Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Multiple Goals as Reference Points: One Failure Makes Everything Else Feel Worse

Evan WeingartenSudeep BhatiaBarbara Mellers

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–16
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Multiple Goals as Reference Points: One Failure Makes Everything Else Feel Worse

Evan Weingarten,[a] Sudeep Bhatia,[b] Barbara Mellers[b]

[a] University of California San Diego, La Jolla, California 92093; [b] University of Pennsylvania, Philadelphia, Pennsylvania 19104

Contact: eweingarten@ucsd.edu, http://orcid.org/0000-0001-9392-0102 (EW); bhatiasu@sas.upenn.edu, http://orcid.org/0000-0001-6068-684X (SB); mellers@wharton.upenn.edu (BM)

Received: February 13, 2017
Revised: November 5, 2017; February 4, 2018
Accepted: February 18, 2018
Published Online in Articles in Advance: July 17, 2018

https://doi.org/10.1287/mnsc.2018.3097

Abstract. It is well known that goals serve as reference points, and their influence on pleasure can be understood with prospect theory's value function. We examine how people feel about their progress on two goals (i.e., academics and fitness). What happens when they achieve one goal but fail to reach another? In four studies, we test the assumptions needed to explain hedonic reactions to progress along two goals. Loss aversion and diminishing sensitivity hold on each variable separately. However, we find violations of additivity in the integration of the emotions about outcomes. A success in one goal and a failure in another feel worse than the sum of the pleasure and pain associated with the gain and loss, respectively.

History: Accepted by Elke Weber, judgment and decision making.
Funding: The authors thank the Wharton Risk Management and Decision Process Center for generous funding via a Russell Ackoff Doctoral Student Grant. Funding for S. Bhatia was received from the National Science Foundation [Grant SES-1626825].
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2018.3097.

Keywords: goals • reference points • prospect theory • loss aversion • additivity

Introduction

Each day we pursue goals for an array of activities, including exercising, working, relaxing, eating, and saving money. We balance time, effort, and money to achieve these goals. Occasionally, we achieve them all, but more likely, we fall short on one or more. In this paper, we examine how people feel about successes and failures in settings with multiple goals.

Others have studied how people feel about success or failure on a single target goal (Heath et al. 1999). Goals function as reference points in a manner consistent with prospect theory's value function (Heath et al. 1999). For example, Markle et al. (2018) asked marathon runners to forecast or report their satisfaction with exceeding or falling short of their goals for a race. Predicted and actual satisfaction were both consistent with loss aversion and diminishing sensitivity. The impact of a loss was greater than a gain of equal magnitude (loss aversion), and a fixed change in performance relative to a goal had a greater impact on pleasure when it occurred closer to the goal than when it was further away (diminishing sensitivity).

Does prospect theory extend to settings with multiple goals? Do loss aversion and diminishing sensitivity occur along multiple variables (i.e., two academic goals for two separate classes)? Do feelings about one goal combine additively with feelings about another? That is, are outcomes aggregated independently, such that the overall emotional response is predicted by the sum of the two separate feelings?

Here, we test prospect theory's value function and additive forms of multiattribute utility theory as a representation of feelings about progress on multiple goals (Keeney and Raiffa 1976). We find evidence of loss aversion and diminishing sensitivity along each variable (i.e., different academic goals for two separate classes). However, feelings about one goal are not independent of feelings about the other. Instead, the pleasure of reaching (or exceeding) one goal and falling short on another is less than predicted from the sum of the feelings associated with an independent loss and gain.

Pleasure and Single Reference Points

Reference points are discrete levels above which people perceive outcomes as gains and below which they perceive outcomes as losses (Kahneman 1992). Such breakpoints serve as the cornerstone of prospect theory's value function and other reference-dependent theories of choice (Kahneman and Tversky 1979, Kőszegi and Rabin 2006). A reference point could be the endowment of a gift or product (Kahneman, Knetsch, and Thaler 1990, Brenner et al. 2007), current wealth (Kahneman and Tversky 1979), familiar prices (Weaver and Frederick 2012), expectations (Cherry et al. 2003, Novemsky and Kahneman 2005), counterfactuals (Mellers 2000), the status quo (Samuelson and Zeckhauser 1988), or social comparisons (Festinger 1954).

In a classic paper, Heath et al. (1999) showed how target levels of performance, or goals, serve as reference points in a manner consistent with the value function in prospect theory. That is, prospect theory's value function is applied to satisfaction. Using scenario-based studies, Heath et al. found evidence of loss aversion. In one scenario, both Sally and Trish performed 35 sit-ups. Sally was *above* her goal of 31 sit-ups, whereas Trish was *below* her goal of 39 sit-ups. Sally exceeded her goal by 4 sit-ups, and Trish fell short of hers by 4 sit-ups. When asked, "Who is experiencing more emotion?" (p. 86), participants said it was Trish, which is consistent with the premise that losses have greater impact than equivalent gains.

Heath et al. (1999) also found support for diminishing sensitivity, which means a change in progress closer to the reference point has greater impact than the same change further away. For example, in two scenarios, participants read about Charles and David, both of whom had done either 28 sit-ups (first scenario) or 42 sit-ups (second scenario). In both cases, Charles had a goal of 30, and David had a goal of 40. When asked which person would work harder to meet his goal, participants said it was the individual who was closer to his goal, consistent with diminishing sensitivity.

Happiness and Multiple Reference Points

Multiple reference points have been studied in two ways. First, multiple goals can occur on a single variable. Markle et al. (2018) demonstrated how satisfaction for running times was influenced by multiple time goals, including best finish times, personal goals, and most recent finishes. In another example, corporate investment decisions of managers were influenced by the company's current performance and its target performance (Sullivan and Kida 1995). Finally, Ordóñez et al. (2000) found that satisfaction with a given salary offer depended on the offer, as well as offers made to others.

Theories of multiple reference points on a single variable include trireference point theory and decision affect theory (Mellers et al. 1997, 1999; Wang and Johnson 2012). Trireference point theory describes the effects of three reference points on risky decision-making for gambles: the minimum amount needed or required, the status quo, and the aspiration level (Wang and Johnson 2012). Decision affect theory describes the pleasure of outcomes when counterfactual outcomes such as disappointment (outcomes under alternative states of the world) and regret (outcomes from different choices) are known to the decision maker (Mellers et al. 1997, 1999).

Multiple reference points also occur when single goals exist along more than one variable. Such contexts are the focus of our analysis. Decision makers strive for outcomes in professional, financial, or family domains. Progress along one variable may influence the motivation to pursue another. The goal-setting literature has examined such interactions (Locke and Latham 1990). When people set a specific goal for an activity, they often put greater effort toward that activity and less effort toward another (Schmidt et al. 1984). Studies also show that people attempt a second goal after they have made some progress on the first (focal) goal, a process known as "balancing" (Dhar and Simonson 1999). Balancing is more common when people think that success on the focal goal is either likely or impossible to reach (Fishbach and Dhar 2005, Louro et al. 2007). However, when people continue to pursue a focal goal instead of a second goal, a process known as "highlighting," they often view their effort toward a focal goal as a sign of commitment (Fishbach and Dhar 2005, Finkelstein and Fishbach 2010).

Violations of Additivity in Judgment and Choice

The simplest approach to modeling the pleasure of two outcomes on different variables is to assume additivity of feelings about each outcome. Additivity is a common assumption and is supported by normatively compelling axioms (Keeney and Raiffa 1976, Fishburn and Wakker 1995). However, researchers have documented numerous violations in judgment and choice. For example, when participants rate the likeableness of people described by two adjectives, the worse adjective has greater impact (Birnbaum 1974). A person described by a positive and a negative adjective is only slightly more likeable than one described by two negative adjectives, and that person is judged very differently from someone described by two positive adjectives. Similarly, when participants judge the morality of others, a negative deed paired with a positive deed leads to a precipitous drop relative to two positive deeds, just slightly higher than two negative deeds. Yet when participants judge their own morality, a negative deed paired with a positive deed is "excused" and is much closer to two positive deeds. Individuals seem to forgive their own sins, although they are unlikely to forgive the sins of others (Birnbaum 1972, 1973; Riskey and Birnbaum 1974).

Violations of additivity occur in other multiattribute contexts. Fischer (1976) examined the evaluations of jobs based on salary, occupation, and location in both riskless and risky settings. He tested independence (preference rankings for an attribute are the same across all other attributes), joint independence (preference ranks for two attributes, holding another attribute constant, should be independent from the other attribute), and the cancellation assumption in conjoint measurement. Although there were occasional violations of simple independence and joint independence, violations of cancellation were frequent (Fischer 1976). More recently, Bhatia (2018) found similar joint independence violations in choice, as well as joint independence violations in settings where rewards are learned from experience. Finally, Birnbaum (2008) offers a model of risky decision making that permits violations of cancellation and a related property known as coalescing. He shows that a large number of effects in judgment and decision making require more flexibility than additivity.

Different types of interactions require different models. We focus on two models that we will call the minimum model and maximum model. The minimum model predicts that falling short on either (or both) goals will generally result in dissatisfaction. The pleasure of a success is spoiled by the displeasure of a failure; failures have greater impact. The minimum model predicts a divergent interaction: feelings about a success and failure are more similar to feelings about two failures than to feelings about two successes. By contrast, the maximum model predicts that achieving or exceeding either goal will generally result in satisfaction. The displeasure of a failure is buffered by a success; successes have greater impact. The maximum model predicts a convergent interaction: feelings about a gain and a loss are more similar to feelings about two gains than to feelings about two losses.

Our models are inspired by nonadditive theories of multiattribute decision making previously introduced by Dawes (1964) and Einhorn (1970). Dawes proposed a "minimum evaluation function," in which the utility of a multiattribute option is determined by the worst attribute, and a "maximum evaluation function," in which the utility of an option is determined by the best attribute. Einhorn (1970) called these models "conjunctive" and "disjunctive," respectively. According to the conjunctive model, an option must be above a minimum standard on all attributes to be desirable. According to the disjunctive model, an option is evaluated based on the best attribute, regardless of the other attributes.

Our minimum model resembles Dawes' minimum evaluation function and Einhorn's conjunctive model; it emphasizes the impact of the worst variable. The decision maker experiences overall disutility if at least one variable entails a loss (i.e., falls below the goal or standard). Our maximum model resembles Dawes' maximum evaluation function and Einhorn's disjunctive model. It emphasizes the impact of the best variable. The decision maker experiences greater overall happiness if at least one variable entails a gain (i.e., reaches or surpasses the goal). We prefer the terms "minimum" and "maximum" models because they refer to the focus of attention on the overall emotion.

Minimum and maximum models have much in common with the configural-weight model proposed by Birnbaum (1974) and his colleagues (i.e., Birnbaum and Stegner 1979, Birnbaum et al. 1992). Configural weighting predicts that people are more attentive to positive or negative information depending on their point of view. For example, the buyer of a used car is likely to be more sensitive to the negative attributes than the positive attributes of the car. In this case, configural weighting resembles the minimum model and predicts a divergent interaction between attributes. A seller of that same car is likely to focus on the car's positive features more than the negative ones. Now, configural weighting resembles the maximum model and predicts a convergent interaction between attributes.

Theoretical Assumptions

To formalize the models, we begin with the prospect theory value function. We represent the value of outcome of x_i on variable i for a subject with a goal g_i as $V_i(x_i|g_i)$. For simplicity, we define $x = (x_1, x_2, \ldots)$ and $g = (g_1, g_2, \ldots)$ as vectors of outcomes and goals, respectively. Adopting the Tversky and Kahneman (1991) value function, $V(x_i|g_i)$ is expressed as follows:

$$V_i(x_i|g_i) = \begin{cases} (x_i - g_i)^\alpha & \text{if } x_i \geq g_i, \\ -\lambda \cdot (g_i - x_i)^\alpha & \text{if } g_i > x_i, \end{cases} \tag{1}$$

where the coefficient α is the exponent in the power functions for utility and disutility. It represents diminishing sensitivity when $\alpha < 1$. The coefficient λ represents loss aversion in prospect theory. Losses have greater impact than gains when λ is greater than 1.

In multiattribute utility theory (Keeney and Raiffa 1976), values may be combined additively, such that the total value of the two outcomes relative to two goals, $U(x|g)$, is

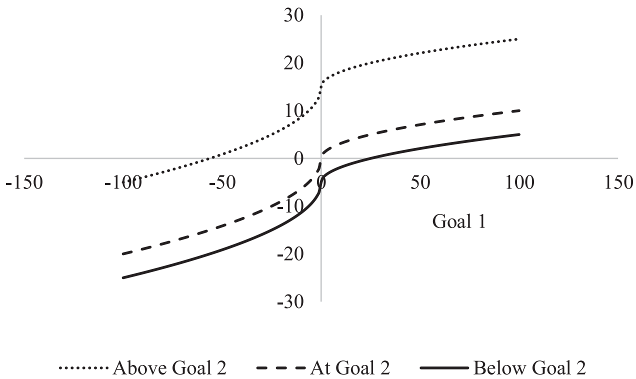
$$U(x|g) = V_1(x_1|g_1) + V_2(x_2|g_2). \tag{2}$$

Equations (1) and (2) lead to clear predictions:

1. *Loss aversion*: If $\lambda > 1$, a negative outcome relative to a goal should have a bigger impact on happiness than an equally large positive outcome. Loss aversion is illustrated in Figure 1, in which the slope of the line below zero is steeper than it is above zero.

2. *Diminishing sensitivity*: If $\alpha < 1$, changes in outcomes should have a greater effect on happiness when those changes are closer to a goal than when they are

Figure 1. Multiattribute Prospect Theory Value Function



Notes. The slopes of the curves in the loss domain are steeper than those in the gain domain, showing loss aversion. The curves also exhibits diminishing sensitivity: changes closer to 0 (i.e., goal attainment) have a bigger impact on happiness than changes farther from 0. Furthermore, there is additivity across dimensions: additivity across variables implies that the dotted, dashed, and solid lines should be parallel.

further from it. Diminishing sensitivity is also illustrated in Figure 1. Here, the line is concave in the gain domain and convex in the loss domain.

3. *Additivity*: The emotional effects of outcomes on one variable do not depend on outcomes on the other variable. Additivity is illustrated in Figure 1. The dotted, dashed, and solid black lines correspond to falling short of, achieving, and exceeding a second goal, respectively. The slopes of the lines reflect performance relative to a first goal, shown on the abscissa. Additivity implies that the lines depicting progress toward the second goal should be parallel across levels of progress toward the first goal.

As discussed above, we also consider models that relax the assumption of additivity. These models allow for added displeasure from falling short on one or both goals. Using an indicator function T, we represent these models as follows:

$$U(x|g) = V_1(x_1|g_1) + V_2(x_2|g_2) + \omega \cdot T, \qquad (3)$$

where T takes on a value of either 1 or 0. This binary notation simplifies the interpretation of the parameters and makes them more intuitive. The notation also is reminiscent of the cutoffs in conjunctive and disjunctive models from Dawes (1964) and Einhorn (1970).

With the minimum model, one or more failures decreases pleasure. The function T is defined by

$$T_{\min} = \begin{cases} 1 & \text{if } x_1 \geq g_1 \text{ AND } x_2 \geq g_2, \\ 0 & \text{otherwise.} \end{cases} \qquad (4)$$

Here, the coefficient ω in Equation (3) is positive and represents the pleasure of successes on both goals. A failure on either goal spoils the pleasure of a success. This pattern can be seen when comparing Figure 2, panel (a), in which additivity is present, to Figure 2, panel (b) (i.e., the minimum model). In Figure 2, panel (b), the happiness associated with one gain and one loss (B and C) is closer to that of two losses (D) than two gains (A). In Figure 2, panel (a), however, the lines are parallel.

With the maximum model, T is defined by

$$T_{\max} = \begin{cases} 1 & \text{if } x_1 \geq g_1 \text{ OR } x_2 \geq g_2, \\ 0 & \text{otherwise,} \end{cases} \qquad (5)$$

where the weight ω in Equation (3) is positive and represents the pleasure of a success, regardless of the other outcome. This pattern appears in Figure 2, panel (c): the happiness associated with one gain and one loss (B and C) is closer to that of two gains (A) than two losses (D). Here, feelings associated with a loss are buffered by a gain. Just one success makes the overall experience more positive.

To predict happiness ratings, we need a response function that maps feelings into the response scale. We also need to specify errors to accommodate variability in our data. For tractability, we assume that the response function is linear and that error, ε, is normally distributed around zero, with standard deviation σ. This is shown in Equation (6). Here, $U(x|g)$ could be obtained from either Equation (2) or Equation (3), or from the various alternative models we outline in the Web appendix:

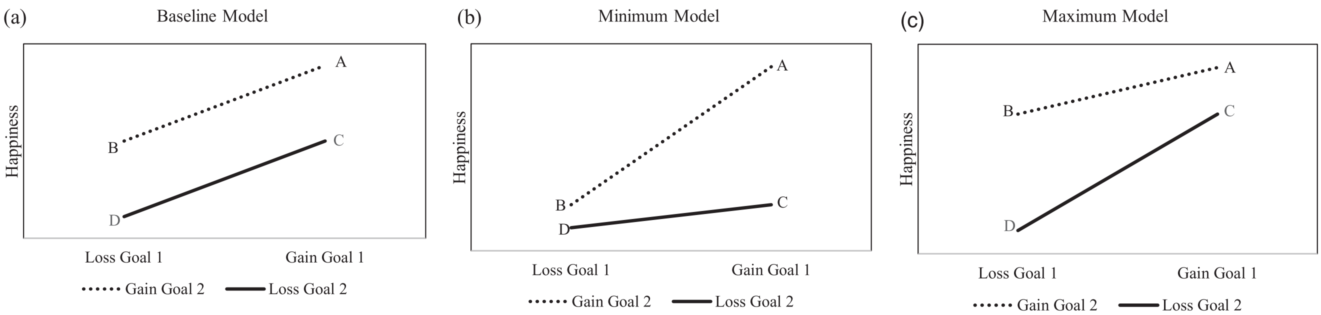
$$R(x|g) = a + b \cdot U(x|g) + \varepsilon. \qquad (6)$$

Equations (3)–(6) are not the only ways to represent the data. As discussed in the previous section, the configural-weight model (e.g., Birnbaum 1974) provides a closely related framework for studying the interactions. With this model, ω incorporates the magnitude of the interaction. Simpler models could also apply, including those that describe the overall emotion based on the best or the worst outcome relative to its corresponding goal.

The Present Research

Following Heath et al. (1999), we employ scenario studies to study multiple goals. Participants are instructed to rate their own happiness or another person's happiness after progress on two outcomes relative to two goals along different variables. We test additivity and the predictions of prospect theory. In Studies 1a and 1b, we use small designs and examine aggregate data. These studies offer simple tests of loss aversion and additivity, and participants make very few responses. Study 1a finds patterns consistent with loss aversion and reference dependence, and Study 1b finds violations of additivity that are consistent with a minimum model.

Figure 2. Additivity in Judged Happiness With Progress Along Two Goals (a) Compared With Violations of Additivity According to a Minimum Model (b) or a Maximum Model (c)



Notes. Point A represents the feelings associated with two gains, points B and C represent the feelings associated with one gain and one loss, and point D represents the feelings associated with two losses. Additivity across variables in panel (a) shows that the dotted and solid lines are parallel. Panel (b) presents a divergent interaction for the minimum model, whereas panel (c) presents a convergent interaction for the maximum model.

Studies 2 and 3 simultaneously test loss aversion, diminishing sensitivity, and additivity using the fits of models to aggregate and individual data. These designs are larger and require more responses. We replicate the results with different variables (i.e., grades in two classes and exercises of two types) to demonstrate generalizability. The results provide further evidence for loss aversion, diminishing sensitivity, and a divergent interaction that violates additivity and is consistent with a minimum model.

The Web appendix explores alternative models, including separate weights for each goal in Equations (3)–(5), in addition to an interaction term. We also examine a model proposed by Markle et al. (2018), who found that achieving a goal was pleasurable. They represented this jump at the goal with an additive constant. Our analyses suggest the minimum model that allows a divergent interaction outperforms alternative models without interactions or with convergent interactions (i.e., maximum models).

In all of the studies, we use a 101-point scale (−50 = extremely unhappy to 50 = extremely happy) to measure feelings. Such scales are common in the literature (e.g., Mellers et al. 1997, 1999). However, we recognize that prior research has voiced concerns about bipolar scales for measuring loss aversion. McGraw et al. (2010) contend that bipolar scales may not force people to directly compare a gain to a loss, masking loss aversion. They claim that unipolar scales (or relative intensity scales) more directly compare losses to gains. When people judge combinations of outcomes (i.e., one loss and one gain), we cannot assume they will feel either good or bad, which complicates the use of unipolar scales. Yet that means our tests of loss aversion with bipolar scales are more conservative, so any evidence we observe for loss aversion should be more compelling. We also acknowledge that some researchers have concerns about whether happiness ratings are ordinal or cardinal scales (Frey and Stutzer 2002a, 2002b). We treat the scales as cardinal and recognize that if we used a standard ordinal regression, we would need one parameter for each rating, requiring more than 100 parameters in the regression model. However, nonparametric techniques that allow nonlinearities without requiring many additional parameters could be used as an alternative approach (e.g., see the spline-based techniques outlined in Markle et al. 2018). We will investigate these ideas in future research.

Study 1a: A Simple Test of Loss Aversion

In Study 1a, we test loss aversion in a multiple goal setting. We randomly assign participants to one of two goal conditions with two classes, math and English. Participants rate how happy they would feel about receiving the grades. The different goals are: (1) goal achievement in both classes, (2) a small loss on a goal in one class and a small gain of equivalent magnitude in the other, and (3) a large loss on a goal in one class and a large gain of equivalent magnitude in the other. Loss aversion means that ratings for $(0, 0) > (+x, -x) > (+X, -X)$, where $0 > x$ (small loss) $> X$ (big loss). Loss aversion predicts that participants should be happiest when they achieve both goals, less happy with a smaller loss and gain relative to their goals, and least happy with a larger loss and gain relative to their goals. Changing goals across participants should lead to changes in relative happiness.

Method

One hundred and sixty-nine participants from Amazon Mechanical Turk completed this study in exchange for $0.20.

Participants were randomly assigned to one of two goals in English and math classes. One pair was grades of 90 and 60 on English and math quizzes, respectively, and the other was grades of 60 and 90 on English and math quizzes, respectively. They rated their feelings about pairs of grades relative to one of the two goal pairs on a 101-point scale (−50 = extremely unhappy to 50 = extremely happy). Grade pairs were (1) 90 in English and 60 in math, (2) 75 in English and 75 in and math, and (3) 60 in English and 90 in math. For those with goals of 90 in English and 60 in math, the first pair of grades corresponds to goal achievement, the second pair corresponds to a small gain and a small loss, and the third pair corresponds to a large gain and a large loss. This mapping is reversed for participants with goals of 60 in English and 90 in math.

Results and Discussion

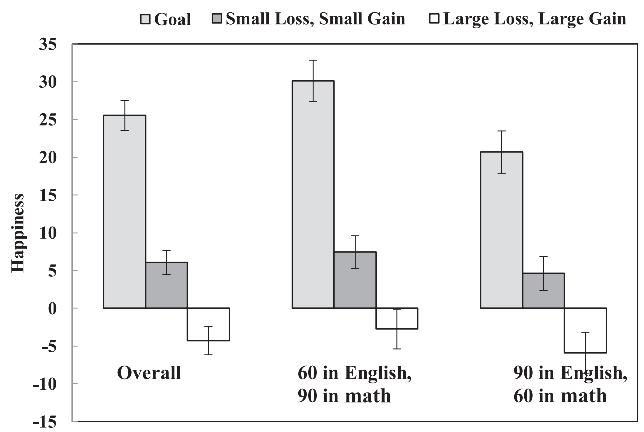
We tested predictions about happiness ratings using an analysis of variance (ANOVA) run in PROC MIXED with unstructured covariance (Wolfinger and Chang 1998). Participants were happier with grades of 60 in English and 90 in math when those scores represented goal achievement (M = 30.13, SE = 2.72). But when the same grades implied a large loss and large gain (i.e., with goals of 90 in English and 60 in math), participants were much less happy (M = −5.88, SE = 2.71; $F(1, 167) = 88.09$, $p < 0.001$). Similarly, participants were happier with grades of 90 in English and 60 in math when the grades implied goal achievement (M = 20.68, SE = 2.80). But when those grades were a large loss and large gain relative to goals, participants were much less pleased (M = −2.75, SE = 2.63; $F(1, 167) = 37.23$, $p < 0.001$). These differences in emotions assigned to the same pair of grades with differing goals show that reference points systematically influenced emotional responses.

Results were consistent with loss aversion, as shown in Figure 3, which plots ratings for the three pairs of grades with both set of goals. The figure also shows ratings for each pair of outcomes pooled across goals. Participants were happiest when they achieved their goals, less happy with the small loss and gain, and least happy with a large loss and gain (omnibus: $F(2, 168) = 89.17$, $p < 0.001$; goal versus small: $F(1, 168) = 80.93$, $p < 0.001$; goal versus large: $F(1, 168) = 178.07$, $p < 0.001$; small versus large: $F(1, 168) = 30.95$, $p < 0.001$).

Averages can mask individual subject data (see Hutchinson et al. 2000), but that was not true here. The majority of participants were happier with goal achievement than with a small gain and loss (127/169, 75%; $\chi^2(1) = 42.75$, $p < 0.001$), happier with goal achievement than with a large gain and large loss (140/169, 83%; $\chi^2(1) = 72.91$, $p < 0.001$), and happier with a small gain and loss than with a large gain and large loss (113/169, 67%; $\chi^2(1) = 19.22$, $p < 0.001$).

Emotional responses to small equivalent gains and losses were slightly above the neutral point on the response scale. Participants may have viewed this pair of grades as a "compromise," which was preferred over the two more extreme outcomes (i.e., Simonson 1989). Alternatively, if participants thought that one of

Figure 3. Study 1a: Loss Aversion in Emotional Ratings for Multiple Goals



Notes. Error bars represent ±1 SE. Results followed predictions for loss aversion: despite each grade pair providing an average score of 75, participants were happiest when the grade pairs meant they achieved both goals, were less happy when they had a small loss despite an equal magnitude gain, and were the least happy when they had a large loss with an equal magnitude gain.

the goals was harder to achieve and they did not achieve it, they may have expected it. An expected loss is not as painful as a surprising loss of the same magnitude (Mellers et al. 1997). Finally, participants may have assigned different weights to their English and math classes. We address this possibility in Studies 2 and 3.[1]

Study 1b: A Simple Test of Additivity

In Study 1b, we test the additivity assumption. Is the overall happiness of two outcomes described as the sum of the pleasure of each separate outcome? Or does a gain have relatively little influence on happiness in the presence of a loss?

Participants were given a pair of grades in English and math classes. They rated how they would feel about those grades given different pairs of goals. There were four goal pairs: (A) exceeding both goals (two gains), (B) exceeding one goal and falling short on the other (a gain and a loss), (C) falling short on one goal and exceeding another (a loss and a gain), and (D) falling short on both goals (two losses). If additivity holds, feelings of happiness generated by a grade in one class should not depend on the grade in the other class. Panels (a), (b), and (c) of Figure 2 show feelings about the same pair of grades with different goals (A, B, C, and D). If additivity holds, the difference between happiness ratings when goals are A and C (gain and gain versus loss and gain) should be the same as the difference between happiness ratings when goals were B and D (gain and loss versus loss and loss) (see Figure 2, panel (a)).

If happiness ratings were consistent with a minimum model, a loss and a gain should be more similar to two losses than two gains. The difference in happiness ratings for two identical grades should be larger with goals A and C (a change from a gain to a loss in the presence of a gain) than with goals B and D (a change from a gain to a loss in the presence of a loss; see Figure 2, panel (b)).

If ratings are consistent with a maximum model, the difference in happiness ratings for two identical grades should be smaller with goals A and C (a change from a gain to a loss in the presence of a gain) than with goals B and D (a change from a gain to a loss in the presence of a loss). With the maximum model, participants would tend to feel happy despite the loss (see Figure 2, panel (c)).

Method

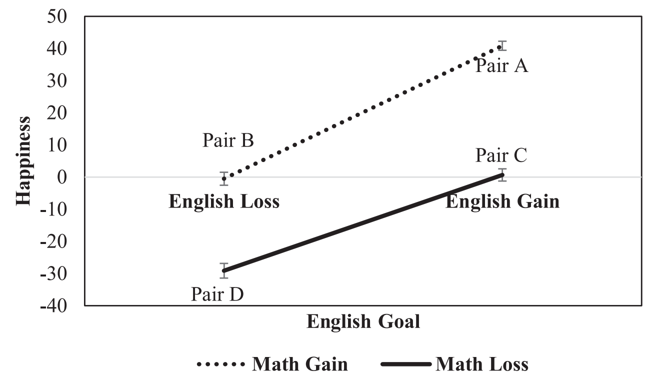
Ninety-four participants from Amazon Mechanical Turk completed this study in exchange for $0.35. Participants read about a woman named Alice who received grades of 60 in English and math. She had one of four sets of goals presented across four different pages in a randomized order. These goals were (A) 30 in English and 30 in math (two gains), (B) 90 in English and 30 in math (one gain and one loss), (C) 30 in English and 90 in math (one gain and one loss), and (D) 90 in English and 90 in math (two losses). Participants rated how happy Alice would feel about the scores of 60 in English and math relative to the goals using the same response scale as in Study 1a.

Results and Discussion

Figure 4 shows average feelings about grades of 60 in both classes across different goals. Ratings are plotted against English goals, with separate curves for math goals. Ratings display a divergent interaction, consistent with a minimum model. A single loss makes a gain less pleasurable. The interaction between goals was statistically significant (M = 11.57; $F(1, 93) = 17.58$, $p < 0.001$), which represents a larger gap between goals A and C than goals B and D. Participants indicated a sharper decline in happiness when moving from no losses (M = 40.86, SE = 1.41, Mdn = 48.5) to one loss (English loss: M = –0.51, SE = 2.02, Mdn = 0; math loss: M = 0.67, SE = 1.90, Mdn = 0) than they did when moving from one loss to two losses (M = –29.13, SE = 2.28, Mdn = –38). For the majority of participants, the data showed divergent interactions: 63% indicated a larger gap in happiness between two gains and one gain and one loss compared with the gap between one gain and one loss versus two losses.

A divergent interaction is a violation of additivity in line with a minimum model and several previous findings from the literature. Birnbaum et al. (1992) found divergent interactions for buyers' values of gambles varying across two rewards and their probabilities: a gamble with one low outcome and one high outcome is not valued much more than a gamble with two low outcomes. Birnbaum and Stegner (1979) also found that buyers' estimates followed a divergent pattern.

Figure 4. Study 1b: Violations of Additivity Consistent With a Minimum Model



Notes. Error bars represent ±1 SE. Results were consistent with a minimum model: there was a larger gap in happiness between achieving two gains and having one gain and one loss than between having one gain and one loss and two losses.

The divergent interaction is consistent with the notion that greater focus is placed on the worst outcome. We know from the goal literature that incomplete goals may persist and retain focus (Osviankina 1928, Masicampo and Baumeister 2011). Unattained goals may have a similar effect. The divergent interaction has not been discussed in riskless choice (Tversky and Kahneman 1991). Loss aversion on each variable alone may understate the displeasure from the mere presence of falling short of one or two goals.

Study 1b does not rule out the possibility of differential weighting of English and math classes. However, differential weighting does not predict divergence. We address models with differential weighting of classes in Study 2. Furthermore, similar to Study 1a, the one gain/one loss combination is also not rated as extremely painful; this may again be due to the goal of scoring 90 being unexpectedly difficult (Mellers et al. 1997).

Study 2: Parametric Tests of Loss Aversion, Diminishing Sensitivity, and Additivity

Here, we fit and compare models with our data at both the aggregate and the individual levels. The first is a baseline model (Equation (6)), with estimated parameters for only the response transformation and error (a, b, and σ). The parameters λ, α, and ω are fixed to 1, 1, and 0, respectively. We then examine the effects of adding each parameter (λ, α, and ω) to the baseline model. Finally, we explore full models: one with five parameters (a, b, σ, λ, and α) with no interaction (the additive prospect theory model) and two with six parameters (a, b, σ, λ, α, and ω) representing a minimum model and a maximum model.

Method

Two hundred and forty-two participants from Amazon Mechanical Turk completed Study 2 in exchange for \$1.00. Outcome pairs differed between subjects, and goal pairs were manipulated within subjects. Participants again read about Alice, who scored either 50 in English and 70 in math, 60 in both English and math, or 70 in English and 50 in math. Within-subject trials were constructed from a factorial design of English goals (30, 40, 50, 60, 70, 80, and 90) by math goals (30, 40, 50, 60, 70, 80, and 90). Participants rated Alice's happiness on the same scale as used in Study 1a and 1b. Alice was either at, above, or below her goals across the 49 trials.

This design provided tests of diminishing sensitivity, loss aversion, and additivity. It was also closer to the studies of Heath et al. (1999) in which the outcome remained constant and the participant saw different goals. An example trial is depicted in Figure 5. Underlined numbers in bold represent goals that vary across trials. The second line shows the outcome for each class that remained constant across trials. The third line refers to the happiness rating.

Results

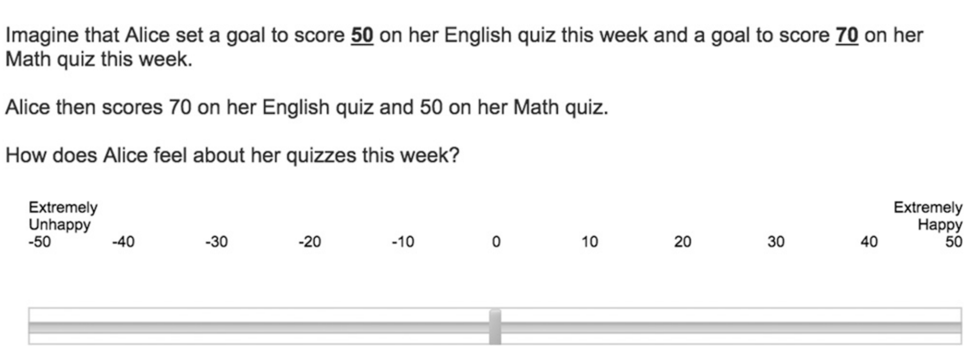
Happiness Ratings. Figure 6 shows the effects of goals on judged happiness for the same outcome pair within panels and different outcome pairs across panels. The horizontal axis depicts larger math goals on the left (where participants may be in the loss domain) and smaller math goals on the right (where participants may be in the gain domain). Dotted lines show gains on English goals, dashed lines show achievement on English goals, and solid lines show losses on English goals. Each panel in Figure 6 shows a kink where the math goal is achieved (70, 60, and 50 in the left, center, and right panels, respectively). Figure 7 inverts this mapping (i.e., English on the x axis and curves representing math goals) and shows similar patterns when the English goal is achieved. Slopes of the curves are steeper when goals are missed (solid lines below the kink in each panel) than when goals are met or exceeded (dashed and dotted lines above the kink in each panel).

Each panel shows feelings about two gains (top right portion), one gain and one loss (middle portions), and two losses (bottom left portion). Lines are not parallel as predicted by the additivity assumption. Interactions between English and math goals in the 7 (English goals) × 7 (math goals) ANOVA are significant in all three panels of Figure 6 ($F(36, 72) = 6.34$, $p < 0.001$; $F(36, 76) = 9.24$, $p < 0.001$; and $F(36, 91) = 5.68$, $p < 0.001$). People were happiest when they achieved or exceeded both goals. There was a large drop in happiness when they fell short on one goal and a relatively smaller drop when they missed both goals (compared with missing only one). Feelings about progress on one goal varied with feelings about progress on the other. Contrary to a maximum model, achieving or exceeding a goal did not make up for failing to achieve another one (i.e., the two losses do not generate additional displeasure).

Models. Models were fit to aggregate results by finding parameters that maximized log likelihood on the aggregate ratings data. Outcomes were represented as distances from goals (e.g., outcomes x_1 and x_2 under the influence of goals g_1 and g_2 were represented as $x_1 - g_1$ and $x_2 - g_2$). We used a similar technique to fit individual-level data.

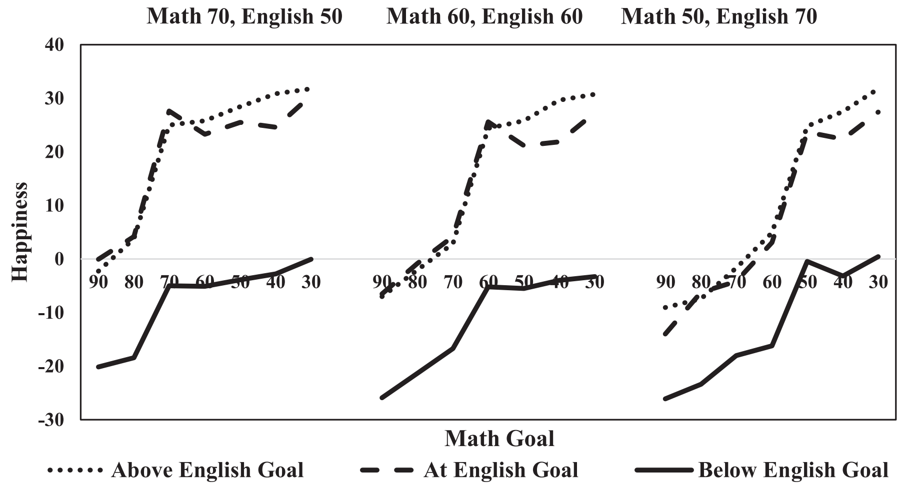
Table 1 shows best-fitting parameters and log likelihoods to the seven models for aggregate data. To compare the models while permitting penalties for model flexibility, we also computed the Akaike information criterion for each model, formalized as AIC = $2k - 2\text{LL}$, where k is the number of parameters in the model and LL is the best-fit log

Figure 5. Study 2: Screenshot for the Task



Notes. Numbers in bold and underlined represent goals manipulated within subject that vary on each of the 49 judgments. The outcomes on the second line are constant for each outcome condition but are manipulated between subjects.

Figure 6. Study 2: Emotion Ratings for Grades Showing Loss Aversion and Nonadditivity



Notes. Emotion ratings show outcome pairs of 50 in English and 70 in math on the left, 60 in English and 60 in math in the center, and 70 in English and 50 in math on the right. In each panel, judgments are made with respect to factorial combinations of English and math goals (both are 90, 80, 70, 60, 50, 40, and 30). Math goals are on the x axis, and curves represent English scores relative to English goals: dotted lines represent exceeding the English goal, dashed lines represent hitting the English goal, and solid lines represent falling short of the English goal.

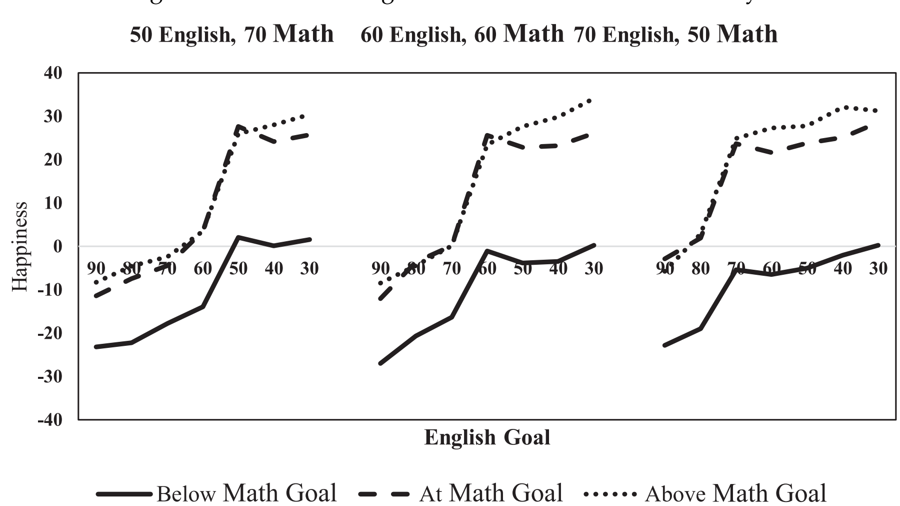
likelihood. For the models in Table 1, we focus on the Bayesian information criterion, formalized as BIC = $\log(n) \times k - 2\text{LL}$, where n is the number of observations.

When we assume linear and symmetric values and additivity of goal variables (i.e., baseline model), the BIC was 521 (AIC = 514, LL = −254). Now we investigate the extent to which loss aversion, diminishing sensitivity, and nonadditivity increase model fit relative to this baseline model.

The model that allowed nonlinear values supported the assumption of diminishing sensitivity ($\alpha < 1$). The estimated value of α was 0.52, with a BIC of 506 (AIC = 497, LL = −245). Diminishing sensitivity increased the fit of happiness ratings. The model with linear and asymmetric values supported loss aversion ($\lambda > 1$). The estimated value of λ was 2.81, and the BIC was 488 (AIC = 479, LL = −236). Finally, the model that permitted nonadditivity of outcome pairs was consistent with a minimum process ($\omega > 0$ when one or both failures occurred). The value of ω was 57.09, and the BIC was 423 (AIC = 413, LL = −203). This value indicated that a loss and a gain were closer to two losses than to two gains.

We therefore observe evidence of loss aversion, diminishing sensitivity, and nonadditivity when adding λ, α, and ω to the baseline model, respectively. These

Figure 7. Study 2: Emotion Ratings for Grades Showing Loss Aversion and Nonadditivity



Notes. Emotion ratings show outcome pairs of 50 in English and 70 in math on the left, 60 in English and 60 in math in the center, and 70 in English and 50 in math on the right. In each panel, judgments are made with respect to factorial combinations of English and math goals (both are 90, 80, 70, 60, 50, 40, and 30). English goals are on the x axis, and curves represent math goals: dotted lines represent exceeding the math goal, dashed lines represent hitting the math goal, and solid lines represent falling short of the math goal.

Table 1. Study 2: Models of Emotional Ratings for Aggregate-Level Results

Model	λ	α	a	b	ω	σ	k	LL	D^{h}	AIC	BIC
Baseline	1	1	2.55	0.55	0	7.17	3	−254.16		514.33	521.28
Test of diminishing sensitivity (α)	1	0.52	2.55	2.74	0	6.31	4	−244.57	19.18*	497.15	506.42
Test of loss aversion (λ)	2.81	1	13.47	0.29	0	5.59	4	−235.56	37.21*	479.12	488.39
Test of nonadditivity (ω): Maximum[a]	1	1	1.84	0.54	1.60	7.17	4	−254.12	0.09	516.24	525.51
Test of nonadditivity (ω): Minimum[b]	1	1	−3.19	0.34	57.09	3.61	4	−202.73	102.88*	413.45	422.72
No-LA[c]	1	0.82	−2.87	0.65	28.51	3.54	5	−201.29	80.18*	412.58	424.17
No-DS[d]	2.35	1	4.00	0.22	74.56	2.67	5	−180.19	148.72*	370.37	381.96
Additive prospect theory[e]	5.67	0.45	19.74	1.03	0	3.68	5	−204.18	85.97*	418.37	429.96
Maximum[f]	5.57	0.45	19.64	1.05	6.83×10^{-6}	3.68	6	−204.19		420.37	434.28
Minimum[g]	3.46	0.62	8.90	0.66	19.13	2.08	6	−161.20		334.39	348.30

Note. Let AIC = $2k - 2LL$, where k is the number of parameters in the model, and LL is the best-fit log likelihood; let BIC = $\log(n) \times k - 2LL$, where n is the number of observations, k is number of parameters in the model, and LL is the best-fit log likelihood.

[a]The test of nonadditivity with a maximum model has an interaction term (ω) associated with Equation (5).

[b]The test of nonadditivity with a minimum model has an interaction term (ω) associated with Equation (4).

[c]The no-LA model includes parameters to test diminishing sensitivity (α) and an interaction term for nonadditivity (ω) associated with Equation (4); it does not include a parameter for loss aversion.

[d]The no-DS model includes parameters to test loss aversion (λ) and an interaction term (ω) associated with Equation (4); it does not include a parameter for diminishing sensitivity.

[e]The additive prospect theory model includes parameters to test loss aversion (λ) and diminishing sensitivity (α) but has no interaction term for nonadditivity.

[f]The maximum model includes parameters to test loss aversion (λ), diminishing sensitivity (α), and an interaction term for nonadditivity (ω) associated with Equation (5).

[g]The minimum model includes parameters to test loss aversion (λ), diminishing sensitivity (α), and an interaction term for nonadditivity (ω) associated with Equation (4).

[h]The likelihood ratio test (D) is computed relative to the baseline, three-parameter model for the four-parameter models; D is computed relative to the six-parameter maximum model for the five-parameter models.

*$p < 0.001$.

models have lower BIC values than the baseline model. Adding loss aversion ($p < 0.001$), diminishing sensitivity ($p < 0.001$), and nonadditivity ($p < 0.001$) bolstered the model fits. Relaxing the additivity assumption provided the single greatest improvement in BIC relative to loss aversion and diminishing sensitivity.

Next, we fit full models: one with both λ and α without an interaction and two six-parameter models that allowed interactions according to the minimum or maximum models. Consistent with the four-parameter models, Study 1b, and the means in Study 2, the minimum model (BIC = 348, AIC = 334, LL = −161) outperformed the maximum (BIC = 434, AIC = 420, LL = −204) and additive prospect theory models (BIC = 430, AIC = 418, LL = −204) with the lowest BIC value. The difference in fits between the additive prospect theory model and the minimum model shows the need for nonadditivity ($p < 0.001$). In this case, the value of α was 0.62 (less than 1, consistent with diminishing sensitivity), the value of λ was 3.46 (greater than 1, consistent with loss aversion), and the value of ω for the minimum model was 19.13 (consistent with greater impact of a loss than a gain in the integration of feelings). This minimum model had an adjusted R^2 of 0.99, which was greater than that of the baseline (0.85), the maximum model (0.96), and additive prospect theory model (0.96).

We also compared a five-parameter minimum model without loss aversion and diminishing sensitivity to the six-parameter full minimum model. Allowing α to vary for diminishing sensitivity from the five-parameter model improved the BIC from 382 to 348 (AIC = 370, LL = −180; $p < 0.001$). Allowing λ to vary improved the BIC from 424 to 348 (AIC = 413, LL = −201; $p < 0.001$). These models demonstrate improved model fit by including parameters for loss aversion and diminishing sensitivity.

Individual-level data, reported in the Web appendix, exhibited results that were similar to aggregate-level data. The full minimum model supported a better fit than the full maximum model or the additive prospect theory model. Median parameter values for α, λ, and ω showed diminishing sensitivity, loss aversion, and nonadditivity. In short, our results hold at both the individual and aggregate levels.

In the Web appendix, we discuss nine additional models, including a power function with a jump, configural-weight models, differential-weight models, and models where the overall emotion is dictated solely by the worst outcome (minimum) or the best outcome (maximum). Although some of the models are not outperformed by those in Table 1, they nevertheless support violations of additivity. For example, the configural-weight model has a negative interaction parameter, consistent with our minimum model.

Other models also provide theory-consistent values for α and λ, indicating that support for loss aversion and diminishing sensitivity is not due to model misspecification.

Discussion

Study 2 supports loss aversion and diminishing sensitivity in gains and losses for multiple goals as reference points. However, additivity of feelings about goal progress was not supported. The emotional reaction to one variable was not independent of the other. Results were consistent with a minimum model; one failure spoiled (at least to some degree) pleasure from the other outcome.

Study 3: Multiple Goals in a New Context

Study 3 uses an experimental design similar to that in Study 2 with different goal trade-offs. Participants made 49 happiness judgments about outcomes on two exercise goals for push-ups and lunges.

Method

Two hundred and forty participants on Amazon Mechanical Turk completed this study in exchange for $1.00. Outcomes were manipulated between subjects, whereas goals were varied within subject. Participants were assigned to one of three outcome pairs achieved by Charles who was exercising at the gym: 75 push-ups and 45 lunges, 60 push-ups and lunges, and 45 push-ups and 75 lunges. Trials were constructed from a factorial design of push-up goals (15, 30, 45, 60, 75, 90, and 105) by lunge goals (15, 30, 45, 60, 75, 90, and 105). Participants used the same scale as before to rate the happiness of Charles, who achieved different combinations of outcomes relative to the goals. Figure 8 shows an example trial from this task.

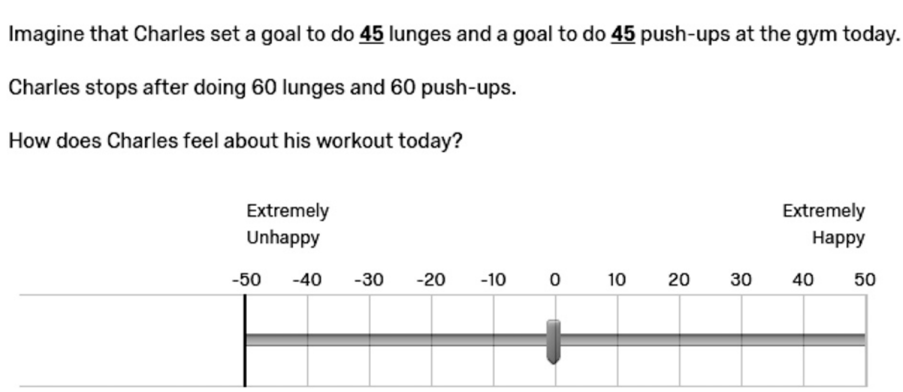
Results

Happiness Ratings. Each panel of Figure 9 shows the effects of goal pairs on the happiness associated with the same outcome pairs. The horizontal axis shows the push-up goal from larger to smaller. Each panel shows a kink in each curve when the push-up goal is achieved (75, 60, and 45 push-ups in the left, center, and right panels, respectively). This effect at goal attainment was, similar to Study 2, starker than movements along other parts of the graph in each panel. Solid lines represent losses on the lunges goal, dashed lines show goal achievement, and the dotted lines show gains (outcomes above the lunges goal). Figure 10 inverts this mapping (i.e., the lines represent the push-up goal, and the horizontal axis depicts progress along the lunge goal) and shows similar patterns. As in Study 2, the lines for being above the goal (gains) and the lines for being below (losses) are averaged over relevant combinations of goals and outcomes.

Slopes below the goals were steeper than those above the goals, as expected with loss aversion. The larger impact of losses compared with gains was also evident. Loss curves (solid lines) are lower than goal achievement (dashed lines) or gains (dotted lines). Second, the interaction between goal progress on one variable and goal progress on another variable reveals divergent interactions. Achieving one goal does not fully ameliorate failure on another, supporting a minimum model. The omnibus interaction in the 7 (push-ups goal) × 7 (lunges goals) ANOVA was significant in each of the panels ($F(36, 87) = 6.29$, $p < 0.001$; $F(36, 86) = 3.73$, $p < 0.001$; and $F(36, 64) = 12.29$, $p < 0.001$).

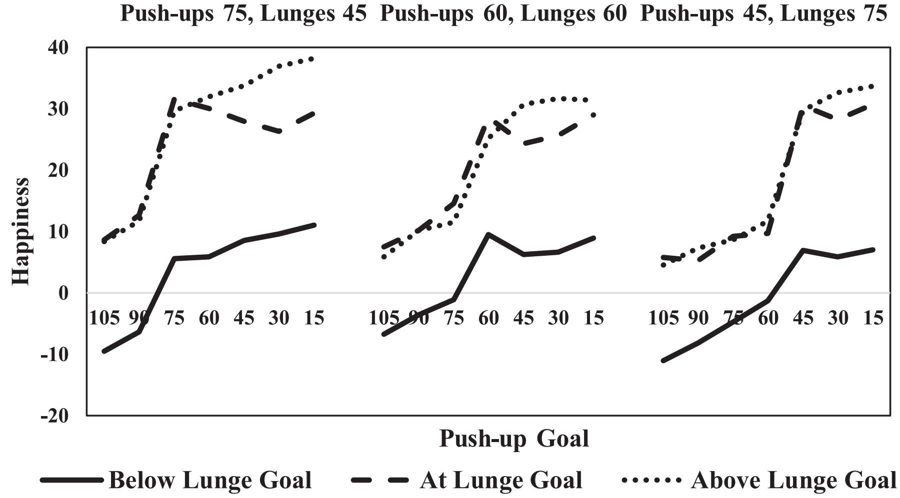
Model Fits. Table 2 shows the log likelihoods, AIC, BIC, and best-fitting parameters for the seven models. The baseline model fit to the aggregate data had a BIC of 479 (AIC = 472, LL = –233). When models allowed the addition of parameters to account for diminishing sensitivity (BIC = 460, AIC = 450, LL = –216; $p < 0.001$),

Figure 8. Study 3: Screenshot for the Task



Notes. Numbers in bold and underlined represent goals manipulated within subject that vary on each of the 49 judgments. The outcomes on the second line are constant for each outcome condition but are manipulated between subjects.

Figure 9. Study 3: Emotional Ratings for Exercises Showing Loss Aversion and Nonadditivity



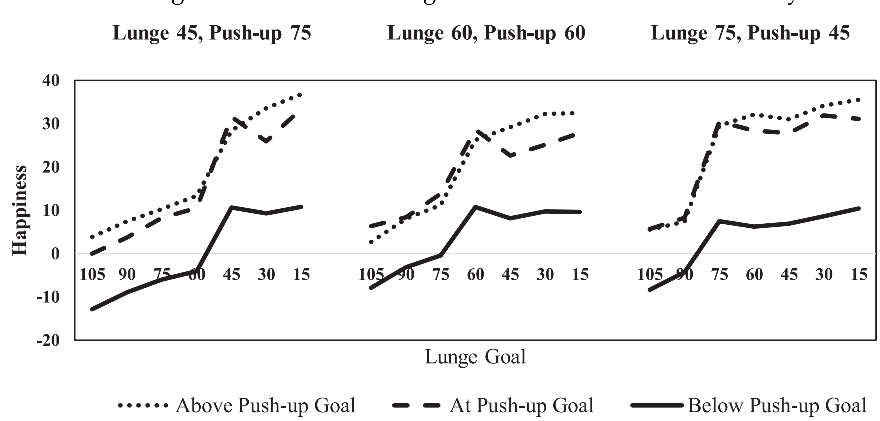
Notes. Mean emotion ratings of outcome pairs with 75 push-ups and 45 lunges on the left, 60 push-ups and 60 lunges in the center, and 45 push-ups and 75 lunges on the right. In each panel, judgments represent factorial combinations of lunge and push-up goals (both were 105, 90, 75, 60, 45, 30, and 15). Numbers of push-ups are on the abscissa, and curves are outcomes relative to lunge goals (dotted for above lunge goal, dashed for at lunge goal, and solid for below lunge goal).

loss aversion (BIC = 450, AIC = 441, LL = −216; $p < 0.001$), and nonadditivity (BIC = 400, AIC = 391, LL = −192; $p < 0.001$), fits were improved, with much lower BIC values compared with the baseline model. Values were 0.47, 2.65, and 77.73 for α, λ, and ω, respectively, all of which were consistent with the diminishing sensitivity, loss aversion, and nonadditivity results from Study 2. Moreover, improvement was once again greatest with relaxation of the additivity assumption.

The six-parameter minimum model with diminishing sensitivity, loss aversion, and nonadditivity had an adjusted R^2 value of 0.98 and BIC value of 334 (AIC = 320, LL = −154), outperforming the additive prospect theory (BIC 380, AIC = 368, LL = −179; $p < 0.001$, adjusted $R^2 = 0.96$) and maximum (BIC 384, AIC = 370, LL = −179; adjusted $R^2 = 0.96$) models. These R^2 values were greater than 0.85, the value for the baseline model. Although all models fit reasonably well, the six-parameter minimum model had the best fit with the lowest BIC value. Parameter values for the minimum model were 0.52, 3.72, and 13.40 for α, λ, and ω, respectively, consistent with diminishing sensitivity and loss aversion. People felt displeasure with a loss on any goal, and the impact of the second loss was much smaller than that of the first loss.

As in Study 2, the full minimum model fit substantially better than did five-parameter models with

Figure 10. Study 3: Emotional Ratings for Exercises Showing Loss Aversion and Nonadditivity



Notes. Mean emotion ratings of outcome pairs with 75 push-ups and 45 lunges on the left, 60 push-ups and 60 lunges in the center, and 45 push-ups and 75 lunges on the right. In each panel, judgments represent factorial combinations of lunge and push-up outcomes (both were 105, 90, 75, 60, 45, 30, and 15). Numbers of lunges are on the abscissa, and curves are outcomes relative to push-up goals (dotted for above push-up goal, dashed for at push-up goal, and solid for below push-up goal).

Table 2. Study 3: Emotion Models for Aggregate-Level Results

Model	λ	α	a	b	ω	σ	k	LL	D[h]	AIC	BIC
Baseline	1	1	12.24	0.28	0	5.42	3	−233.15		472.31	479.26
Test of diminished sensitivity (α)	1	0.47	12.24	2.02	0	4.62	4	−221.18	23.94***	450.37	459.64
Test of loss aversion (λ)	2.65	1	20.15	0.15	0	4.34	4	−216.45	33.41*	440.89	450.17
Test of nonadditivity (ω): Maximum[a]	1	1	10.18	0.26	9.43	5.37	4	−232.48	1.34	472.96	482.23
Test of nonadditivity (ω): Minimum[b]	1	1	8.12	0.18	77.73	3.11	4	−191.49	83.32*	390.98	400.25
No-LA[c]	1	0.71	8.55	0.57	22.03	2.98	5	−188.25	68.11*	386.49	398.08
No-DS[d]	2.20	1	13.39	0.12	97.34	2.55	5	−176.64	44.90*	363.29	374.87
Additive prospect theory[e]	5.49	0.41	25.25	0.79	0	2.64	5	−179.21	50.03*	368.42	380.01
Maximum[f]	5.67	0.41	25.40	0.77	3.08×10^{-9}	2.65	6	−179.22		370.44	384.35
Minimum[g]	3.72	0.52	18.62	0.57	13.40	1.89	6	−154.19		320.39	334.29

Note. Let AIC = $2k - 2LL$, where k is the number of parameters in the model, and LL is the best-fit log likelihood; let BIC = $\log(n)\times k - 2LL$, where n is the number of observations, k is number of parameters in the model, and LL is the best-fit log likelihood.

[a]The test of nonadditivity with a maximum model has an interaction term (ω) associated with Equation (5).

[b]The test of nonadditivity with a minimum model has an interaction term (ω) associated with Equation (4).

[c]The no-LA model includes parameters to test diminishing sensitivity (α) and an interaction term for nonadditivity (ω) associated with Equation (4); it does not include a parameter for loss aversion.

[d]The no-DS model includes parameters to test loss aversion (λ) and an interaction term (ω) associated with Equation (4); it does not include a parameter for diminishing sensitivity.

[e]The additive prospect theory model includes parameters to test loss aversion (λ) and diminishing sensitivity (α) but has no interaction term to test nonadditivity.

[f]The maximum model includes parameters to test loss aversion (λ), diminishing sensitivity (α), and an interaction term for nonadditivity (ω) associated with Equation (5).

[g]The minimum model includes parameters to test loss aversion (λ), diminishing sensitivity (α), and an interaction term for nonadditivity (ω) associated with Equation (4).

[h]The likelihood ratio test (D) is computed relative to the baseline, three-parameter model for the four-parameter models; D is computed relative to the six-parameter maximum model for the five-parameter models.

*$p < 0.001$.

fixed values of α and λ (and a minimum model interaction). Allowing α to vary in the full minimum model (from the five-parameter model with it fixed at 1; BIC = 375. AIC = 363, LL = −177) improved fit compared with the full minimum model (BIC 334, AIC = 320, LL = −154; $p < 0.001$). The same was true for the five-parameter model with λ fixed at 1 (BIC = 398, AIC = 386, LL = −188) compared with the full minimum model ($p < 0.001$). These tests provide consistent evidence for improvements via inclusion of parameters for loss aversion and diminishing sensitivity. Consistent with Study 2, the individual-level results were similar to aggregate-level results. The full minimum model also fit better than the maximum model or the additive prospect theory model. Additional models are explored in the Web appendix.

Discussion

Study 3 shows that features of the prospect theory value function appear with multiple goals. Loss aversion and diminishing sensitivity were supported at the aggregate and individual levels, but additivity was not. Individual-level fits from the emotions data showed that most participants exhibited a $\lambda > 1$ and $\alpha < 1$, and participants were well fit individually by the six-parameter minimum model. Using a scenario akin to the workout vignettes from Heath et al. (1999), we demonstrated the generalizability of the results.

General Discussion

For decades, researchers have recognized the importance and impact of reference points in judgment and decision making (Heath et al. 1999, Kőszegi and Rabin 2006, Allen et al. 2017). With laboratory and field data, researchers have accumulated systematic support for loss aversion and diminishing sensitivity across a wide array of contexts (e.g., Pope and Schweitzer 2011, Wang and Johnson 2012).

Goals can serve as reference points, and most research has focused on single goals along one variable (Heath et al. 1999). Relatively little work has been done to understand single goals on multiple variables. In these cases, goals may conflict, and trade-offs may be required (Dhar and Simonson 1999, Fishbach and Dhar 2005). We find evidence of loss aversion and diminishing sensitivity, but not additivity. Individual evaluations of outcomes relative to goals were not independent. The pleasure of an achievement of one goal did not make up for the displeasure of falling short on another goal. Simple studies (Studies 1a and 1b) and more complex designs with quantitative model fits (Studies 2 and 3) show the effects with aggregate and individual data. These individual-level analyses suggest that our results are not due to unobserved heterogeneity or cases in which averages mask individual patterns (Hutchinson et al. 2000).

We find evidence of a sharp increase in happiness at goal attainment. This increase is consistent with the study by Markle et al. (2018) of marathon runners in which the best description of runner satisfaction is the prospect theory value function with a jump at 0 (i.e., goal attainment). Our minimum model and the Markle et al. model suggest that having single goals as reference points may require more than the prospect theory value function.

The present results have implications for multiattribute riskless decision making (Tversky and Kahneman 1991, Kőszegi and Rabin 2006). The nonadditivity of emotions should send warnings to others about additivity assumptions in riskless choice. The interactions were similar to those found by Birnbaum and his colleagues (e.g., Birnbaum and Stegner 1979, Birnbaum et al. 1992, Birnbaum 2008). The divergent interactions in Studies 1b, 2, and 3 suggest that people place greater attention on the worst outcome of the experience. Future research could manipulate these interactions as done by Birnbaum et al. (1992) and Birnbaum and Stegner (1979).

The divergent interactions in our data indicate a greater impact of loss aversion beyond the value function. The mere presence of losses on either variable contributes to overall unhappiness beyond the claim that "losses loom larger" (Kahneman and Tversky 1979, p. 279). Previous work on riskless reference-dependence models has examined choices, not happiness, and has not tested the additivity assumption (Tversky and Kahneman 1991).

Our results have managerial implications for employee satisfaction. Although there is some evidence that multiple goals yield heightened performance (Locke and Latham 1990), the present results suggest that multiple goals can decrease the pleasure of a given success (see also King and Burton 2003, Ordóñez and Wu 2014). If a loss on any goal diminishes the pleasure of a gain on another, employees might be better off attending to each goal separately or reframe the goal along one dimension. Sequential and focused goal achievements might better motivate employees. That is, employee morale might increase by actively segregating different goal outcomes (e.g., staggered deadlines). This idea is related to mental accounting recommendations about when to segregate or integrate losses and gains (Thaler 1985, 1999). In that literature, cancelation of losses against larger gains (mixed gains) is recommended (i.e., integration). Segregating a small gain against a larger loss is also recommended (Thaler 1985, Thaler and Johnson 1990). Yet if multiattribute outcomes on different variables are not cancelled in the same way as a large monetary gain and a small monetary loss, the large gain might instead be contaminated by the large loss.

There are other avenues for future investigations. We used hypothetical scenarios to test theoretical models. But real goals should be further examined (e.g., Markle et al. 2018). Research should also be done on multiple goals that are self-generated (Fulford et al. 2010, Markle et al. 2018). Self-set goals may also produce results consistent with the minimum model if individuals are similarly displeased by a loss for which they are more personally invested. Alternatively, self-set goals may show sharper effects resulting from higher personal relevance that are more consistent with the Markle et al. (2018) model. Finally, the multiple goals that we examined were on similar variables (i.e., two classes). What happens when the variables are vastly different (e.g., a financial goal and an exercise goal)? Do the same models hold? It is possible that different functional forms would be necessary when integrating across less alignable variables. Studies that wish to investigate this variation must control for vastly different units across variables, across which it may be complicated for participants to integrate.

Furthermore, one larger question concerns the extent to which the present results—specifically, with the minimum model—generalize to scenarios with gains and losses more broadly. It is possible that performance could conform to the maximum model. For example, an athlete who succeeds at a sporting competition in one specialized event may sacrifice performance in another niche event, but that athlete may feel satisfaction consistent with the maximum model. It may be better to be the expert in one area than to be only moderately successful in multiple areas.

Conclusions

Researchers have found that features of prospect theory's value function apply to many behaviors in many contexts. These results help explain findings in the goal-setting literature (Heath et al. 1999, Markle et al. 2018). Our results extend previous results about a single goal on one variable to single goals on multiple variables. The emotional impact of progress on two goals is not an additive function of the progress made on each goal. Rather, success in one goal and failure in another are less pleasurable than the sum of the separate feelings associated with these outcomes would predict. Understanding emotional reactions to progress on multiple goals is critical to motivation. How we feel about the fruits of our efforts has an enormous effect on our desire to set and pursue even bigger goals in the future.

Acknowledgment

The authors thank George Wu and Daniel Bartels, among other attendees of the 2016 Society for Judgment and

Decision-Making conference for their generous comments and feedback.

Endnote

[1] Study 1a is conceptually similar to other studies such as in Ordóñez et al. (2000) or in Tversky and Kahneman (1991), in the latter of which subjects were given choices between jobs described by commute time and social interaction. However, unlike Tversky and Kahneman (1991), we examine emotions rather than choices. Furthermore, unlike those two papers, we examine diminishing sensitivity and additivity.

References

Allen EJ, Dechow PM, Pope DG, Wu G (2017) Reference-dependent preferences: Evidence from marathon-runners. *Management Sci.* 63(6):1657–1672.

Bhatia S (2018) Decision making in environments with non-independent dimensions. *J. Behav. Decision Making* 31(2): 294–308.

Birnbaum MH (1972) Morality judgments: Tests of an averaging model. *J. Experiment. Psych.* 93(1):35–42.

Birnbaum MH (1973) Morality judgments: Tests of an averaging model with differential weights. *J. Experiment. Psych.* 99(3): 395–399.

Birnbaum MH (1974) The nonadditivity of personality impressions. (Monograph.) *J. Experiment. Psych.* 102(3):543–561.

Birnbaum MH (2008) New paradoxes of risky decision making. *Psych. Rev.* 115(2):463–501.

Birnbaum MH, Stegner SE (1979) Source credibility in social judgment: Bias, expertise, and the judge's point of view. *J. Personality Soc. Psych.* 37(1):48–74.

Birnbaum MH, Coffey G, Mellers BA, Weiss R (1992) Utility measurement: Configural-weight theory and the judge's point of view. *J. Experiment. Psych. Human Perception Performance* 18(2): 337–346.

Brenner L, Rottenstreich Y, Sood S, Bilgin B (2007) On the psychology of loss aversion: Possession, valence, and reversals of the endowment effect. *J. Consumer Res.* 34(3):369–376.

Cherry B, Ordóñez LD, Gilliland SW (2003) Grade expectations: the effects of expectations on fairness and satisfaction perceptions. *J. Behav. Decision Making* 16(5):375–395.

Dawes RM (1964) Social selection based on multidimensional criteria. *J. Abnormal Soc. Psych.* 68(1):104–109.

Dhar R, Simonson I (1999) Making complementary choices in consumption episodes: Highlighting vs. balancing. *J. Marketing Res.* 36(1):29–44.

Einhorn HJ (1970) The use of nonlinear, noncompensatory models in decision making. *Psych. Bull.* 73(3):221–230.

Festinger L (1954) A theory of social comparison processes. *Human Relations* 7(2):117–140.

Finkelstein SR, Fishbach A (2010) When healthy food makes you hungry. *J. Consumer Res.* 37(3):357–367.

Fischer GW (1976) Multidimensional utility models for risky and riskless choice. *Organ. Behav. Human Decision Processes* 17(1): 127–146.

Fishbach A, Dhar R (2005) Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *J. Consumer Res.* 32(3):370–377.

Fishburn PC, Wakker PP (1995) The invention of the independence condition for preferences. *Management Sci.* 41(7):1130–1144.

Frey BS, Stutzer A (2002a) What can economists learn from happiness research? *J. Econom. Literature* 40(2):402–445.

Frey BS, Stutzer A (2002b) *Happiness and Economics: How the Economy and Institutions Affect Human Well-Being* (Princeton University Press, Princeton, NJ).

Fulford D, Johnson SL, Llabre MM, Carver CS (2010) Pushing and coasting in dynamic goal pursuit: Coasting is attenuated in bipolar disorder. *Psych Sci.* 21(7):1021–1027.

Heath C, Larrick RP, Wu G (1999) Goals as reference points. *Cognitive Psych.* 38(1):79–109.

Hutchinson JW, Kamakura WA, Lynch JG (2000) Unobserved heterogeneity as an alternative explanation for "reversal" effects in behavioral research. *J. Consumer Res.* 27(3):324–344.

Kahneman D (1992) Reference points, anchors, norms, and mixed feelings. *Organ. Behav. Human Decision Processes* 51(2):296–312.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):263–291.

Kahneman D, Knetsch J, Thaler R (1990) Experimental tests of the endowment effect and the Coase theorem. *J. Political Econom.* 98(6):1325–1348.

Keeney RL, Raiffa H (1976) *Decisions with Multiple Objectives* (Cambridge University Press, Cambridge, UK).

King LA, Burton CM (2003) The hazards of goal pursuit. Chang EC, Sanna LJ, eds. *Vice, Virtue, and Personality: The Complexity of Behavior* (American Psychological Association, Washington, DC), 53–69.

Kőszegi B, Rabin M (2006) A model of reference-dependent preferences. *Quart. J. Econom.* 121(4):1133–1165.

Locke EA, Latham G (1990) *A Theory of Goal Setting and Task Performance* (Prentice-Hall, Englewood Cliffs, NJ).

Louro MJ, Pieters R, Zeelenberg M (2007) Dynamics of multiple-goal pursuit. *J. Personality Soc. Psych.* 93(2):174–193.

Markle A, Wu G, White R, Sackett A (2018) Goals as reference points in marathon running: A novel test of reference-dependence. *J. Risk Uncertainty*, ePub ahead of print February 23, https://doi.org/10.1007/s11166-018-9271-9.

Masicampo EJ, Baumeister RF (2011) Consider it done! Plan making can eliminate cognitive effects of unfulfilled goals. *J. Personality Soc. Psych.* 101(4):667–683.

McGraw AP, Larsen JT, Kahneman D, Schkade D (2010) Comparing gains and losses. *Psych. Sci.* 21(10):1438–1445.

Mellers BA (2000) Choice and the relative pleasure of consequences. *Psych. Bull.* 126(6):910–924.

Mellers BA, Schwartz A, Ritov I (1999) Emotion-based choice. *J. Experiment. Psych. General* 128(3):332–345.

Mellers BA, Schwartz A, Ho K, Ritov I (1997) Decision affect theory: Emotional reactions to the outcomes of risky options. *Psych Sci.* 8(6):423–429.

Novemsky N, Kahneman D (2005) The boundaries of loss aversion. *J. Marketing Res.* 42(2):119–128.

Ordóñez LD, Wu G (2014) Goals and decision making. Highhouse S, Dalal RS, Salas E, eds. *Judgment and Decision Making* (Routledge, New York), 123–139.

Ordóñez LD, Connolly T, Coughlan R (2000) Multiple reference points in satisfaction and fairness assessment. *J. Behav. Decision Making* 13(3):329–344.

Osviankina M (1928) Die Wiederaufnahme unterbrochener Handlungen. *Psych. Forsch.* 11(1):302–379.

Pope D, Schweitzer M (2011) Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *Amer. Econom. Rev.* 101(1):129–157.

Riskey DR, Birnbaum MH (1974) Compensatory effects in moral judgment: Two rights don't make up for a wrong. *J. Experiment. Psych.* 103(1):171–173.

Samuelson W, Zeckhauser R (1988) Status quo bias in decision making. *J. Risk Uncertainty* 1(1):7–59.

Schmidt K-H, Kleinbeck U, Brockmann W (1984) Motivational control of motor performance by goal setting in a dual-task situation. *Psych. Res.* 46(1):129–141.

Simonson I (1989) Choice based on reasons: The case of attraction and compromise effects. *J. Consumer Res.* 16(2):158–174.

Sullivan K, Kida T (1995) The effect of multiple reference points and prior gains and losses on managers' risky decision making. *Organ. Behav. Human Decision Processes* 64(1):76–83.

Thaler R (1985) Mental accounting and consumer choice. *Marketing Sci.* 4(3):199–214.

Thaler R (1999) Mental accounting matters. *J. Behav. Decision Making* 12(3):183–206.

Thaler R, Johnson EJ (1990) Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Sci.* 36(6):643–660.

Tversky A, Kahneman D (1991) Loss aversion in riskless choice: A reference-dependent model. *Quart. J. Econom.* 106(4):1039–1061.

Wang XT, Johnson JG (2012) A tri-reference point theory of decision making under risk. *J. Experiment. Psych. General* 141(4): 743–756.

Weaver R, Frederick S (2012) A reference price theory of the endowment effect. *J. Marketing Res.* 49(5):696–707.

Wolfinger R, Chang M (1998) Comparing the SAS GLM and MIXED procedures for repeated measures. White paper, SAS Institute, Cary, NC.