Goals as Reference Points in Marathon Running: A Novel Test of Reference-Dependence

Alex B. Markle* George Wu† Rebecca White‡ Aaron M. Sackett§

November 12, 2014

Abstract

Although many empirical investigations have documented reference-dependent preferences, most studies of reference dependence have considered only status quo reference points. In a large-scale field study of marathon runners, we test whether goals, a non-status quo reference point, act similarly to status quo reference points. We find that satisfaction as a function of relative performance (the difference between a runner's time goal and her finishing time) exhibits loss aversion and diminishing sensitivity, consistent with the Prospect Theory value function. Unlike the Prospect Theory value function, however, we also find a discontinuity (or jump) at the reference point. We further find that loss aversion is moderated by goal importance, that multiple reference points simultaneously impact runner satisfaction, and that loss aversion is overestimated in predictions of satisfaction, but still present in actual experienced satisfaction.

1 Introduction

A recent article in *The Chronicle of Higher Education* documented the enormous scholarly impact of Prospect Theory, highlighting Prospect Theory's extensive reach beyond economics and psychology into areas as disparate as engineering, humanities, law, and medicine (Goldstein 2011). Perhaps the greatest contributor to Prospect Theory's widespread influence is its proposition that individuals evaluate outcomes in relative terms, as changes from a neutral reference point, rather than as final states as classical economic models assume (Kahneman and Tversky 1979, Tversky and Kahneman 1992). Indeed, reference dependence can account for a wide range of empirical phenomena, including the endowment effect (Kahneman et al. 1990), status quo bias (Samuelson and Zeckhauser 1988), disposition effect (Shefrin and Statman 1985), equity premium puzzle (Benartzi and Thaler 1995), and extreme risk aversion for small stakes prospects involving the possibility of a loss (Rabin 2000).

Empirical investigations into Prospect Theory, and reference dependence more generally, have largely taken the reference point to be the status quo. This choice is sensible in many domains, such as studies of the endowment effect, because the status quo is the most reasonable or even the only reasonable reference point an individual may use to evaluate a particular outcome. However, in other domains, points of comparison besides the status quo are also plausible reference points. Indeed, Tversky and Kahneman (1991) suggested that "aspirations, expectations, norms, and social comparisons" might also serve as

*Fordham University, Gabelli School of Business, 5 Columbus Circle, Room 11-18, New York, NY 10019 (amarkle@fordham.edu)

†University of Chicago, Booth School of Business, 5807 S. Woodlawn Avenue, Chicago, IL 60637 (wu@ChicagoBooth.edu)

‡University of Chicago, Booth School of Business, 5807 S. Woodlawn Avenue, Chicago, IL 60637 (rebecca.white@ChicagoBooth.edu)

§University of St. Thomas, Opus College of Business (sackett@stthomas.edu)

reference points. While these alternative reference points have been the subject of some empirical study, such investigations face conceptual and methodological challenges. Conceptually, the field currently lacks the principles to guide a researcher faced with identifying which of any number of plausible reference points an agent is using to evaluate an outcome. Empirically, even once identified reference points may be difficult to observe and measure. Expectations, for example, are likely to be subjective and thus hidden from the researcher.

In this paper, we investigate reference dependence in marathon running, a setting distinct from those traditionally used to study Prospect Theory, and one which can provide useful insights into the operation of non-status quo reference points. While people run marathons for a variety of reasons (e.g., to support a charity or compete with friends), achieving or surpassing a predetermined time goal defines success for many marathoners. Indeed, marathon training guides often emphasize the importance of goals as a motivational tool and provide heuristics and calculators for setting appropriate goals (e.g., Higdon 2011). In the present research, we test the specific proposition that time goals serve as reference points in the evaluation of marathon performance.[1] In support of our proposition, we find that runner satisfaction as a function of relative performance (the difference between a runner's finishing time and her time goal) exhibits loss aversion and diminishing sensitivity, mirroring the classic S-shape of the Prospect Theory value function.

These results are significant, both empirically and theoretically. Empirically, our results contribute to the list of field demonstrations of Prospect Theory (Camerer 2005, Camerer et al. 1997, Fryer et al. 2012, Genesove and Mayer 2001, Pope and Schweitzer 2011, Post et al. 2008; see Allen et al. 2014 for a more extensive list). This growing body of work suggests that reference dependence extends beyond small-stakes, stylized laboratory experiments, contrary to the criticism of List (2003). Moreover, we provide an estimation of the Prospect Theory value function using data collected outside of the laboratory. Although many papers have estimated the Prospect Theory value function from laboratory data (e.g., Abdellaoui 2000, Gächter et al. 2007, Ho and Zhang 2008, Wu and Markle 2008), very few have done so using field data (for exceptions, see Post et al. 2008, Tovar 2009).

Our novel setting also yields a number of conceptual insights, filling gaps in the existing literature and generalizing the scope of Prospect Theory in several ways. In particular, we demonstrate that a goal, a non-status quo reference point, operates in the same basic manner as a status quo reference point, corroborating the laboratory findings of Heath et al. (1999) in a field environment. Unlike the Prospect Theory value function, however, we also find a jump in satisfaction at the reference point, resulting in a discontinuous value function. Although Diecidue and Van de Ven (2008) axiomatized an expected utility model that incorporates a jump at a level of aspiration and suggested that such a discontinuity can be seen as an extreme form of loss aversion, empirical demonstrations of such a jump are limited. For marathon runners, aversion to losses thus appears to have two sources: a value function that is more steeply sloped in the loss than in the gain domain, as in the classical Prospect Theory account, and a discrete jump at the reference point itself. We further find that loss aversion in both forms is moderated by goal importance. Runners who view beating their time goal as more important exhibit a greater degree of loss aversion than those for whom surpassing their time goal is of lesser importance.

Our paper also examines the role of multiple reference points. In our context, a time goal is one of several possible referents from which a runner may evaluate her performance. Experienced marathoners, for instance, may compare their current performance against past performance (e.g. their most recent marathon time or best marathon time). Indeed, we find that satisfaction is driven by comparing performance against both their time goal and previous marathon time.

[1]The proposition that goals serve as reference points has broader implications. Heath et al. (1999) propose that goals also influence effort provision. For evidence of this, see Sackett et al. (2014a), Allen et al. (2014), and Gómez-Miñambres et al. (2013).

Finally, we employ satisfaction as our dependent measure rather than the more typical dependent measures of choice or pricing. This allows us to directly investigate the relationship between predicted and experienced utility (Kahneman 1999). In our study, we contrast pre-marathon predictions of satisfaction with post-race measures of satisfaction and find that while runners typically overestimate the magnitude of their reaction to falling short of their goal, consistent with research on immune neglect (Gilbert et al. 1998), loss aversion is still evident in their experienced satisfaction. In sum, this study contributes to multiple strands of research on reference dependence.

Our paper proceeds as follows. In Section 2, we review the existing literature and position this paper's contributions relative to that literature. In Section 3, we describe the methods for our empirical investigation. In Section 4, we present the basic results of our study. In Section 5, we estimate the Prospect Theory value function to relate satisfaction to relative performance. We conclude in Section 6, by summarizing our contributions and discussing some implications and future directions for this research.

2 Background

In this section, we review the literature on non-status quo reference points, as well as previous findings on the effects of multiple reference points on the valuation of outcomes, moderators of loss aversion, and differences between loss aversion in decision and experienced utility.

2.1 Non-Status Quo Reference Points

While most studies of reference dependence take the decision maker's reference point to be the status quo, an increasing number have examined alternative reference points. Some have utilized a lagged status quo as the reference point. In studies of the disposition effect, for example, an asset's purchase price serves as the reference point from which later prices are evaluated. Consistent with that interpretation, investors are significantly more likely to sell winning stocks than losing stocks (Barberis and Xiong 2009, Genesove and Mayer 2001, Odean 1998, Shefrin and Statman 1985). In a study of employee stock options, Heath et al. (1999) used a different lagged status quo, the yearly high price, as the reference point and documented a dramatic increase in the likelihood of exercising a stock option once a stock exceeded the yearly high.

In other cases, certain reference points are intuitively plausible. For example, Pope and Schweitzer (2011) found that professional golfers performed worse when putting for a birdie than for a par. In this case, the reference point is a convention known as "par," and thus a putt for a birdie (one stroke below par) is coded as a possible gain and a putt for a par is seen as avoiding a loss, which stimulates greater effort. In addition, Pope and Simonsohn (2011) proposed that round numbers serve as reference points and provided evidence from baseball and the Scholastic Aptitude Test that individuals exert enough effort to perform just above, rather than just below, a level of performance expressed as a round number. Similarly, Allen et al. (2014) and found that finishing times of marathon runners bunch just below round numbers such as 4 hours.

Other inquiries have relied upon expectations as reference points, with much of the empirical work building on Kőszegi and Rabin's (2006) reference point model of rational expectations. In a modified endowment effect study, Ericson and Fuster (2011) manipulated participants' expectations that they would have an opportunity to trade the item with which they had been endowed. Those facing a greater chance to trade demanded a higher price, consistent with the model of Kőszegi and Rabin. Abeler et al. (2011) also took an experimental approach, manipulating participants' expectations concerning the wage they would earn for performing a boring task, and documented a pattern of effort provision consistent with a model in which expectations serve as a reference point.

Goals, aspirations, and targets are natural candidates for reference points, and have been incorporated into a number of theoretical choice models (Diecidue and Van De Ven 2008, Lopes 1987, Lopes and Oden 1999, March and Shapira 1992). In the literature on motivation, setting a goal is a "discrepancy creating process" that produces "discontent with one's present condition and the desire to attain an object or outcome" (Locke and Latham 2006). Implicit in this formulation is that goals operate as reference points, providing a point of comparison from which to evaluate present performance. This notion of goals was made explicit by Heath et al. (1999), who proposed that both effort and satisfaction in goal pursuit tasks could be explained by the characteristics of the Prospect Theory value function.

In support of this proposition, Camerer et al. (1997) found that, contrary to the standard economic prediction, New York City taxi drivers work longer hours on slower days (when the average hourly wages are less), and finish work earlier on busier days (when hourly wages are higher). Drivers, the authors argue, establish daily income targets and throughout the day compare their accumulated earnings against these income targets. Earnings below the targets are judged to be losses, and thus drivers exert additional effort to meet their targets. Further support that income targets behave as reference points comes from Fehr and Goette (2007), Crawford and Meng (2011), Farber (2008), and Oettinger (1999), although Farber (2005) provides conflicting evidence.

There are challenges, both conceptual and methodological, to studying non-status quo reference points. While Heath et al. (1999) proposed that "whenever a specific point of comparison is psychologically salient, it will serve as a reference point," we have few guiding principles to help us discriminate between any number of salient candidate reference points. Indeed, Barberis (2013) suggested that identifying the appropriate reference point is one of the main challenges to applying Prospect Theory in field settings: "it is often unclear how to define precisely what a gain or loss is, not least because Kahneman and Tversky offered relatively little guidance on how the reference point is determined" (p. 178).

The difficulties extend beyond identification to measurement. While historical stock prices are readily observable and round numbers are round for everyone, expectations and aspirations are both heterogeneous and difficult to observe. One approach is to experimentally manipulate the reference point as in Ericson and Fuster (2011) or Abeler et al. (2011). The second is to infer the reference point from something that is observable. Crawford and Meng (2011), for example, proxy a taxi driver's earnings expectations by the average wage earned per day. In this study we take a third approach, directly eliciting reference points (in this case goals). This has the advantage of not having to rely on a proxy variable, but the disadvantage that because participants are not randomly assigned, inferred relationships may reflect systematic differences between participants who set difficult goals and those who set easier ones. We address this issue directly in Section 5.3.

2.2 Multiple Reference Points

A student with a test score of "B+" could compare that result against her typical grade (say, a "B-"), her expected grade for that test (say, an "A-"), or the performance of her friend (say, a "B"). Which comparison will drive her evaluation of her performance on the test? In some cases, one of these reference points may be distinctly more psychologically salient than the others. But what happens if this is not the case? In contrast to much of the existing literature on multiple reference points, which has assumed either that the reference point is some weighted average of a set of reference points (e.g., Briesch et al. 1997, Ordóñez 1998, Winer 1986), or that individuals direct their attention to only one reference point at a time (Lopes and Oden 1999, March and Shapira 1992), Kahneman (1992) proposed that each of the possible reference points could simultaneously influence the evaluation of an outcome. In other words, an outcome could produce genuine mixed feelings or ambivalence, being viewed simultaneously as a gain relative to one reference point and a loss relative to another (Larsen et al. 2004).

A handful of empirical studies have found support for Kahneman's simultaneous reference point account. Ordóñez et al. (2000) asked participants to rate their satisfaction with a hypothetical salary offer, displayed alongside offers to other qualified job candidates. They found lower satisfaction when the focal salary lay between two peer salaries than when the peers were offered the same salary as the subject, a pattern of satisfaction consistent with the two reference points operating separately. Sullivan and Kida (1995) found that managers evaluated the performance of an investment option by simultaneously comparing that investment's return to that of previous investments, as well as to the target level of return. Finally, Wang and Johnson (2012) developed a tri-reference point extension to Prospect Theory, including the status quo, a goal level, and a minimum requirement, and provided empirical support for that model.

Much like the student example above, a marathon runner might compare her performance against her time goal, her most recent marathon time, her personal best marathon time, or a number of other possible points of comparison. While we anticipate that comparisons between actual performance and time goal will be the primary driver of a runner's satisfaction, we also test whether these alternative comparisons play a complementary role.

2.3 Moderators of Loss Aversion

A frequent observation in the decision making literature is that losses are "approximately twice as painful as gains are pleasurable" (Hastie and Dawes 2010, p. 276). Indeed, many empirical studies of decision making under risk have estimated loss aversion coefficients of around two (e.g. Abdellaoui et al. 2007). This does not appear to be a universal constant, however, as a number of boundary conditions and moderators of loss aversion have been identified (Ariely et al. 2005, Camerer 2005, Novemsky and Kahneman 2005). Within the endowment effect literature these include characteristics of the endowed good such as its duration of ownership (Strahilevitz and Loewenstein 1998), whether it is a good intended for exchange (Novemsky and Kahneman 2005), whether it is hedonic or utilitarian (Dhar and Wertenbroch 2000), its desirability (Brenner et al. 2007), and its numerosity (Burson et al. 2013).[2]

Additionally, loss aversion may vary across the attributes of a single good. Tversky and Kahneman (1991) provided a theoretical model for evaluating objects with multiple attributes, where each attribute has its own value function. They proposed that the value function for each attribute may differ, with different degrees of loss aversion for each attribute. Specifically, they posited that differences in loss aversion across attributes reflect differential importance or prominence of those dimensions. Johnson et al. (2006) found support for differential loss aversion. In the context of hypothetical car purchases, they found that loss aversion coefficients varied significantly across vehicle dimensions such as safety, comfort, and fuel economy, and critically that the magnitude of loss aversion was correlated with the importance that the consumers ascribed to that attribute. Runners participate in marathons for many reasons, only one of which is to beat a predetermined time goal. In our study, we asked runners to rate a number of objectives in terms of importance. We test for an association between loss aversion and the importance that runners attach to meeting their time goal.

List (2003) and other scholars have expressed skepticism about the real-world relevance of reference dependence, arguing that experience, among other factors, is likely to reduce or eliminate loss aversion. In a study of trading behavior at a collectible sports card show, for example, List (2003) found that the endowment effect is smaller and, indeed, negligible for experienced card dealers relative to inexperienced non-dealers, and furthermore that inexperienced non-dealers demonstrated a diminished endowment

[2] Most studies that have looked at moderators of loss aversion have used the endowment effect paradigm, although both rejection of positive expected value mixed gambles and disparities between selling and buying price in the endowment effect paradigm are interpreted as evidence for loss aversion.

effect after one year.[3] In research on the disposition effect, the tendency to sell investments that have appreciated in price and the reluctance to sell those that have depreciated in price, Dhar and Zhu (2006) found that the effect is attenuated among those investors with greater financial literacy and greater experience, and Shapira and Venezia (2001) found a weakened disposition effect among professional, compared to amateur, investors. In a study of the disposition effect in housing markets, Genesove and Mayer (2001) found that investor-owners exhibited a diminished disposition effect relative to owner-occupiers. However, an experimental study by Haigh and List (2005) found greater myopic loss aversion among professional traders than among undergraduate subjects. Our participants vary considerably in their prior marathoning experience, and we use this natural variation to test whether loss aversion is attenuated by experience.

2.4 Loss Aversion as an Affective Forecast

"Utility," as most commonly used in economics and decision making research today, refers to the weight given to an outcome as inferred from observed choices. This is distinct from Bentham (1789), for whom utility referred to the actual affective experience (pain or pleasure) aroused by an outcome. Kahneman et al. (1997) termed the earlier notion of utility *experienced utility* and the later and more common conception *decision utility.* There are two possible reasons for a discrepancy between decision and experienced utility: (1) the decision maker has an objective other than maximizing happiness, or (2) the decision maker has tried and failed to predict what would make him or her happy.

Expected utility has generally been taken as the normative model of rational choice (Luce and Raiffa 1957), and behaviors that depart from it, such as those that Prospect Theory is invoked to explain, are cast as irrational "anomalies." Kahneman (1999) suggested that experienced utility might be one criterion by which to judge the rationality of choices. By this metric, choices based on reference-dependent preferences might be rational if the experienced utility of its outcomes are also reference dependent (see, also Bell et al. 1988). The Prospect Theory value function is explicitly a representation of the decision utility of outcomes, rather than the actual hedonic experience of those outcomes (Kahneman 1999). Whether the characteristics of the value function (such as loss aversion) also apply to experienced utility is largely an open question. Studies of choice often implicitly assume an equivalence between experience and decision utility, but choice data can at best only reveal the decision utility of an outcome.

There is significant evidence that people often fail to correctly anticipate and predict their affective reactions to events, and thus reason to expect that decision utility might indeed depart from experienced utility. Research on affective forecasting has shown that, while people typically predict the valence of their emotional reaction to an event correctly, they often incorrectly predict either the intensity or the duration of that reaction, or both (Gilbert et al. 1998, Wilson and Gilbert 2003). Furthermore, the prevailing finding is systematic overestimation of both the intensity and duration of emotional reactions, termed the *impact bias.* Of particular relevance for this study, Kermer et al. (2006) found that individuals overestimate the hedonic impact of losses, suggesting that loss aversion is a type of affective forecasting error, although Boyce et al. (2013) have found differently. Our study elicits reports of both predicted satisfaction before the marathon and actual satisfaction after the marathon and thus provides a field replication of Kermer et al.'s findings.

In sum, our study extends research on reference dependence and loss aversion into a novel domain, one that allows us to pose and answer questions that are difficult to address in more traditional decision-making settings. As a result, our findings contribute to the understanding of the impact of non-status

[3] Note that while this may suggest that loss aversion is attenuated through experience, these findings are also consistent with a formulation of reference dependence that adopts the decision-maker's expectation as the reference point from which trades are evaluated. If more experienced traders have a higher expectation of selling their wares than inexperienced traders, they will also be more likely to experience a sale as less of a loss (Kőszegi and Rabin, 2006).

quo reference points, the moderators of this impact, and the normative status of reference dependence and loss aversion, while providing evidence of the real-world relevance of these phenomena in a setting with highly motivated and experienced participants.

3 Methods

3.1 Participants

We recruited 2,652 participants registered for one of 15 target marathons conducted from 2007 to 2009.[4] Each marathon was among the 20 largest U.S. marathons, with the number of finishers ranging from 6,875 for the 2008 Grandma's Marathon to 38,557 for the 2007 New York Marathon.[5] Participants were compensated by being entered into a random drawing for prizes. We awarded one grand prize (a prize worth approximately \$300, such as a GPS watch, an iPod, or a Bose SoundDock), three second prizes (a prize worth approximately \$100, such as a running jacket, an iPod Shuffle, or a heart rate monitor watch), and 10 third prizes (a prize worth approximately \$10, such as running socks, running gloves, or a winter hat). A unique drawing was conducted for each marathon. Participants received no other compensation.

We dropped participants from our sample who did not complete the entire study (23.5%), who did not start (2.3%) or finish (4.6%) the marathon they entered, whose data could not be matched to official marathon results (2.5%), or who had participated in a previous study (0.6%). Additionally, we dropped participants from one of our 15 marathons, the 2007 Chicago Marathon.[6] This marathon took place on a day of record heat (reaching a peak of 87 degrees, the hottest October 7th on record, compared with a daily average high of 66 degrees for that day), causing hundreds of runners to fall ill or drop out, and prompting the organizers to stop the race 3.5 hours after it began.[7] Of those who finished the race before it was stopped, a significantly smaller proportion met their time goal relative to other runners in our sample ($M = 3.33\%$ versus $M = 26.93\%$, $\chi^2(1) = 35.67$, $p < .001$), and runners in that marathon also finished the marathon further behind their goal than other runners ($M = 50.75$ minutes versus $M = 14.80$ minutes, Welch's t-test, $t(165.53) = 13.83$, $p < .001$).[8] These exclusions left us with 1,633 runners.

Our sample was reasonably representative of the overall population of marathon finishers used in our study. We use the 258,128 marathon finishers in our 14 sampled marathons as the "population" and the 1,633 runners described above as our "sample."[9] Our sample ($M = 272.94$) had slightly faster finishing times than the population of marathoners ($M = 279.55$). The mean age of our runner sample, 37.47, was almost identical to the mean age of runners in the marathon population, 37.27. However, our sample had more women ($M = 56.7\%$) than the population ($M = 39.8\%$) consistent with the general finding that women are more likely to participate in surveys than men (Curtin, Presser, and Singer 2000).

[4]The marathons surveyed were Marine Corps (2007-2009), New York City (2007), Portland (2007), Twin Cities (2007-2009), Boston (2008), Grandma's (2008), Los Angeles (2008), San Diego Rock 'n' Roll (2008), and Chicago (2007-2009).

[5]http://www.marathonguide.com/Features/Articles/2007RecapOverview.cfm. Referenced on November 12, 2014.

[6]When the data from the 2007 Chicago Marathon are included the results are directionally consistent with those presented in this paper, with wider confidence intervals. These results can be accessed here: https://faculty.chicagobooth.edu/george.wu/research/withchicago.pdf

[7]http://www.nytimes.com/2007/10/08/us/08chicago.html. Referenced on November 12, 2014.

[8]Welch's t-test is an adaptation of the standard t-test for samples of unequal variances. The adjustment results in a non-integer value for degrees of freedom.

[9]We created weighted averages by weighting the relevant statistics by the proportion of our sample in each marathon. For example, 20.33% of our participants ran the 2009 Marine Corps Marathon. To compute the weighted average finishing time, we multiplied the finishing time for all runners in the 2009 Marine Corps Marathon (275.87 minutes) by 20.33%, repeated this process for the other 13 marathons, and summed the 14 products.

3.2 Conditions

Participants were randomly assigned to one of six conditions created by crossing three pre-marathon conditions with two post-marathon conditions. Participants in our study were blind to condition and were unaware of our specific research hypotheses other than that we were examining the "relationship between marathon performance and satisfaction." We describe the pre- and post-marathon conditions in more detail below.

3.2.1 Pre-Marathon Conditions

The pre-marathon conditions consisted of a "goal-not-asked" condition, an early "goal-asked" condition and a late "goal-asked" condition. For the 2009 marathons, we simplified the design and dropped the early goal-asked condition.

Participants in the goal-not-asked condition were surveyed approximately two weeks prior to the marathon and asked to provide some basic demographic information, including their age, gender, and information on their training and running background. Runners indicated whether they had previously completed a marathon, and provided their fastest and most recent marathon times, and their fastest half-marathon and 10 kilometer times. Participants in this condition were not asked about goals or any other objectives they might have had for their upcoming marathon.

Participants in the early goal-asked condition completed two surveys prior to the marathon, administered approximately 6 and 2 weeks before the marathon, while participants in the late goal-asked condition completed only one survey, administered approximately 2 weeks prior to the marathon. We asked participants in both of the goal-asked conditions to report the same demographic and experience information as the goal-not-asked participants. In addition, participants in the goal-asked conditions were asked about time goals and other objectives. In particular, participants were asked if they had a "specific time goal for the marathon," and if so, to indicate it. They also estimated the likelihood of reaching the goal and were asked to predict their satisfaction on a 1 to 7 scale for finishing 1, 10, and 20 minutes ahead of or behind their time goal. Participants in the early condition answered these questions twice, while participants in the late pre-condition answered these questions once. The complete battery of questions for the pre- and post-marathon surveys, including the wording and order, is found in Sackett, Wu, White, and Markle (2014b). We included these three conditions to test whether providing a goal prior to the marathon increased the commitment to that goal, possibly leading to increased loss aversion, as well as to test whether providing a goal prior to the marathon affected performance.

3.2.2 Post-Marathon Conditions

The two post-marathon conditions differed only in timing. Participants in the early condition were contacted 1 day after the completion of their marathon, while participants in the late condition were contacted 4 weeks after the marathon finished. In both cases, participants had 2 weeks to complete the survey. We omitted the late post-condition for the 2009 marathons.

Participants were asked if they started the marathon, and if so, whether they completed the race. Runners who completed the marathon then answered the following question: "How satisfied are you with your performance in the Marathon?" on a 1 to 7 Likert scale. Participants later indicated their time goal if they had one. Note that participants in the goal-asked pre-marathon conditions had previously provided a goal, while participants in the goal-not-asked condition had not. Participants also reported their official time, as well as their bib number. We included the two conditions to test whether loss aversion was attenuated over time, as might be predicted by some accounts of affective forecasting (Gilbert et al. 1998).

3.3 Official Performance Data

We used a participant's bib number to match that participant with official marathon results obtained from marathon websites. These results usually included split times, such as times for the half marathon, as well as for other intermediate distances such as 10, 20, 30, and 40 kilometers. (A marathon is approximately 42.2 kilometers or 26.2 miles long.) Marathoners in large races such as the ones in our study are given two times, a "chip time" (also known as "net time") and a "clock time" (also known as "gun time"). The chip time reflects the elapsed time between when a runner crosses the start line and the finish line, while the clock time is the difference between when a runner finishes the race and when that race officially starts. For large races, the difference between the clock and chip time can be sizable. It took runners in our sample 6.73 minutes on average to reach the starting line. We use the chip time throughout as a measure of performance, as this is the number that most runners take to be the most relevant measure of performance (Austen 2001).[10]

4 Basic Results

In this section, we describe some basic results. In Section 5, we organize these results by estimating a Prospect Theory value function to relate satisfaction with relative performance.

4.1 Predicted Satisfaction

We asked participants in the goal-asked conditions to predict their satisfaction from finishing 1, 10, and 20 minutes ahead or behind their goal. The 20 minute questions were added for the 2008 and 2009 surveys. Thus, this analysis only considers 2008 and 2009 participants.

The average of these estimates is shown in Figure 1, where positive values along the x-axis indicate performance in excess of the goal. Note that this figure traces the characteristic shape of the Prospect Theory value function. The slope for performance above the goal is shallower than the slope for performance below the goal. Let S_t be the average predicted satisfaction for finishing t minutes above the goal. To test for loss aversion, we take $S_{+10} - S_{+1}$ as a measure of the incremental satisfaction of finishing an additional 9 minutes better than a goal and $|S_{-10} - S_{-1}|$ to be the incremental dissatisfaction of finishing an additional 9 minutes worse than a goal.

The gain ($M = 0.58$) and loss ($M = 1.43$) slopes are significantly different (Wilcoxon signed-rank Test, $p < .001$).[11] The loss slope exceeds the gain slope, $|S_{-10} - S_{-1}| > |S_{+10} - S_{+1}|$, consistent with loss aversion, for 58.7% of our participants and is in the direction counter to loss aversion for just 12.6% of our participants ($\chi^2(1) = 263.51$, $p < .001$). The ratio of the mean loss slope to the mean gain slope is 2.46, a value in the range of estimates of loss aversion (Camerer 2005).

Predicted satisfaction also exhibited diminishing sensitivity for gains, $S_{+10} - S_{+1} > S_{+20} - S_{+10}$ (Wilcoxon signed-rank Test, $p < .001$), as well as losses, $|S_{-10} - S_{-1}| > |S_{-20} - S_{-10}|$ (Wilcoxon signed-rank Test, $p < .001$). Note that this is a conservative test, because the intervals are not the same length.

[10] The chip time is used as a qualifying time for "elite" races such as the Boston Marathon. However, clock time is the time generally used for determining prize money and other awards.

[11] We use a non-parametric test because the ordinal scale underlying the satisfaction measure violates the assumptions for t-tests (Siegel and Castellan 1988). Later, we use Spearman correlation instead of Pearson correlation for the same reasons. However, the results for parametric and non-parametric tests are nearly identical.

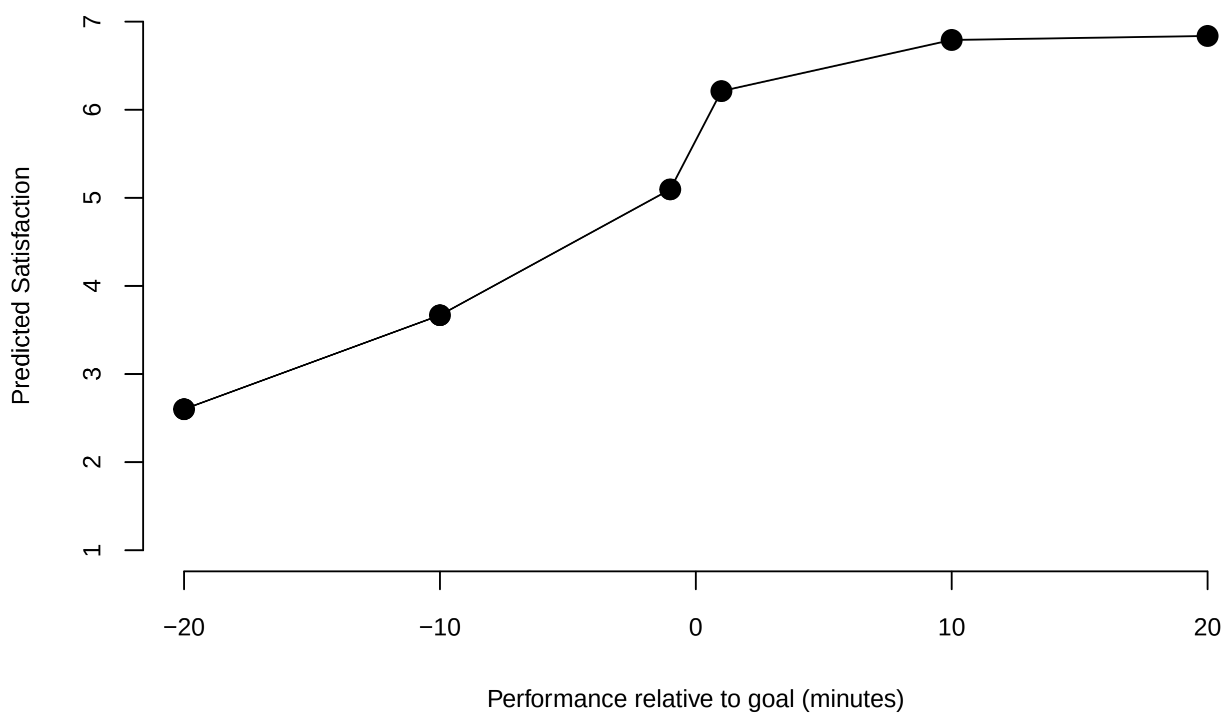

Figure 1: Predicted satisfaction as a function of relative performance

4.2 Time Goals

Overall, 85.7% of our participants had time goals. Experienced marathoners (M = 88.1%) were more likely to have goals than were rookie marathoners (M = 79.7%) ($\chi^2(1) = 18.26$, $p < .001$). 48.5% of experienced marathoners set goals faster than their best marathon time, and 66.1% set goals faster than their last marathon time.

We asked participants to indicate, on a 1 to 7 scale, the importance of attaining their time goal. Runners on average rated the goal as being important ($M = 5.80$), with 86.4% of runners rating the goal as 5 or higher in importance. Participants were also optimistic about the chances of reaching their goal, indicating a 71.6% average chance of achieving their goal.

Recall that we elicited goals from participants in the goal-asked conditions prior to the marathon but asked all participants to indicate their goal after the marathon. Participants in the goal-asked condition exhibited a small but statistically significant tendency to report less ambitious goals after the marathon ($M = 251.78$) than before it ($M = 247.87$) (paired t-test, $t(740) = 2.92$, $p < .005$). Although we use the post-marathon report of goals throughout to maximize our sample size, we will also examine whether our results depend on whether we use pre- or post-marathon goals.

Recall also that our pre-marathon survey was administered at two different times. Participants in the early pre-condition ($M = 242.79$) set more ambitious goals than those set by participants in the late pre-condition ($M = 249.75$) (Welch's t-test, $t(348.61) = 1.97$, $p < .05$). Sackett, Wu, White, and Markle (2014a) explore the effect of the timing of goal elicitation on performance. In most of the analyses in this paper we collapse these two conditions, although we do examine whether the relationship between satisfaction and relative performance is influenced by the timing of goal elicitation.

4.3 Performance

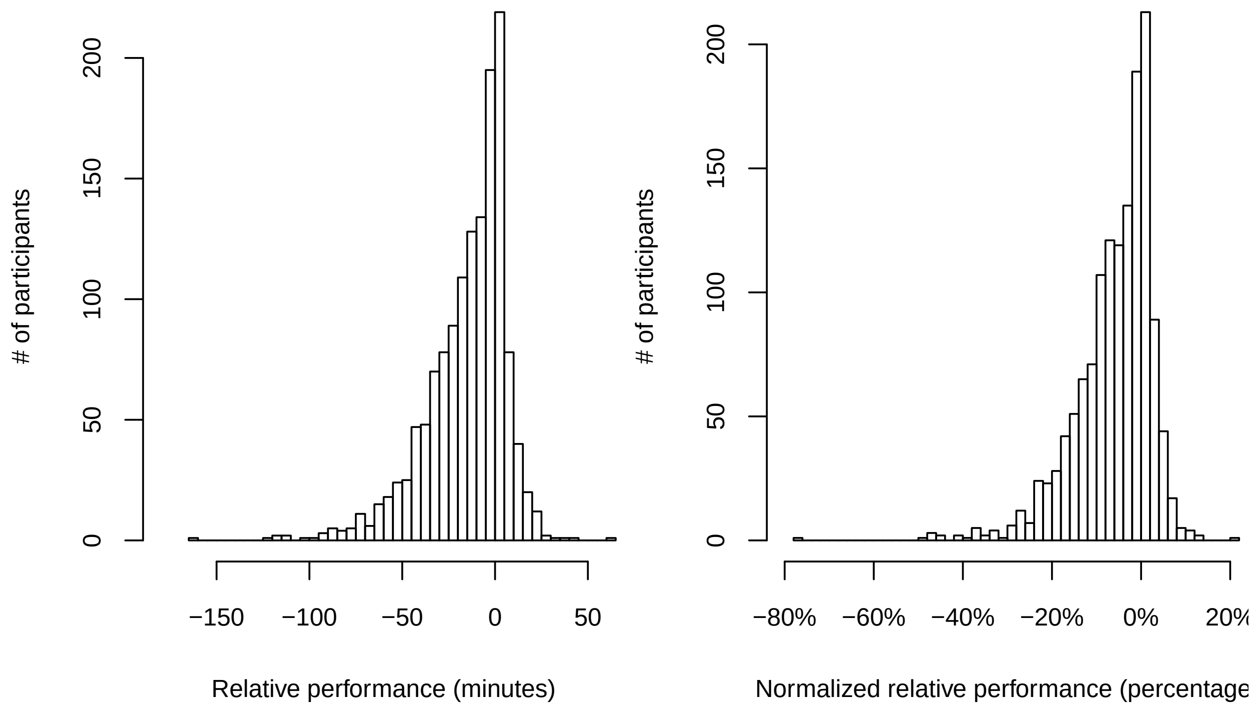

Figure 2: Histograms of finishing time relative to goal, in minutes (time goal - finishing time) (left panel) and percentage ((time goal - finishing time) / time goal) (right panel).

The average finishing time for our runners was 4:32:56 (male: 4:16:40; female: 4:45:21), with finishing times ranging from approximately 2:15:00 to 8:15:00.[12] 26.9% of our runners ran faster than their time goal, with the success rate varying considerably across marathons, ranging from 7.4% for the 2007 Twin Cities Marathon to 45.8% for the 2009 Twin Cities Marathon. 25.8% of runners who rated achieving their time goal as 7 out of 7 in importance met their goal, compared to 27.6% of all remaining runners ($\chi^2(1) = 0.45$, $p = .50$). In addition, 25.4% of experienced (i.e., non-rookie) marathoners beat their time goal, compared to 31.4% of first-time marathoners ($\chi^2(1) = 4.65$, $p < .05$).

We define "relative performance" to be the difference between finishing time and time goal. Relative performance is defined such that exceeding the goal produces a positive value. Runners fell short of their goal by 14.80 minutes on average (Median: 9.69). To deal with the large variation of finishing times, we also create a "normalized relative performance" measure by dividing relative performance by goal time. For example, a runner who ran 4:12 and had a goal of running 4:00 was 5% (12/240) short of their goal. The mean normalized relative performance is 5.92% (Median: 3.99%). Figure 2 shows histograms of both measures of relative performance.

4.4 Satisfaction

Mean satisfaction was 5.00 out of 7. Satisfaction was negatively correlated with finishing time, with faster finishers more satisfied than slower finishers (Spearman ρ = -0.19, $p < .001$). However, the correlation was much stronger if we considered performance relative to a runner's time goal, either relative performance in minutes (Spearman $\rho = 0.66$, $p < .001$) or in percentage terms (Spearman $\rho = 0.67$, $p < .001$).

[12]Precise finishing times are not provided to protect the anonymity of our participants.

5 Estimation of the Value Function

In this section, we test whether the relationship between satisfaction and relative performance follows the shape of the Prospect Theory value function. There are several challenges to fitting a value function to satisfaction data. First, each participant provided only one satisfaction rating, and thus we must estimate an aggregate value function. Second, our satisfaction measure was elicited on a 1 to 7 Likert scale. Such a measure is often interpreted as being a cardinal measure, though more conservatively it should be thought of as merely reflecting an ordinal scale (cf. Long and Freese 2005). The discreteness and boundedness of this scale violates the standard regression assumption of normally distributed errors, making OLS inappropriate. Finally, there is the challenge of measuring loss aversion. Köbberling and Wakker (2005) discuss the difficulties of defining loss aversion, in particular in cases in which the value functions for gains and losses do not have the same shape.

To deal with the first challenge we assume that satisfaction is a function of performance relative to the goal, and use normalized relative performance (as defined in Section 4.3), rather than absolute performance in minutes, in order to make performance across participants more readily comparable and appropriate for aggregation. To deal with the second challenge we employ ordered logistic regression, which does not assume equal spacing between the categories of the response variable.[13] In estimating the value function, we adopt a non-parametric approach using polynomial splines. Polynomial splines allow us to investigate reference dependence and loss aversion using a "light touch" that does not assume any particular functional form, however due to the lack of symmetry in curvature across gains and losses they do not lend themselves naturally to crisp summaries of loss aversion. In addition, standard parametric forms such as those used in Tversky and Kahneman (1992) cannot readily be incorporated into ordered logistic regression models. To deal with the challenge of measuring loss aversion in this scenario, we adapt a definition of loss aversion for use with non-symmetric value functions.

5.1 Spline Regression

We begin by fitting a polynomial spline to our data as a means of summarizing and exploring the relationship between satisfaction and performance without imposing any particular assumptions about the form of that relationship. Splines are piece-wise functions, composed of separate polynomials defined over intervals of the predictor variable and connected across those intervals by knots (Marsh and Cormier 2001). Here we adopt a quadratic spline with a single knot at 0, the assumed reference point. Satisfaction is taken to be a function of a runner's normalized relative performance (x). The spline takes the form:

$$S(x) = a + \begin{cases} b_1 x + c_1 x^2, & x \le 0 \\ b_2 x + c_2 x^2, & x > 0 \end{cases} . \tag{1}$$

The satisfaction function is thus composed of two polynomials, one each for gains and losses. The two polynomials share a constant term and therefore must meet at the reference point, a constraint we later relax. This specification permits a discontinuity at the knot in the function's first and second derivatives. Loss aversion is commonly defined as a kink or a discontinuity in the first derivative of the value function at the reference point, while the "S" shape of the value function arises from an inflection point, or a change of sign in the second derivative, as the function switches from convexity below the reference point to concavity above it (Köbberling and Wakker 2005). A quadratic spline is therefore the minimal order that is required to capture the relevant characteristics of the Prospect Theory value function, and is chosen for parsimony.

[13] All of the following analyses were also conducted using an ordered probit regression model without any change in the basic results.

We fit Equation (1) using ordered logistic regression, which estimates the probability that the dependent variable will take on each of its possible levels (in our case, the Likert scale values 1 through 7). To summarize the results we calculate the expected value of the dependent variable by multiplying the probability of each level by the value of that level, and summing across the 7 levels. Expected satisfaction as a function of normalized relative performance is shown in Figure 3 (left panel), plotted along the domain -24% to 6%. Recall from Section 4.3 that the distribution of performance in our sample is skewed, with the large majority of participants falling short of their time goal, so this domain incorporates an approximately equal share of the observations falling above (95.4%) and below (92.3%) zero. Figure 3 (left panel) also displays the 95% confidence band around the predicted satisfaction values in the shaded area, generated via bootstrapping (Efron and Tibshirani 1993).

We test whether the fitted spline exhibits the three main properties of the Prospect Theory value function: convexity below the reference point, concavity above the reference point, and loss aversion. The spline largely meets the first two criteria. It is convex throughout most of the range below zero (from the lower bound to -0.05%), and is concave everywhere above zero.

To test for and measure loss aversion we employ the definition from Wakker and Tversky (1993), according to which a value function exhibits loss aversion if $u'(-x) > u'(x) \; \forall \, x$. The fitted spline fails to exhibit loss aversion by this definition, holding only for $x > 2.4\%$. A loss aversion parameter can be derived from the ratio of the slope of the function above and below the reference point, $u'(-x)/u'(x)$. This measure varies over x. To arrive at a single loss aversion coefficient, we calculate the above ratio for 60 evenly spaced values of x over the range from 0 to 6%, and take their median. This process yields a loss aversion parameter estimate of $\hat{\lambda} = 1.21$. Using bootstrapping, we calculate 95% confidence intervals for this estimates of [0.90, 1.75].

Recall that the spline specification in Equation (1) requires continuity at the knot. We noted in the introduction that some models of aspiration levels have incorporated a discontinuity, or jump, at the aspiration level itself. If such a jump exists, the continuity constraint will result in an underestimation of loss aversion. Accordingly, we modify Equation (1) to allow for a discontinuity at zero, by adding an additional constant term J, as follows:

$$S(x) = a + \begin{cases} b_1 x + c_1 x^2, & x \leq 0 \\ J + b_2 x + c_2 x^2, & x > 0 \end{cases} . \tag{2}$$

The refitted spline is shown in Figure 3 (right panel), along with its 95% bootstrap confidence interval. The coefficient on J is positive and statistically significant ($\beta = 0.91$, $\chi^2(1) = 20.85$, $p < 0.01$), resulting in a jump in expected satisfaction at the reference point of 0.50 on our 1 to 7 scale. The model incorporating the jump also has a better fit than the one that requires continuity (likelihood ratio test, $\chi^2(1) = 21.08$, $p < 0.01$).

This spline exhibits convexity in most of the region below zero (from the lower bound to -4.45%), and is concave everywhere above zero. Importantly, the spline with a jump now satisfies the definition of loss aversion for all values of x. It produces a loss aversion parameter of $\hat{\lambda} = 2.24$, 95% CI $= [1.33, 5.35]$.[14] The 95% confidence interval for the size of the jump at the reference point is [0.26, 0.71].

5.2 Model Validation

We take a number of steps to verify our model specification and test the robustness of our findings. We refit our model with the addition of control variables for marathon and gender, with alternative reference point locations, and with alternative definitions for performance. We also identify and remove influential outliers from our data, and refit our model to the reduced data set.

[14] Estimated loss aversion coefficients were above 1 for all ranges from 0-1% to 0-10%. Loss aversion coefficients were significantly greater than 1 from 0-5% to 0-9% at the 0.05 level (one-tailed test).

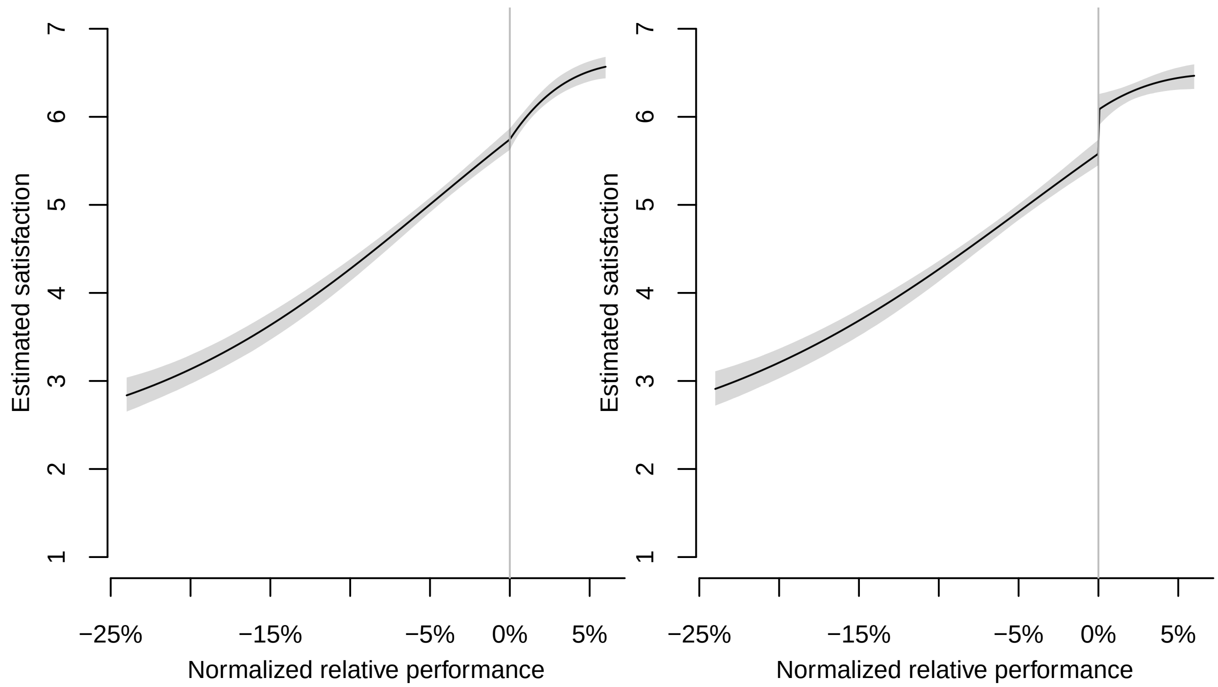

Figure 3: Quadratic spline with a single knot at zero, fitted to runner satisfaction. The spline in the left panel requires continuity at the reference point, while the spline in the right panel allows for a discontinuity at the reference point. The shaded bands indicate the 95% bootstrap confidence interval.

5.2.1 Control variables

We refit Equation (2) with the addition of a dummy variable for gender, and dummy variables for each of the 14 marathons we sampled. We then estimate a satisfaction function by taking all the control variables at their mean levels. The results are nearly identical to our model without control variables. The loss aversion coefficient $\hat{\lambda} = 1.97$, 95% CI = [1.16, 4.57], with a jump at the reference point of size 0.52, 95% CI = [0.27, 0.73].

5.2.2 Location of Reference Point

The above analysis assumed that the time goal serves as a reference point for evaluating performance. While the results support this assumption, here we perform two checks on the placement of the reference point. First we refit the regression splines (both with and without the jump at the reference point), varying the placement of the knot between -5% and 5%. For the model with a jump at the reference point, residual deviance is minimized when the knot is placed at 0.31%. For the model with no jump at the reference point, residual deviance is minimized when the knot is placed at -3.10%. The close proximity of the best fitting knot placement to zero supports our assumption that the time goal serves as a reference point in the sense that outcomes on the two sides of the reference point are viewed as qualitatively different.

5.2.3 Alternative Reference Points

We next contrast our goals-as-reference-points model with two alternative models: one in which performance is defined relative to a runner's best marathon time, and one in which performance is defined relative to a runner's most recent marathon time. Because the latter two models only apply to runners with previous marathon experience, our data set for this analysis is restricted to experienced runners. We further restricted the data set by removing participants whose goal time was identical to either their best or their last marathon times. This leaves 940 observations, or 67.29% of our sample.

The model based on goal time as the reference point provides a notably better fit, with residual deviance of 2807, versus 3196 and 3222 for the models with best marathon time as the reference point and last marathon time as the reference point, respectively.

5.2.4 Alternative Performance Specifications

As noted at the beginning of Section 5, in this analysis we have expressed x (a runner's performance relative to her goal) as a percentage of that runner's time goal as a means of placing the performance of all runners on a common scale. We refit Equation (2) using relative performance (in minutes) and we find that the resulting function satisfies the definition of loss aversion for all values of x, producing coefficient $\hat{\lambda} = 2.12$, 95% CI = [0.63, 4.74], with a jump at the reference point of size 0.58, 95% CI = [0.35, 0.78]. While the point estimate for loss aversion is similar to that obtained using our original specification for relative performance, the confidence interval is wider. This is expected, given that falling 10 minutes behind a 6 hour goal is likely to provoke a milder reaction than falling 10 minutes behind a 3 hour goal. Treating them identically thus leads to a noisier estimate of loss aversion.

5.2.5 Outlier Analysis

Next we identify and drop influential outliers from our data, and refit the models both with and without the discontinuity at the reference point to the reduced data set. DFBETAS (standardized difference of the beta) measures the degree of influence of a single observation on the coefficients of the fitted model, where the larger the value of DFBETAS, the greater the influence of the observation on a particular coefficient.[15] By this method we identify 102 data points that are influential outliers in the model with no discontinuity from Equation (1) (7.3% of our sample). Refitting this model to the reduced data set with outliers removed, we find the function is closer to (but still does not perfectly satisfy) the definition of loss aversion, meeting the definition for $x > 0.1\%$ (relative to $x > 2.4\%$ as reported earlier for the model fit to the entire data set). Calculating a loss aversion coefficient for the model refit to the reduced data set with outliers removed produces $\hat{\lambda} = 2.37$, 95% CI = [1.27, 3.36].

We identify 92 data points that are influential outliers in the model with a jump at the reference point from Equation (2) (6.59% of our sample). This model, when refit to the data set with outliers removed, continues to satisfy the definition of loss aversion. Refitting this model to the reduced data set produces $\hat{\lambda} = 2.87$, 95% CI = [1.62, 6.94], and a jump of size 0.33, 95% CI = [0.08, 0.56].

5.3 Evidence for Causality

In Section 5.1, we fit an aggregate value function using a representative-agent framework. Because we did not randomly assign goals to marathon participants, it is possible that marathoners who surpass their goals differ from runners who fall short of their goals in respects other than their performance (a similar critique applies to Boyce et al. 2013). In other words, the value function depicted in Figure 5.1 might capture a mixture of agents, none of whom exhibit the characteristic Prospect Theory value function. To

[15] Specifically, the DFBETAS for an observation is the difference between a regression coefficient for a particular variable calculated when the model is fit to the entire data set including that observation, and that when the model is fit to a data set with that observation removed, scaled by the standard error from the second fitting. A standard cutoff for larger data sets is DFBETAS$> \pm\frac{2}{\sqrt{n}}$ (Agresti 2007).

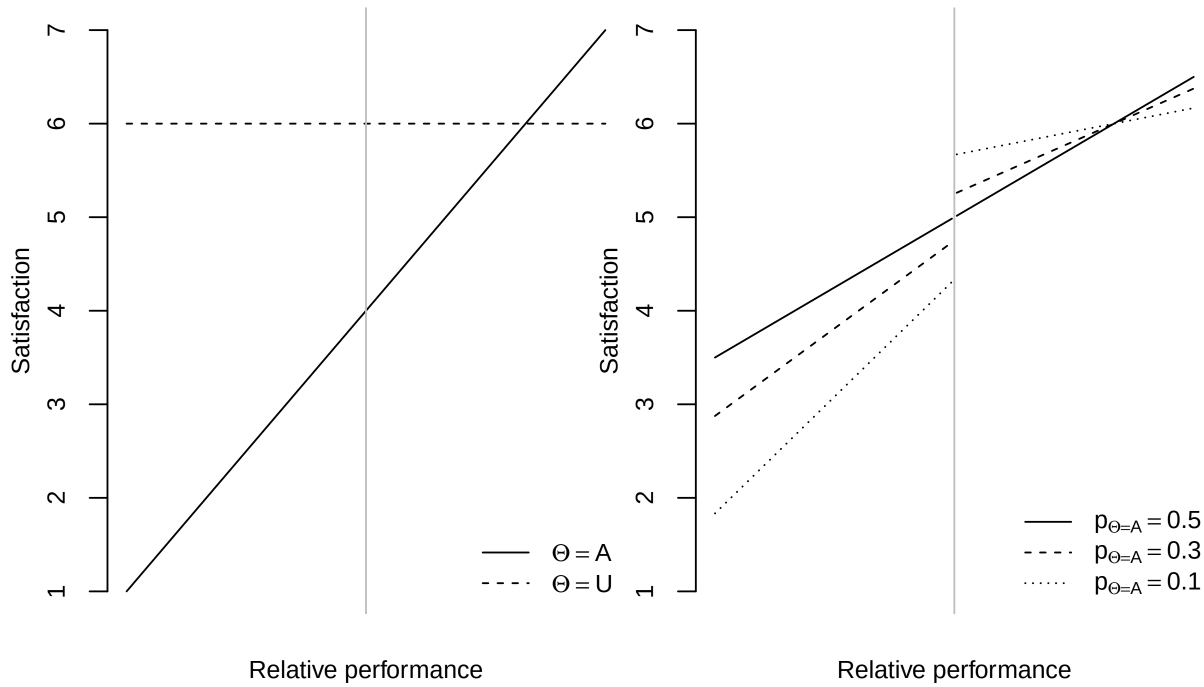

Figure 4: Two-type alternative model. Satisfaction of ambitious and unambitious types as a function of relative performance (left panel), and aggregate satisfaction for mixtures of the two types (right panel).

illustrate, we sketch a two-type model and show how this model could produce a pattern that resembles Figure 3 (with the exception of diminishing sensitivity).

Suppose that there are two types of runners, "ambitious" (denoted $\Theta = A$) and "unambitious" runners (denoted $\Theta = U$). Ambitious runners set aggressive goals and achieve these goals very infrequently. Unambitious runners establish modest goals and mostly achieve these goals. If satisfaction for ambitious runners is linear and monotonically increasing in performance relative to the goal, and satisfaction for unambitious runners is flat and high (above 6 on the 1 to 7 Likert scale), a mixture of these two types could easily produce what appears to be loss aversion and/or a jump in aggregate, even though both types have a linear satisfaction function.

This possibility is illustrated in Figure 4. The left panel shows the satisfaction each type experiences for a range of performance relative to the time goal. The right panel displays the aggregate satisfaction function resulting from averaging the satisfaction of the two types, given the simplifying assumption that the two types are equally prevalent in the population and that the proportion of runners exceeding their time goal differs across the two types.

We hold the success rate for unambitious types, $p_{\Theta=U}$, constant at 0.5, and vary the success rate for ambitious types, $p_{\Theta=A}$. When the proportion of runners meeting or exceeding their goals is equal across the two types, the aggregate satisfaction function is linear with no jump. As fewer of the ambitious types make their goal relative to the unambitious type, a jump emerges at the reference point and the slope in the satisfaction function above the reference point becomes flatter than that below the reference point. When $p_{\Theta=A} = 0.2$, the aggregate satisfaction function exhibits a jump of size 0.89, and $\lambda = 2.50$, values which are similar to the results obtained in our analysis. Of course, it is not critical that the slope of the satisfaction function for unambitious types be flat, but merely more shallow than for ambitious types.

We examine whether our data are consistent with this two-type model by exploiting a natural experiment in our data. Our data sample includes 14 marathons over 3 years. These marathons differ considerably in the overall rate in which marathoners in our sample achieve their goals. Only 7.4% of Twin Cities Marathon in 2007 met their goal, compared to 45.7% of runners in the 2009 Twin Cities Marathon. Not surprisingly, the rates of goal attainment are driven mostly by weather. The high temperature for the Twin Cities Marathon in 2007 was 82 degrees, while the high temperature in 2009 was 55 degrees. Indeed, goal attainment in our sample is closely related to the mean finishing time for that race. For each marathon, we calculate a residual finishing time by regressing finishing times for each of the marathons in our sample (Boston, Chicago, Grandma's, Los Angeles, Marine Corps, New York, Portland, Rock 'n' Roll, Twin Cities) including fixed effects for that marathon and dummies for all years from 2000 to 2010. The correlation between the residuals of that model and our goal attainment rates is .75.

If the two-type model holds and the exogenous shock of weather does not influence the proportion of types $\Theta = A$ and $\Theta = U$, only the likelihood that ambitious runners achieve their goals, then we should see more loss aversion and larger jumps for low success rate marathons and less loss aversion and smaller jumps for high success rate marathons. To test this implication, we fit the value function to each of our 14 marathons separately, using a linear spline to reflect the relatively small size of some of the subsamples. This yields a loss aversion coefficient and a jump coefficient for each of our 14 marathons. We then correlate the loss aversion coefficients and the jump coefficients with the rates of goal attainment. The two-type model predicts a negative correlation between success rate and loss aversion or jump. On the contrary, we find a positive but insignificant correlation between success rate and loss aversion (Spearman $\rho = .34, p = .24$) and between success rate and the size of the jump (Spearman $\rho = .05, p = .88$). This pattern is consistent with relative performance driving satisfaction, rather than a story in which heterogeneous types simultaneously drive a pattern of performance and satisfaction that mimics reference dependence.

5.4 Multiple Reference Points

The analysis in Section 5.2.3 suggests that comparisons between performance and time goal drive satisfaction to a greater degree than do comparisons between current performance and performance in a past marathon. However, this does not preclude the possibility that the latter comparisons are also salient to runners and thus additionally influence their satisfaction as in Kahneman's (1992) multiple reference point account. Therefore we test whether incorporating multiple reference points in the model contributes to its explanatory power. To do so, we add a dummy variable to the goals-as-reference-points model indicating whether the participant outperformed their best marathon time. This model provides a better fit than the model incorporating only the time goal as a reference point (likelihood ratio test, $\chi^2(1) = 28.03$, $p < 0.01$), producing an additional boost in satisfaction of 0.72 on our 1-7 scale. We obtain similar results using last marathon time as a secondary reference point.

5.5 Moderators of Loss Aversion

In the previous sections we have established that runners' satisfaction, as a function of performance relative to their goal, exhibits the key characteristics of the Prospect Theory value function. Here we investigate how runners' experience and the importance of the time goal moderate that relationship.

5.5.1 Goal importance

Participants in our study were asked to rate the importance they attach to meeting their time goal on a 1-7 scale. The average importance rating was 5.80 (Median = 6.00, $SD = 1.26$). We refit Equation (2) with the addition of a dummy variable indicating whether each participant reported a goal-importance rating of of 6 or 7 (66.2% of our sample), or 5 or less (33.8% of our sample), along with the interactions of that variable with the linear and quadratic terms in the spline specification. Loss aversion estimates from this models are presented in Table 1.

Permutation tests reveal that loss aversion is higher for runners who rate meeting their goal a 6 or 7 ($\hat{\lambda} = 3.24$) relative to those whose goal-importance rating was a 5 or below ($\hat{\lambda} = 0.59$) ($p = .06$).[16] The size of the jump at the reference point is also larger for those with a goal-importance rating of 6 or 7 (0.49) than for those with a goal-importance rating of 5 or less (0.06) ($p = .05$).

5.5.2 Runner Experience

Among the runners in our sample, 73.5% had completed at least one prior marathon, and on average had completed 5.3 prior marathons (Median = 2.00, $SD = 11.40$). The correlation between reported goal-importance and the number of previously completed marathons is positive and significant (Spearman $\rho = 0.06$, $p < .05$). We fit the model in Equation (2), adding a dummy variable to indicate whether each participant had previously completed a marathon and the interactions of that dummy variable with the linear and quadratic terms in the spline specification. Loss aversion estimates from this model are presented in Table 1.

Permutation tests reveal no significant effect of experience on loss aversion. Loss aversion for novice marathoners ($\hat{\lambda} = 2.77$) is not significantly different from that for experienced marathoners ($\hat{\lambda} = 2.14$) ($p = .72$). The magnitude of the jump at the reference point for novice marathoners (0.54) is also not significantly different from that for experienced marathoners (0.47) ($p = .65$)

5.5.3 Timing of Elicitation of Goals and Satisfaction

Recall from Section 3.2 that participants in some conditions were asked to report their time goal ahead of the marathon, while others were asked only for demographic information prior to the marathon. All participants were asked for their time goal after the marathon, and it is this goal that has been utilized in the above analyses. Here we examine whether explicitly asking for a goal prior to the marathon influences the degree of reference dependence.

We fit the model in Equation (2), adding a dummy variable to indicate whether the participant's goal had been elicited prior to, or only after the marathon, and its interactions with the other terms in the model. The results are displayed in Table 1. We find no significant difference in the loss aversion coefficients for the participants in the goal-asked condition relative to those in the goal-not-asked condition. There is also no significant difference in the size of the jump for the goal-asked and goal-not-asked condition.

Recall also that we varied the timing of the post-marathon survey. Here we examine whether the timing of the elicitation of satisfaction matters. We fit the model in Equation (2), adding a dummy variable to indicate whether the participant's satisfaction had been elicited one day (early condition) or four weeks (late condition) after the marathon, and its interactions with the other terms in the model. The results are displayed in Table 1. We find no significant difference in the loss aversion coefficients for the participants in the early post-marathon condition relative to those in the late post-marathon condition.

[16] In a permutation test the distribution of the test statistic under the null hypothesis is obtained by randomly rearranging the labels of the data points and calculating the test statistic for each permutation. The observed value of the test statistic (in this case the difference between the loss aversion estimates of the two groups) is then compared against this distribution to calculate a p-value. See Good (2005) for details.

Moderator	Group	$\hat{\lambda}$	significance	Jump	significance
Goal Importance	Goal importance = 6-7	3.24	$p = .06$	0.49	$p = .05$
	Goal importance = 1-5	0.59		0.06	
Runner Experience	Novice marathoners	2.77	$p = .72$	0.54	$p = .65$
	Experienced marathoners	2.14		0.47	
Timing of Goal Elicitation	Goal-asked condition	2.53	$p = .36$	0.53	$p = .39$
	Goal-not-asked condition	1.21		0.45	
Timing of Satisfaction Elicitation	Early condition	1.43	$p = .93$	0.48	$p = .59$
	Late condition	4.62		0.51	
Predicted vs. Experienced	Predicted satisfaction	3.73	$p < .001$	0.88	$p < .05$
	Experienced satisfaction	1.54		0.53	

Table 1: Moderators of loss aversion

5.6 Predicted vs. Actual Satisfaction

In Section 4.1 we plotted participants' predictions about their satisfaction with several hypothetical outcomes: finishing the marathon 20, 10, or 1 minute behind their goal, or 1, 10, or 20 minutes ahead of their goal. Here we compare predicted and actual satisfaction. Recall from Section 4.1 that the 20 minutes questions were only added for the 2008 and 2009 surveys. We therefore restrict our analysis to that reduced sample. Because we have analyzed satisfaction as a function of performance expressed as a percentage of the goal time, we divide the six hypothetical outcomes by the goal time for each participant. We then regress the predicted satisfaction with those outcomes on the newly translated outcomes, now expressed as normalized relative performance, according to the model in Equation (2). The results are displayed in Figure 5.

Predicted satisfaction also conforms to the Prospect Theory value function. It is convex throughout most of the range below zero (from the lower bound to -4.4%), and is concave everywhere above zero. We estimate a loss aversion coefficient for predicted satisfaction of $\hat{\lambda} = 3.73$, 95% CI = [3.09, 4.53]. We find greater loss aversion in predicted satisfaction than in actual satisfaction by permutation test ($\hat{\lambda} =$ 3.73 vs. 1.54, $p < .001$). The difference in the size of the jump at the reference point between predicted and actual satisfaction is also significant by permutation test (0.88 vs. 0.53, $p < .05$)

6 Discussion

Investigations into reference dependence have typically involved two domains: risky decision making and the endowment effect. While these domains differ in critical ways (e.g., one is risky and the other riskless), they share important similarities. Specifically, studies in both areas have generally taken the decision maker's status quo – either their current wealth level or state of ownership of a good – to be the reference point. Also, studies in both risky decision making and the endowment effect typically employ choice or pricing as dependent measures.

While the focus on status quo reference points is sensible in these studies, for many decisions the status quo may not be a meaningful, or the only meaningful, reference point (e.g., Barberis 2013, Baucells et al. 2011). In addition, the use of choice and pricing as dependent measures constrains those studies to making inferences about decision utility alone, yielding no insight into the welfare implications of those choices (for notable exceptions, see Boyce et al. (2013) and Ockenfels et al. 2014).

Our study takes place in a domain distinct from most prior studies. Because we examine how goals influence subsequent performance and satisfaction, our study departs from most research on reference dependence both in its use of a non-status quo reference point, and in its dependent measure of satisfac-

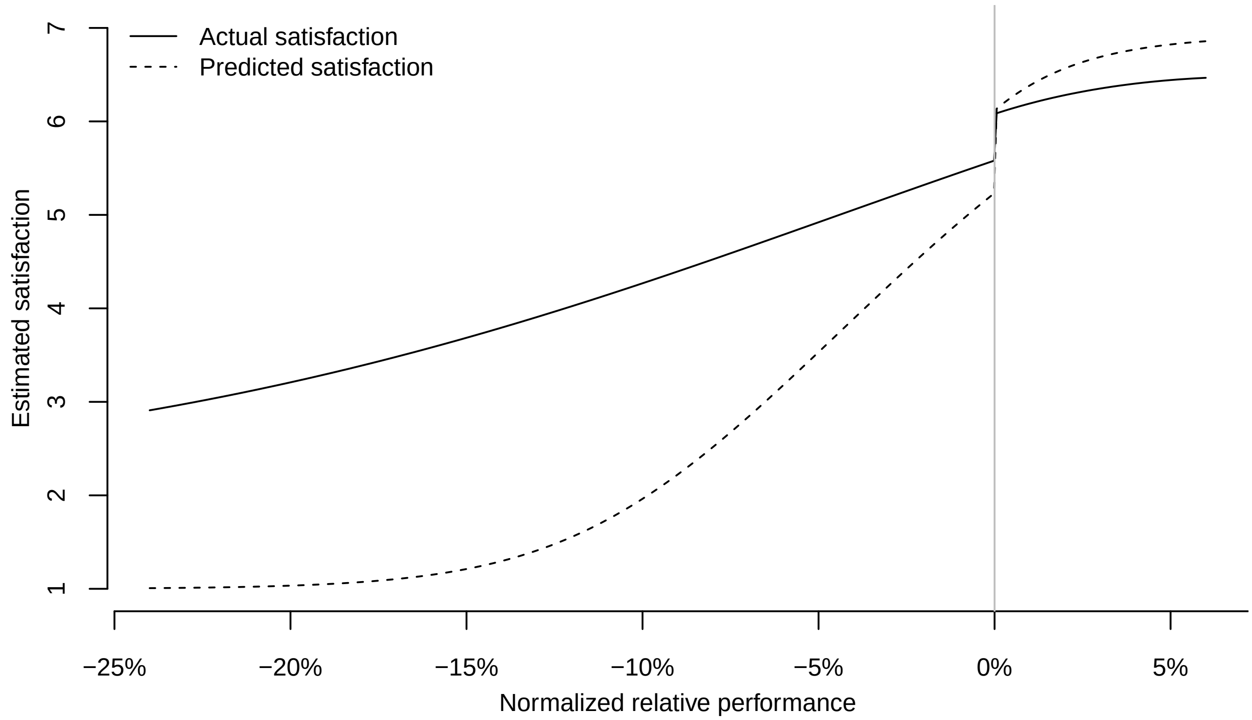

Figure 5: Comparison of predicted versus actual satisfaction

tion. These departures provide insight in a domain in which the reference point is not the status quo. Investigations of non-status quo reference points are rare, perhaps because it is not clear what reference point is pertinent to a particular decision (Barberis, 2013). For example, the debate about whether taxi drivers exhibit reference-dependent preferences is confounded by the lack of clarity about what reference point a taxi driver might be using (Camerer et al. 1997, Crawford and Meng 2011, Farber 2008).

While experimental findings have mapped out the form of the Prospect Theory value function, we are unaware of studies that map out the shape of the value function from experienced utility or for a non-status quo reference point. Our study raises the natural question of whether the utility experienced in our motivational context differs from utility experienced in a gambling context or any other setting in which Prospect Theory has been evoked. On the one hand, there are psychological reasons that the setting we have studied may differ from investigations such as Abdellaoui et al. (2007). Intuitively, barely surpassing a performance goal is likely to feel characteristically different from barely falling short, relative to the distinction between minute gains or losses in one's wealth, for example. The notion that goals may produce a more dichotomous categorization of outcomes than other types of reference points suggests that the value function around that goal may have a distinct shape. Indeed, we find evidence for a discontinuity or jump in satisfaction at the goal. Comparing our models with and without this jump, we find that the jump is accompanied by a value function that is both flatter and less curved in both losses and gains, relative to models with no jump.

Our study is silent about whether a jump is limited to goals, or whether it extends to other types of reference points. Few studies have demonstrated such a jump, and studies that fit a parametric model for the value function, such as the common power function, are likely to miss the presence of a jump. These models require continuity at the value function, producing a confound between a jump at the reference point and more extreme curvature in the region around the reference point. However, a study by Schneider and Lopes (1986) suggests that jumps might not be unique to goals. Schneider and Lopes found risk-seeking in losses involving a chance at breaking even but no risk-seeking when loss gambles were displaced from zero. This pattern is consistent with a value function that is linear in losses with a jump at the reference point of 0. One other question that requires further investigation is whether the two sources of loss aversion that we identify, the jump at the reference point and the difference in slopes above and below the reference point, stem from the same or from different psychological processes.

Our findings do suggest that not all types of reference points operate identically, and also that not all goal reference points operate identically. Specifically, more important goals result in a greater degree of loss aversion, consistent with the findings of Johnson et al. (2006). While other studies have identified moderators of loss aversion, these have typically been qualities of a traded good rather than a characteristic of the reference point itself. Furthermore, we find that multiple reference points can concurrently influence satisfaction with an outcome. Surpassing one's best previous marathon time provides an additional boost in satisfaction above and beyond meeting one's time goal. While earlier studies have also found evidence that multiple reference points can operate simultaneously (Ordóñez et al. 2000), our study does so in a naturalistic manner, without specifically drawing participants' attention to possible alternative reference points. Although our findings document the moderation of loss aversion and the role multiple reference points have on the evaluation of satisfaction, our research also highlights a number of open research questions. For example, what other attributes of a reference point, besides importance, might moderate loss aversion? And do multiple reference points create a feeling of ambivalence, as in Larsen et al. (2004), or do they compete for attention as suggested by Kahneman's (1992) frequency account?

The implications of our study also extend beyond the domain of goals-as-reference-points, to reference dependence more generally. Our use of satisfaction as the dependent measure provides an answer to an important and underexplored question: whether decision utility, the weight given to a outcome in choice, matches experienced utility, or the actual satisfaction felt when that outcome occurs. We find that the predictions participants made about their satisfaction with hypothetical outcomes were more extreme on both the positive and negative side than their actual reported satisfaction, consistent with the literature on affective forecasting. Furthermore, we find that loss aversion was more pronounced in predicted satisfaction than in actual satisfaction. However, in contrast to the findings of Kermer et al. (2006) and in support of Boyce et al. (2013) and Ockenfels et al. (2014), loss aversion is still very much pronounced in experienced satisfaction.

This study is motivated by two more general questions about reference points. First, what psychological processes govern the formation and adaptation of reference points? Second, do non-status quo reference points act similarly to the status quo reference points. This paper provides partial answers to these important questions.

We propose one answer to the first question. Little is known about what psychological processes drive reference point formation and adaptation (for some exceptions, see Arkes et al. 2008 and Baucells et al. 2011). We propose that motivational processes can be implicated in the selection of reference points. While this study provides insight into how goals operate once they have been formed, it is silent about how goals are determined, or how they might change over time. Decision makers who self-set goals face a dilemma. While there is substantial evidence that higher goals lead to higher performance (Locke and Latham 2006), those who perform better, paradoxically, often feel worse because their performance falls below those lofty goals (Mento et al. 1992). Goal formation may result from an attempt to reconcile the self-regulatory value of difficult goals with a desire to avoid disappointment with one's performance. Recent theoretical models in economics such as Hsiaw (2013) have tried to resolve this paradox by studying goals within a dual self framework. We hope that a better understanding of the relationship between goals and satisfaction, as provided in this paper, provides a starting point for determining how goals are updated over time. In other settings, reference points might be determined by other psychological processes, including basic cognitive process such as attention, memory, or categorization, as well as hedonic and motivational considerations.

While all of the psychological processes described above most likely play a role in reference point formation, the second question asks whether the value function that results is mechanism-dependent. Our study suggests that the some properties of the value function such as loss aversion are likely to generalize across a host of psychological antecedents. However, our investigation also suggests that some characteristics of reference dependence, either qualitative like the presence of a jump or quantitative like the degree of loss aversion, might very well depend on how a reference point is determined. We hope that this paper takes a first step toward answering these important questions by providing some conceptual and empirical insights about how non-status quo reference points operate.

References

Abdellaoui, M. 2000. Parameter-free elicitation of utility and probability weighting functions. *Management Science* **46**(11) 1497–1512.

Abdellaoui, M., H. Bleichrodt, C. Paraschiv. 2007. Loss aversion under prospect theory: A parameter-free measurement. *Management Science* **53**(10) 1659–1674.

Abeler, J., A. Falk, L. Goette, D. Huffman. 2011. Reference points and effort provision. *American Economic Review* **101**(2) 470–492.

Agresti, A. 2007. *An Introduction to Categorical Data Analysis*. Wiley, Hoboken, NJ.

Allen, E. J., P. M. Dechow, D. G. Pope, G. Wu. 2014. Reference-dependent preferences: Evidence from marathon runners. Natonal Bureau of Economic Research working paper 20343.

Ariely, D., J. Huber, K. Wertenbroch. 2005. When do losses loom larger than gains? *Journal of Marketing Research* **42**(2) 134–138.

Arkes, H. R., D. Hirshleifer, D. Jiang, S. Lim. 2008. Reference point adaptation: Tests in the domain of security trading. *Organizational Behavior and Human Decision Processes* **105**(1) 67–81.

Austen, I. 2001. You clocked what? for marathon runners, it's gun vs. chip. *The New York Times*, 1 November 2001. Available: http://www.nytimes.com/2001/11/01/technology/you-clocked-what-for-marathon-runners-it-s-gun-vs-chip.html.

Barberis, N., W. Xiong. 2009. What drives the disposition effect? an analysis of a long-standing preference-based explanation. *Journal of Finance* **64**(2) 751–784.

Barberis, N. C. 2013. Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives* **27**(1) 173–195.

Baucells, M., M. Weber, F. Welfens. 2011. Reference-point formation and updating. *Management Science* **57**(3) 506–519.

Bell, D. E., H. Raiffa, A. Tversky. 1988. Descriptive, normative, and prescriptive interactions in decision making. D. E. Bell, R. Raiffa, A. Tversky, eds., *Decision making: Descriptive, normative, and prescriptive interactions*. (Cambridge University Press, Cambridge, England UK), 9–32.

Benartzi, S., R. H. Thaler. 1995. Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics* **110**(1) 73–92.

Bentham, J. 1789. *An Introduction to the Principles of Morals and Legislation*. Clarendon Press, Oxford, UK.

Boyce, C. J., A. M. Wood, J. Banks, A. E. Clark, G. D. Brown. 2013. Money, well-being, and loss aversion does an income loss have a greater effect on well-being than an equivalent income gain? *Psychological Science* **24**(12) 2557–2562.

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

Briesch, R. A., L. Krishnamurthi, T. Mazumdar, S. P. Raj. 1997. A comparative analysis of reference price models. *Journal of Consumer Research* **24**(2) 202–214.

Burson, K., D. Faro, Y. Rottenstreich. 2013. Multiple-unit holdings yield attenuated endowment effects. *Management Science* **59**(3) 545–555.

Camerer, C. 2005. Three cheers – psychological, theoretical, empirical – for loss aversion. *Journal of Marketing Research* **42**(2) 129–133.

Camerer, C., L. Babcock, G. Loewenstein, R. Thaler. 1997. Labor supply of New York City cabdrivers: One day at a time. *Quarterly Journal of Economics* **112**(2) 407–441.

Crawford, V. P., J. Meng. 2011. New York City cab drivers' labor supply revisited: Reference-dependent preferences with rational expectations targets for hours and income. *American Economic Review* **101**(5) 1912–1932.

Curtin, R., S. Presser, E. Singer. 2000. The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly* **64**(4) 413–428.

Dhar, R., K. Wertenbroch. 2000. Consumer choice between hedonic and utilitarian goods. *Journal of Marketing Research* **37**(1) 60–71.

Dhar, R., N. Zhu. 2006. Up close and personal: Investor sophistication and the disposition effect. *Management Science* **52**(5) 726–740.

Diecidue, E., J. Van de Ven. 2008. Aspiration level, probability of success and failure, and expected utility. *International Economic Review* **49**(2) 683–700.

Efron, B., R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY.

Ericson, K. M. M., A. Fuster. 2011. Expectations as endowments: Evidence on reference-dependent preferences from exchange and valuation experiments. *Quarterly Journal of Economics* **126**(4) 1879–1907.

Farber, H. S. 2005. Is tomorrow another day? The labor supply of New York City cab drivers. *Journal of Political Economy* **113**(1).

Farber, H. S. 2008. Reference-dependent preferences and labor supply: The case of New York City taxi drivers. *American Economic Review* **98**(3) 1069–1082.

Fehr, E., L. Goette. 2007. Do workers work more if wages are high? evidence from a randomized field experiment. *American Economic Review* **97**(1) 298–317.

Fryer Jr, R. G., S. D. Levitt, J. List, S. Sadoff. 2012. Enhancing the efficacy of teacher incentives through loss aversion: a field experiment. Natonal Bureau of Economic Research working paper 18237.

Gächter, S., E. J. Johnson, A. Herrmann. 2007. Individual-level loss aversion in riskless and risky choices. Tech. rep., IZA Discussion Papers.

Genesove, D., C. Mayer. 2001. Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics* **116**(4) 1233–1260.

Gilbert, D. T., E. C. Pinel, T. D. Wilson, S. J. Blumberg, T. P. Wheatley. 1998. Immune neglect: a source of durability bias in affective forecasting. *Journal of Personality and Social Psychology* **75**(3) 617–638.

Goldstein, E. R. 2011. The anatomy of influence. *Chronicle of Higher Education* **58**(13) B6–B10.

Gómez-Miñambres, J., B. Corgnet, R. Hernán-Gonzales. 2014. Goal setting and monetary incentives: When large stakes are not enough. Working paper.

Good, P. I. 2005. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd ed. Springer, New York, NY.

Haigh, M. S., J. A. List. 2005. Do professional traders exhibit myopic loss aversion? an experimental analysis. *Journal of Finance* **60**(1) 523–534.

Hastie, R., R. M. Dawes. 2010. *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. SAGE, Thousand Oaks, CA.

Heath, C., S. Huddart, M. Lang. 1999. Psychological factors and stock option exercise. *Quarterly Journal of Economics* **114**(2) 601–627.

Heath, C., R. P. Larrick, G. Wu. 1999. Goals as reference points. *Cognitive Psychology* **38**(1) 79–109.

Higdon, H. 2005. *Marathon: The Ultimate Training Guide*. Rodale, Emmaus, PA.

Ho, T.-H., J. Zhang. 2008. Designing pricing contracts for boundedly rational customers: Does the framing of the fixed fee matter? *Management Science* **54**(4) 686–700.

Hsiaw, A. 2013. Goal-setting and self-control. *Journal of Economic Theory* **148**(2) 601–626.

Johnson, E. J., S. Gächter, A. Herrmann. 2006. Exploring the nature of loss aversion. Tech. rep., IZA Discussion Papers.

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

Kahneman, D. 1999. Objective happiness. D. E. Kahneman, E. E. Diener, N. E. Schwarz, eds., *Well-being: The Foundations of Hedonic Psychology*. (Russell Sage Foundation, New York, NY), 3–25.

Kahneman, D., J. L. Knetsch, R. H. Thaler. 1990. Experimental tests of the endowment effect and the Coase Theorem. *Journal of Political Economy* **98**(6) 1325–1348.

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

Kahneman, D., P. P. Wakker, R. Sarin. 1997. Back to Bentham? Explorations of experienced utility. *Quarterly Journal of Economics* **112**(2) 375–406.

Kermer, D. A., E. Driver-Linn, T. D. Wilson, D. T. Gilbert. 2006. Loss aversion is an affective forecasting error. *Psychological Science* **17**(8) 649–653.

Köbberling, V., P. P. Wakker. 2005. An index of loss aversion. *Journal of Economic Theory* **122**(1) 119–131.

Kőszegi, B., M. Rabin. 2006. A model of reference-dependent preferences. *Quarterly Journal of Economics* **121**(4) 1133–1165.

Larsen, J. T., A. P. McGraw, B. A. Mellers, J. T. Cacioppo. 2004. The agony of victory and thrill of defeat: Mixed emotional reactions to disappointing wins and relieving losses. *Psychological Science* **15**(5) 325–330.

List, J. A. 2003. Does market experience eliminate market anomalies? *Quarterly Journal of Economics* **118**(1) 41–71.

Locke, E. A., G. P. Latham. 2006. New directions in goal-setting theory. *Current Directions in Psychological Science* **15**(5) 265–268.

Long, J. S., J. Freese. 2005. *Regression models for categorical dependent variables using Stata*. StataCorp LP, College Station, TX.

Lopes, L. L. 1987. Between hope and fear: The psychology of risk. *Advances in Experimental Social Psychology* **20** 255–295.

Lopes, L. L., G. C. Oden. 1999. The role of aspiration level in risky choice: A comparison of cumulative prospect theory and SP/A theory. *Journal of Mathematical Psychology* **43**(2) 286–313.

Luce, R., H. Raiffa. 1957. *Games and decisions: Introduction and critical survey*. John Wiley and Sons, New York, NY.

March, J. G., Z. Shapira. 1992. Variable risk preferences and the focus of attention. *Psychological Review* **99**(1) 172–183.

Marsh, L. C., D. R. Cormier. 2001. *Spline regression models*. Sage, Thousand Oak, CA.

Mento, A. J., E. A. Locke, H. J. Klein. 1992. Relationship of goal level to valence and instrumentality. *Journal of Applied Psychology* **77**(4) 395–405.

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

Ockenfels, A., D. Sliwka, P. Werner. Published online: 2014. Bonus payments and reference point violations. *Management Science* URL `http://dx.doi.org/10.1287/mnsc.2014.1949`.

Odean, T. 1998. Are investors reluctant to realize their losses? *Journal of Finance* **53**(5) 1775–1798.

Oettinger, G. S. 1999. An empirical analysis of the daily labor supply of stadium venors. *Journal of Political Economy* **107**(2) 360–392.

Ordóñez, L. D. 1998. The effect of correlation between price and quality on consumer choice. *Organizational Behavior and Human Decision Processes* **75**(3) 258–273.

Ordóñez, L. D., T. Connolly, R. Coughlan. 2000. Multiple reference points in satisfaction and fairness assessment. *Journal of Behavioral Decision Making* **13**(3) 329–344.

Pope, D., U. Simonsohn. 2011. Round numbers as goals evidence from baseball, SAT takers, and the lab. *Psychological Science* **22**(1) 71–79.

Pope, D. G., M. E. Schweitzer. 2011. Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review* **101**(1) 129–157.

Post, T., M. J. van den Assem. 2008. Deal or no deal? Decision making under risk in a large-payoff game show. *American Economic Review* **98**(1) 38–71.

Rabin, M. 2000. Risk aversion and expected-utility theory: A calibration theorem. *Econometrica* **68**(5) 1281–1292.

Sackett, A. M., G. Wu, R. J. White, A. B. Markle. 2014. Harnessing optimism: How eliciting goals improves performance. Working paper.

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

Schneider, S. L., L. L. Lopes. 1986. Reflection in preferences under risk: Who and when may suggest why. *Journal of Experimental Psychology: Human Perception and Performance* **12**(4) 535–548.

Shapira, Z., I. Venezia. 2001. Patterns of behavior of professionally managed and independent investors. *Journal of Banking & Finance* **25**(8) 1573–1587.

Shefrin, H., M. Statman. 1985. The disposition to sell winners too early and ride losers too long: Theory and evidence. *Journal of Finance* **40**(3) 777–790.

Siegel, S., N. J. Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, New York, NY.

Strahilevitz, M. A., G. Loewenstein. 1998. The effect of ownership history on the valuation of objects. *Journal of Consumer Research* **25**(3) 276–289.

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

Tovar, P. 2009. The effects of loss aversion on trade policy: Theory and evidence. *Journal of International Economics* **78**(1) 154–167.

Tversky, A., D. Kahneman. 1991. Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* **106**(4) 1039–1061.

Tversky, A., D. Kahneman. 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* **5**(4) 297–323.

Wakker, P., A. Tversky. 1993. An axiomatization of cumulative prospect theory. *Journal of Risk and Uncertainty* **7**(2) 147–175.

Wang, X. T., J. G. Johnson. 2012. A tri-reference point theory of decision making under risk. *Journal of Experimental Psychology: General* **141**(4) 743–756.

Wilson, T. D., D. T. Gilbert. 2003. Affective forecasting. *Advances in Experimental Social Psychology* **35** 345–411.

Winer, R. S. 1986. A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research* **13**(2) 250–256.

Wu, G., A. B. Markle. 2008. An empirical test of gain-loss separability in prospect theory. *Management Science* **54**(7) 1322–1335.