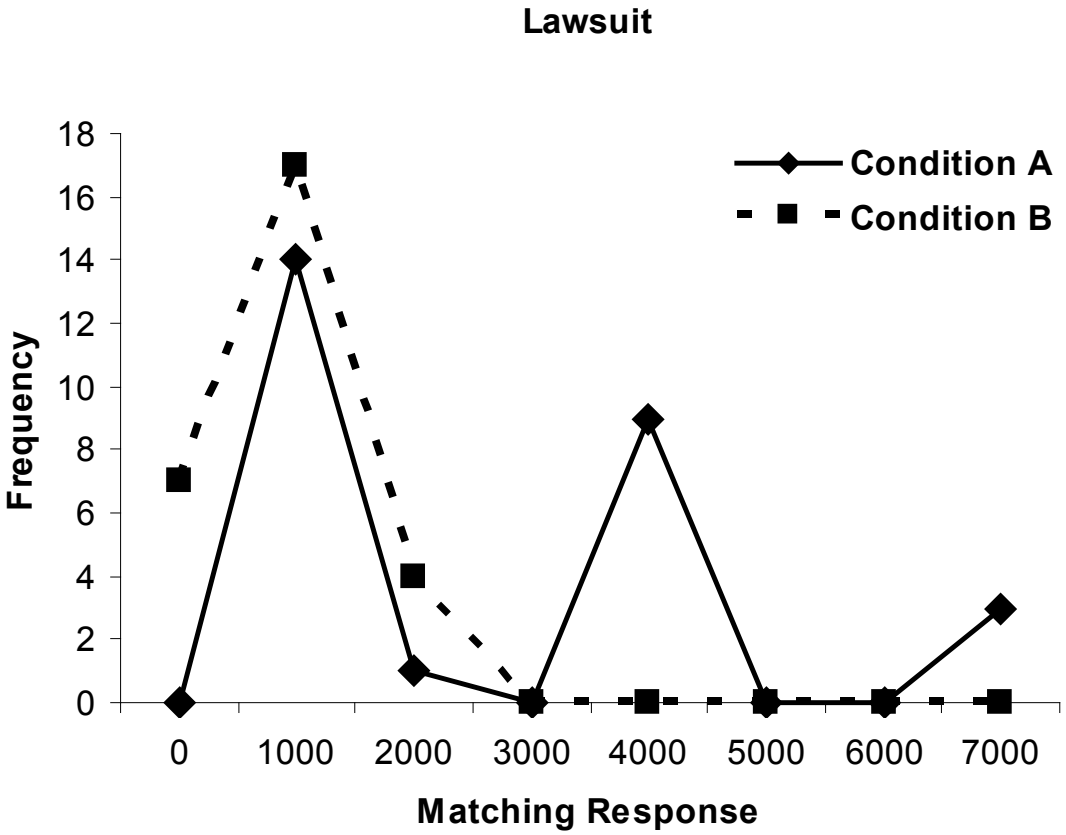


Figure 4 – Frequencies of Matching Responses by Condition for the Paper Item (the absolute number match is 4)

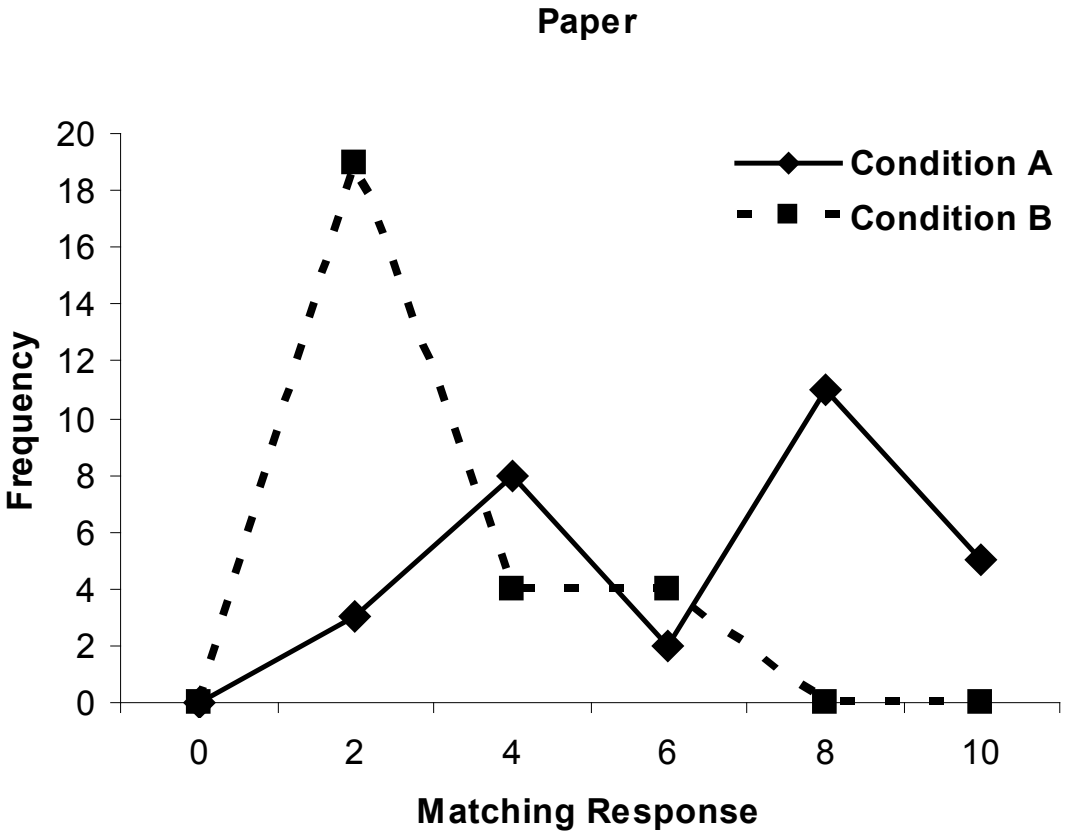


Table 1

Disease	Annual Deaths	Deaths Prevented
X	15,000	10,000
Y	160,000	15,000
Z	290,000	20,000

Table 2

Summary of Normative Perceptions, Correlations between Hindsight Indices and PD

Analysis	Study 2			Study 3		
	NP	*r*	$\eta 2$	NP	*r*	$\eta 2$
Across-Ps	0.22 ***	0.385 ***	0.136	0.11 †	0.301 *	0.073
Anthrax	0.30 ***	0.437 ***	0.179	0.20 **	0.337 *	0.097
Otters	0.15 *	0.447 ***	0.188	0.05	0.423 **	0.163
Paper	0.13 *	0.384 ***	0.135	0.00	0.351 **	0.106
Tuna	0.20 ***	0.315 **	0.086	0.12	0.484 ***	0.219
War	0.25 ***	0.356 **	0.113	n/a	n/a	n/a
Zaire	0.24 ***	0.519 ***	0.259	0.15 †	0.323 *	0.087

$p < .10$ †
$p < .05$ *
$p < .01$ **
$p < .001$ ***

Table 3

Analysis	Study 3 - Proportion Dominance by Thinking Style			
	Experiential	Rational	t	d
Across-Ps	0.52	0.34	2.720 **	0.611
Anthrax	0.48	0.31	1.941 †	0.530
Otters	0.49	0.38	1.075	0.293
Paper	0.48	0.30	2.015 *	0.550
Tuna	0.57	0.30	3.090 **	0.843
Zaire	0.59	0.41	1.872 †	0.511

$p < .10$ †

$p < .05$ *

$p < .01$ **

$p < .001$ ***

Table 4

Department Store Item

Condition	< 10	10	> 10
A	2 (6%)	20 (65%)	**9 (29%)**
B	**9 (36%)**	15 (60%)	1 (4%)

$\chi 2(2) = 12.25, p = .002$

Lawsuit item

Condition	< 1000	1000	> 1000
A	0 (0%)	14 (52%)	**13 (48%)**
B	**7 (25%)**	17 (61%)	4 (14%)

$\chi 2(2) = 14.99, p = .001$

Paper Item

Condition	< 4	4	> 4
A	3 (10%)	8 (28%)	**18 (62%)**
B	**19 (70%)**	4 (15%)	4 (15%)

$\chi 2(2) = 23.90, p < .001$

Appendix A: Scenarios Used in Studies 1-3

Anthrax. Anthrax powder has been weaponized and released into the air above two mid-sized cities. In each city, a number of people are expected to die as a result of anthrax inhalation. There exists a powerful antibiotic that will successfully treat some victims, but there is a limited amount of this treatment. Program A would delegate the treatment to City A, and 225 of the 300 at risk of death will be saved. Program B would delegate the treatment to City B, and 230 of the 920 people at risk of death will be saved. These programs are mutually exclusive and the only two options available.

Birds. An amusement park is nearing the final stages of planning before construction when it is found that construction will destroy some trees where an endangered species of songbird nests. The planners are willing to adopt one of two proposed solutions to the problem. Program A saves 19 of the 25 birds that nest Area A. Program B saves 20 of the 400 birds that nest in Area B. These programs are mutually exclusive and the only two options available.

Jobs. The current recession has forced companies to cut jobs. Your office provides financial support to struggling businesses in the local economy, but limited resources force you to choose which businesses to assist. Program A saves 54 of the 60 jobs that would have otherwise been lost at Factory A. Program B saves 56 of the 560 jobs that would have otherwise been lost at Factory B. These programs are mutually exclusive and the only two options available.

Otters. An oil spill around Puget Sound is threatening the sea otter populations in two areas of the bay. Two cleanup plans are proposed, but there is only enough money to support one plan. So, there are only enough resources to save otters in one of these areas of the bay. Program A will save 120 of the 150 otters near the north end of the bay. Program B will save 124 of the 800 otters near the south end of the bay. These programs are mutually exclusive and the only two options available.

Paper. You are on a committee at a major paper company with two factories on a mid-sized river. These factories use water from the river to cool their machines. Once used, the water is exhausted back into the stream. This polluted water causes a number of fish to die every year near the factory from which it is exhausted. Filters can be installed that will save a number of fish, but filter installation is expensive, and there is only enough money in the budget to install filters at one factory. Program A filters the water exhausted from Factory A, resulting in the prevention of 245 of the annual 350 fish deaths due to pollution. Program B filters the water exhausted from Factory B, preventing 251 of the annual 980 fish deaths due to pollution. These programs are mutually exclusive and the only two options available.

Plants. A species of plant found only in a remote area of New Guinea is threatened with extinction by a recently introduced species of vine. You have access to two treatments that kill the vines and save the plants, but you only have enough money to fund one program. If you implement Program A, you will save 160 of the

200 plants located in Quadrant A. If you implement Program B, you will save 164 of the 820 plants located in Quadrant B. These programs are mutually exclusive and the only two options available.

Preserves. The city council is auctioning off two former nature preserves for commercial development. The corporations that buy the land will bulldoze the land and erect office buildings. Your organization protects public spaces by purchasing lots like these and preserving them. There is only enough money to purchase a portion of the land up for auction. If you implement Program A, you can purchase 22 of the 70 acres in Plot A, saving them from development. If you implement Program B, you can purchase 24 of the 30 acres in Plot B, saving them from development. These programs are mutually exclusive and the only two options available.

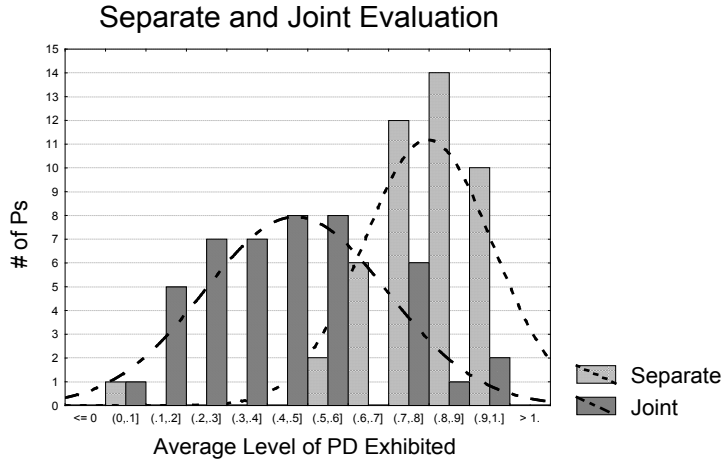
Tuna. Two areas off the southeast coast of Florida are heavily populated with dolphins and tuna. Tuna fishermen accidentally catch a number of dolphins in these areas every year. Dolphins that get caught in the tuna nets drown, because they cannot surface to breathe. To combat this problem, new nets have been designed that will save a number of dolphins. The tuna fishing industry has agreed to fish with the new nets in only one of these two areas. Program A would require boats in Area A to use a different type of net, which would save 180 of the 240 dolphins that die in that area each year. Program B would require boats in Area B to use a different type of net, which would save 187 of the 900 dolphins that die in that area each year. These programs are mutually exclusive and the only two options available.

War. Emergency medical aid is needed at two different sites in a war-torn country. You are the head of the only medical convoy in the area. There is only enough time, and there are only enough supplies, to visit one camp. Treating Camp A will save the lives of 68 of the 80 patients who would otherwise die. Treating Camp B will save the lives of 70 of the 680 patients who would otherwise die. These treatment programs are mutually exclusive and the only two options available.

Zaire. Recent political developments in Zaire have severely marginalized some of the population. These refugees are clustered about in two camps, struggling to survive, because very little clean water is available. A plane with water treatment capabilities will be sent. There is only enough fuel, supplies, and time to visit one camp. Program A would treat enough water to save 102 refugees in the camp of 115. Program B would treat enough water to save 105 refugees in the camp of 700. These programs are mutually exclusive and the only two options available.

Appendix B: "Wiggle-Room" in Separate and Joint Evaluation

As is the case with many psychological effects, PD is at least somewhat influenced by its elicitation. Study 1 elicited responses on the same scale for both separate evaluation and joint evaluation. The scale was univalent for separate evaluation and bivalent for joint evaluation. The figure below depicts Ps' average scale responses in separate evaluation and joint evaluation.



Notice that the distribution of joint evaluation responses appears less peaked and more spread out than the distribution of separate evaluation responses (averaged across the small- and large reference group conditions for each participant). This qualitative observation is supported by a comparison finding a reliable difference in the variances of these samples. It is worth noting that although this *F*-test, $F\left(\frac{s^2_{JE}}{s^2_{SE}}, df_{JE}, df_{SE}\right)$, suggests a significant difference ($F(43,43) = 4.108$, $p < .001$), the comparison across joint- and separate evaluation is hindered by the possibility of scale-end effects in the separate evaluation distribution. Although the distribution of separate evaluation responses does not significantly differ from normality, potential scale-end effects may have reduced variability in the separate evaluation distribution. Nevertheless, the data yield qualitative and some (more tenuous) quantitative evidence that separate evaluation designs can afford Ps less "wiggle-room" than joint evaluation designs.