Title: "*When Falling Just Short is a Good Thing: the Effect of Past Performance on Improvement.*"

Mariya Burdina (corresponding author)
Associate Professor of Economics
University of Central Oklahoma
100 North University Dr.
Edmond, OK 73034
(405)974-2829
mburdina@uco.edu

Scott Hiller
Associate Professor of Economics
Fairfield University
1073 North Benson Rd., Fairfield, CT, 06824.
(203)254-4000, ext. 2795.
rhiller@fairfield.edu,

Abstract: Models of reference-dependent preferences show that an individual's utility depends on the difference between the outcome and a 'neutral' reference point. Our paper investigates how distance from reference points affects future performance. We find that round numbers and personal bests motivate runners, and that missing the goal by a small amount improves future performance. For those who achieve their goal, future performance suffers slightly. In empirical analysis, we use an extensive panel of marathon data, which contains a past running history for every runner in our sample and allows us to estimate runner's ability and experience.

Keywords: reference dependence; round numbers; performance; marathons

JEL Codes: D84, J22, L83

1. INTRODUCTION

Prospect theory assumes that decision makers adopt reference points and evaluate outcomes relative to those important comparisons (Kahneman and Tverskey 1979). The outcomes that fall short of reference points are treated as losses, and outcomes that exceed these reference points are then considered gains. Tversky and Kahneman (1991; 1992) present a formal analysis of the effect of reference-dependence on effort and discuss the change in preferences based on realized gains or losses. If these reference points are relevant to decision-makers two additional questions become important. First, in order to predict behavior one must know which reference points are used. Second, how do realized gains and losses affect future behavior?

In order to determine 'gains' and 'losses' the reference points against which these changes will be calculated must first be clear. Barberis (2013) points out the challenges associated with determining reference points. In many settings, reference points are unobservable or can't be easily identified (Crawford and Meng 2011; Fehr and Goette 2007). Koszegi and Rabin (2006; 2007; and 2009) examine expectations as reference points while Heath, Larrick, and Wu (1999) provide a formal theory that explains the goals acting as reference points. Wu, Heath, and Larrick (2008) show that more challenging goals lead to improved performance. They also show that more challenging goals result in greater performance variance compared to less challenging goals. In their paper mere goals (those that are not set with the potential for reward if achieved) exhibit the properties of the value function, i.e. goals serve as reference points, there is a loss aversion in achieving the goal, and diminishing sensitivity.

The decision maker can also be faced with more than one reference point when making decisions (Boles and Messick 1995; Huang and Tzeng 2007). For example, a worker evaluating a bonus can compare it to bonuses received in the past, any expected bonus, bonuses received by others (Kahneman 1992; Koszegi and Rabin 2006; Card, Mas, Moretti, and Saez 2012), or as discussed in Allen et al (2015) a sum of money equivalent to some arbitrary round number. Similarly, an investor could be evaluating gains and losses against overall wealth, against the value of the overall stock market, initial purchase price of an asset, its peak value, or against the value of some specific stocks (Odean 1997; Odean 1998; Barber 2007; Baker, Pan, and Wurgler 2012).

The previous empirical literature investigates determinants of reference points by examining how individuals respond to supposed ‘gains’ and ‘losses’. For example, Camerer, Babcock, Loewenstein, and Thaler (1997) used cab-driver earnings to illustrate the labor effects of reference-dependent workers. The authors examine how cab drivers respond to an increase in daily wages. The results show that as daily wages decrease, the number of hours worked increases. This finding rejects the standard model of the labor supply and provides the support for the reference dependent model in which worker’s daily earnings serve as a reference point. Grant (2016) examined efforts of ultra-marathon runners during a 100-mile race to illustrate the importance of threshold incentives. The paper uses within-race results to illustrate that runners who are not threshold motivated may diminish their effort and underperform. Anderson and Green (2018) examined the effect of personal bests on effort exerted by players. In their analysis, the authors developed a theoretical model and tested the predictions using a large data set of more than 133 million online chess game matches. The authors found that the player’s personal best rating acts as a reference point. Allen, DeChow, Pope, and Wu (2015) present evidence for round numbers serving as reference points for marathon runners, while Pope and Simonsohn (2011) discuss the effect of round numbers for baseball players and high school students taking SAT.

We employ similar ideas and examine how a runner's performance during a marathon affects her performance in the future. Literature on reference-dependent preferences has mostly dealt with adjustments in preferences based on realized gains and losses during the particular event. We take this idea a step further and test whether current performance and realized gains and losses will have an effect on future performance. This distinction is important as in many settings the tasks are repetitive in nature, thus it is not only important to know how the worker will continue to perform during the current task once the outcomes are realized, but also how achieving current goals will affect their future performance. For example, if the worker is approaching a sales goal, the goal-setting literature predicts improvement in performance. For workers who realize the goal is unachievable the opposite result is expected.

Our paper examines how achieving the goal will affect the worker's effort during subsequent tasks. In our setting, a runner who has narrowly passed the 4-hour finish time mark may perform differently during their next marathon than the runner who was not yet able to achieve that goal. The personal record of a runner could serve as an additional plausible reference point in this scenario. If a runner is motivated by gains and losses, surpassing a previous personal best may serve as a gain, while running a slower race than something previously achieved could certainly be viewed as a loss.

Using a sample of marathon runners polled on their goal, Markle et al (2018) study their satisfaction from a race when comparing a result against their stated goal. We use marathon data with a large number of individuals' marathon finish times to examine similar question. Our data set contains the past public marathon history for every runner in our sample, so we are able to examine the importance of two goals, a 'round number' finish time and a runner's personal record (PR) in serving as reference points for future races. From those results we can discern what effect achieving or not achieving these goals may have on future performance.

Marathon running holds several advantages in testing reference points. First, we are able to access repeated individual data on a large scale, allowing us to measure change clearly and use controls for ability and potential for improvement. Second, the reference point for any runner is very concrete. Races have clear outcomes that allow for precise and realistic goals. There is little confusion as to improvement and regression in results, and if runners are affected by reference points they will have a clear understanding of that reference point and be able to train for the event. Marathons, therefore, may provide the ideal setting for testing the theory.

We test the dual hypotheses that round numbers and personal records may not only serve as important reference points, but also influence the behavior of a runner. If round numbers are influential, we would expect the runner who has narrowly missed the 4-hour mark to improve more than the runner who has finished a previous race just under that time, all else equal. If the personal record is important as a reference point, we might expect a runner who has not achieved a personal record in the previous race to work harder to improve more than a similar runner who set a new personal best. Our dataset allows us to examine the importance of these potential reference points in absolute and relative terms.

2. DATA AND SUMMARY STATISTICS

The data used in this paper are the same as used in Burdina et. al (2017). The data come from publicly available marathon finish times for many marathons over the last 45 years. This unique dataset was obtained by taking the finishers of three marathons: Oklahoma City Memorial Marathon (2001–2014), California International Marathon (1990–2013), and Grandma's Marathon (2001–2014). These races were chosen as a starting point for our dataset from which we use the names of the runners to obtain their complete racing histories from 1970 to 2015 through Athlinks.com, a repository of race data.

In order to observe the effects of important reference points we have to capture what happens to the change in results given proximity to a reference point. We use two dependent variables. The first is the difference in marathon finishing time for a runner *i* from race *j-1* to race *j*, where *j-1* and *j* are sequential in the finisher's racing career. As an example, if runner *i* finishes race *j-1* in 245 minutes, and race *j* in 240, then the difference in time would be -5. A negative value indicates an improvement in time, and therefore positive coefficients on independent variables in regression results indicate the corresponding variables correlated with a slower finishing time, and a negative coefficient shows the variable to be correlated with a faster time.

Our main variables of interest are indicators for runners finishing the previous race five minutes above and below the reference points. The runners that are above the reference point may see the failure to achieve it as a goal for the next race, where the runner who has reached the reference point but remains within the same distance of it may wish to find a new goal for the upcoming race. If these reference points are a substantial motivator, we would expect to see a greater improvement in race times from those runners narrowly above than below each point.

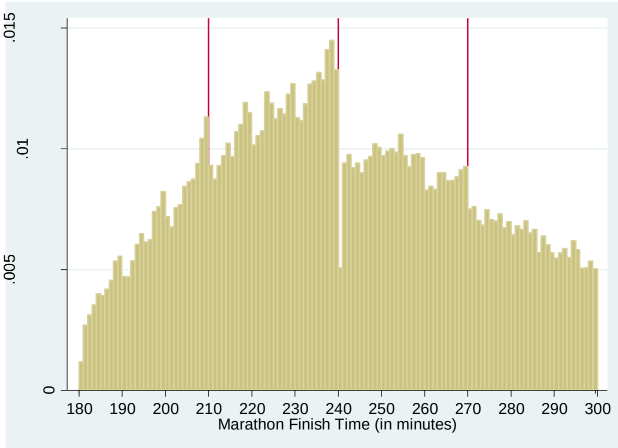
We test if round numbers along with runner's PR serve as reference points. PR is the best finish time (in minutes) of a runner prior to the race in question. If PR is a motivator, the fact that the runner did not achieve a new personal record in the previous race may act as motivation to improve in the current race. Similarly, if a round number is a motivator as well the runners who finish the marathon slightly slower than a certain round number, may be extra motivated to achieve such a goal during the next race.

Allen et. al. (2015) tested for reference dependence in a marathon setting and provided evidence that round numbers serve as reference points, with bunching occurring at these round numbers. For example, there were significantly more people who finished right before 3:00, 3:30, and 4:00, hours than those who finished right after. An examination of our data indicates that bunching occurs at these key round numbers, even for our experienced runners (Figure 1.a). We extend the analysis by using a unique dataset and regression discontinuity to test the effects of reference points over repeated marathons.

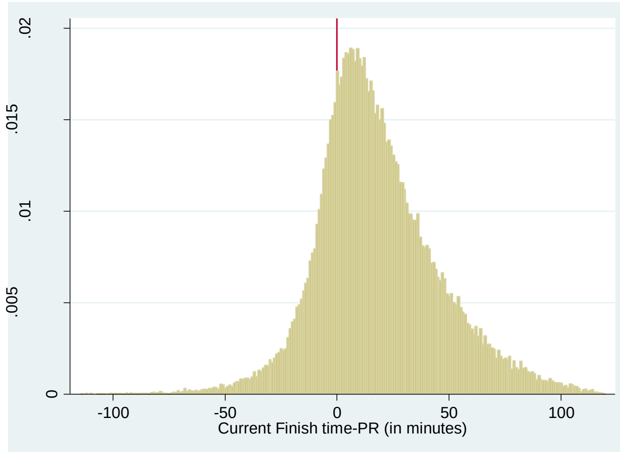
Additionally, the bunching around PR is not as obvious. When looking at the histogram of the difference between current finish time and a current PR (Figure 1.b), we do not observe bunching similar to that occurring at the 4-hour mark. One of the explanations for such a difference is that obtaining a new PR may be more difficult than finishing a marathon under 4-hours, especially for those who have already achieved the goal of finishing under 4-hours. It is also possible that obtaining a PR will be an impossible task for those who have already achieved their maximum physical potential. A runner may not realize such limitations and still aim to set a PR yet not be able to achieve it. This could be especially true for those who began their running career at an earlier age and thus set a PR that could be impossible to beat at an older age. Given this, we cannot rely on bunching in order to test whether PR serves as a reference point and influences future performance. Instead, we rely on regression discontinuity analysis to illustrate importance of PR along with the round numbers when it comes to setting the goals by marathon runners.

FIGURE 1

Histogram of finish times and difference between finish time and current PR.



a. Frequency of marathon finish times in minutes



b. Frequency of the difference between current finish time and current PR

The racing histories are key to our study as they provide a measure of runner ability by observing their full racing history. Many of the important variables require measuring the change from race to race, meaning the first two races of any individual in our sample are necessarily excluded for construction. In order to ensure that we can control for appropriate characteristics, and more importantly to ensure that the runners in our sample are committed to the task we drop any individual in our sample who has finished less than five marathons.[1] Less persistent runners are less likely to be motivated by the same goals, as the achievement of simply finishing may be enough for them.

When running marathons, there is no single formula that will determine how much one can improve. Bar-Eli et al. (1997) used 10%, 20%, 40% improvement targets as a measure of goal difficulty in short distance running, but in a marathon, a 10% improvement is a difficult goal to attain (i.e. a 10% improvement from the most common finish time of 4 hours would be

[1] Five may seem an arbitrary number, but it allows for a large sample while still ensuring that we do not include casual participants. We tried the same regressions with both less than four and six marathons excluded with qualitatively similar results.

24 minutes). A runner close to a reference point may be more motivated than one further away.[2] Additionally, some runners may be at their physical limit in their ability to improve. Our panel data allows us to control for this possibility. We do so in several ways. First, we include overall personal record in the analysis in order to control for a runner's maximum potential. This record could occur either in the past or in the future, depending on which race we are observing. This variable is a control for physical ability and is another reason we exclude any runner with less than five completed marathons as we need a larger sample to be more confident we are properly controlling for ability.

Second, we construct a variable that represents the distance from a runner's previous marathon finish time and her potential goal time (either a round number or a personal record). This variable should help us to control for the `ease' of improving. For example, it could be easier to improve for someone who finishes a marathon in 4 hours and 4 minutes rather than 3 hours and 59 minutes as they have 5 minutes of cushion. As discussed in more detail below, the improvement may not be linearly related to the finish time, for that reason we use natural logarithm of difference from previous race to goal to control for a constant percentage effect.

The other independent variables include a runner's characteristics such as age, gender, and experience. In order to control for a runner's experience, we use the number of marathon and non-marathon races completed by the runner in the past. We use the number of races in the given year as a proxy measuring a runner's preparedness for the marathon. Training for a marathon requires a serious commitment from a runner, but we have no way of observing if the runner was actually seriously training (i.e. completing their scheduled runs). Those who participate in

[2] In testing regressions, we found nonlinearities in the effects of reference points, leading us to use the regression discontinuity approach documented below.

numerous races may be more likely to complete their training plan for the marathon. Additionally, we include a month indicator for the month the race takes place in.

The initial dataset of all runners, from all years and all races contains more than 5.5 million observed race finish times, but this includes many races which are not marathons. When cleared of all races but marathons and narrowing the finish times to between three and five hours with five marathon finish times for each runner, the final sample contains 88,206 observations.[3] However, given the need to establish an initial race time and the set up for the model, the number of observations in each regression is lower, and the individual samples of our regression discontinuity approach (bunched around round numbers) lower the observations still. Variable definitions are presented in Table 1. Summary statistics are included in Table 2

TABLE 1

Variable Description

Variable	**Description**
Time Difference	Time difference (in minutes) between current and previous marathon
Improved	=1 if the runner improved compared to the previous marathon
Above 4-hours	=1 if the runner's finish time was above 4 hours by 5 minutes
Above 3.5 hours	=1 if the runner's finish time was above 3.5 hours by 5 minutes
Above 4.5 hours	=1 if the runner's finish time was above 4.5 hours by 5 minutes
Above PR	=1 if the runner's finish time was above PR time by either 5 minutes
New PR	=1 if the runner achieved a new personal record during the previous race
Running PR	Runners overall personal record, in minutes
Age	Participant's age
Ln(Previous Time - 205)	Natural log of difference between previous finish time and a goal of 3.5 hours
Ln(Previous Time - 235)	Natural log of difference between previous finish time and a goal of 4 hours
Ln(Previous Time - 265)	Natural log of difference between previous finish time and a goal of 4.5 hours
Ln(Previous PR diff + 5**)**	Natural log of difference between previous finish time and PR
Age Squared	Participant's age squared
Gender	=1 if female

[3] We limit the finish times from 3 to 5 hours to remove any participants that are not serious enough, or perhaps too serious for reference points to play a big role. We try to use a middle ground of serious but not professional runners. We also test the same regressions using 2.5 and 7 hours as our cutoff with qualitatively similar results.

Experience	Number of races participant completed in career prior to current marathon
Experience squared	Experience squared
Races in a year	Number of races participant competed in during particular year
Marathons	Number of marathons completed prior to the race
Months since last marathon	Number of month past from the previous marathon

TABLE 2

Summary Statistics

Variable	Observations	Mean	Std. Dev	Min	Max
Improved	88,075	0.495282	0.499981	0	1
Time Difference	88,206	0.5205	25.15089	-118.417	116.1
Above 5 min of PR	88,206	0.078611	0.269133	0	1
Below 5 min of PR	88,206	0.066061	0.248390	0	1
New PR	88,206	0.207492	0.405513	0	1
Above 5 min of 3.5 hours	88,206	0.047945	0.213655	0	1
Below 5 min of 3.5 hours	88,206	0.049055	0.215985	0	1
Above 5 min of 4 hours	88,206	0.043886	0.204842	0	1
Below 5 min of 4 hours	88,206	0.068544	0.252678	0	1
Above 5 min of 4.5 hours	88,206	0.036052	0.186421	0	1
Below 5 min of 4.5 hours	88,206	0.044520	0.206250	0	1
Above 5 min of a Round Number	88,206	0.127883	0.333961	0	1
Ln(Previous Time -205)	75,836	3.422387	0.971317	-4.09459	4.553877
Ln(Previous Time – 235)	46,810	2.933082	1.113482	-4.09459	4.174387
Ln(Previous Time – 265)	19,984	2.383555	1.06891	-4.09459	3.555348
Running PR	88,206	212.2418	22.34356	180	292.6
Months since last marathon	88,206	11.60294	17.09226	0	276
Age Squared	88,206	2116.406	914.0281	289	6724
Marathons	88,206	8.88901	9.595837	2	130
Experience	88,206	40.34308	43.61386	2	377
Experience squared	88,206	3529.711	7874.89	4	142129
Races in a year	88,206	7.754314	6.808564	1	30
Age	88,206	44.90319	10.00552	17	82
Female	86,232	0.298926	0.45779	0	1

The average time difference between the marathon finishes is 0.52 minutes, which means that on average runners slowed from race to race. This is somewhat expected as it is easier to run slower for any particular race than to run it faster. About 50% of all races were finished faster than the previous result. We also observe there are more finishes below the 4-hour mark than above, results consistent with Allen et. al. (2015). Finally, the number of finishes above the PR time is slightly more than the number of finishes below PR, which is also predictable as getting a PR during the race is more challenging than having a slower finish.

3. EMPIRICAL MODEL

We use two specifications to measure the effect of reference points on performance. Our first dependent variable is the change in time from one race to the next. Given that we are trying to measure improvement, a higher number in race time (slower race) may just mean greater room to improve. If this is true, a runner may be able to see a greater improvement simply because they were slower before. We do three things to mitigate this potential bias. First, we employ regression discontinuities, discussed below, to help ensure we are comparing similar runners. Second, we include a variable controlling for slower times within the sample to ensure the effects we find are not due to room for improvement. Finally, using the same regression discontinuity approach we run a random effects logit with a simple dependent variable for whether a runner improved their time from the previous race as the dependent variable. The coefficients in this simpler specification provide a change in probability of improving.

We estimate the effect of proximity to an important reference point using the model of the following general form:

$$TimeDiff_{ijt} = \alpha_0 + \alpha_1 Above\_Round\_Number_{ij-1} + \alpha_2 NewPR_{it-1} + \alpha_3 Gender_i + \alpha_4 Age_{it} + \alpha_5 Age\_Sq_{it} + \alpha_6 Experience_{it} + \alpha_7 Experience\_Sq_{it} + \alpha_8 Races\_in\_Year_i + \alpha_9 TotalPR_{it} + \alpha_{10} LN(Previoustime - Min)_{it} + \alpha_{11} Marathons_i + \alpha_{12} MonthSince_i + \alpha_{13} \sum Month_{ij} + \epsilon_{ijt}$$

Where $TimeDiff_{ijt}$ is the difference in athlete *i*'s finish time in year *t* at marathon *j* from their time in the previous marathon *j-1*, where a positive number indicates a slower time than the previous result. $Above\_Round\_Number_{ij-1}$ is an indicator variable equal to one if runner *i* was above the 3.5, 4, or 4.5-hour[4] finish times by 5 minutes.[5] *NewPR* is a variable indicating whether a runner achieved a new PR in her previous race.

Age_{it} and $Gender_i$ are athlete *i*'s corresponding age and gender (equal to 1 if female, and 0 if male) during the year *t*. We expect Age_{it} variable to be positively related to the change in marathon finish time on average, as older runners have slower finish times compared to younger runners. The effect of $Gender_i$ is not clear as while female runners are usually slower than male runners, the ability to improve may not be related to gender. The $Experience_{it}$ term counts the total number of races the runner has completed before race *j* in our dataset. This includes marathons and other shorter races. The variable $Experience\_squared_{it}$ is the square of $Experience_{it}$. The $Marathons_{ij}$ variable accounts for number of marathons completed before the marathon *j*. The variable $Races\_in\_year_{it}$ provides the total races the runner completes in year *t*. The $TotalPR_{it}$ variable is intended to capture the racer's maximum potential

[4] We also try values of 2 and 10 minutes with similar results. The 2 minute specification lacks the significance of the 5 minute mark, while the 10 minute mark shows an even greater likelihood of improvement than 5. The 2 minutes specification is likely insignificant because of a small sample size.
[5] Each of the three round number periods are estimated separately.

and is equal to the fastest marathon finish time during her visible career. The $MonthSince_{it}$ is a variable that shows how many months have passed since the previous marathon. More serious runners are most likely to have a shorter break between the marathons while more casual runners are more likely to have longer breaks between the marathons. $\sum Month_{ij}$ is a set of dummy variables indicating in which month runner *i* participated in marathon *j* used to control for seasonal differences in races as temperature on race day can play an important role in finish time.

The error term is assumed to be normally distributed, and estimation is performed with a random-effects least squares regression. Several models are estimated with the independent variables slightly altered as described above. The greater distance from a reference point for a runner in their previous race, the less of an impact we expect for that reference point to have on the current race. Given the potential non-linear effects in distance from the round number (or PR) time, we test those runners near the potential reference points. This analysis lends itself to regression discontinuity. Using the entire sample for each regression would allow for many more observations, but at the risk of not comparing runners of similar ability and motivation.

As discussed above, we use three strategies to control for the potential improvement benefit of having a slower finish time. First, we employ regression discontinuities to test runners on each immediate side of a reference point. Second, we include a variable that helps to control for the fact that runners above a round number simply have more room to improve compared to those below. The variable, $Ln(Previoustime\text{-}Min)_{it}$ takes the natural log of the difference between the previous finish time and the bottom value of the regression discontinuity range.[6] So for

[6] We tested other possible variables that control for the cushion provided for a slower time, but most had a problem passing the falsification tests discussed below. This variable was generally significant and passed falsification tests.

runners within 5 minutes of 3:30, the bottom of the regression discontinuity range would be 205 minutes (finish time of 3 hours and 25 minutes), for the runners within 5 minutes of 4 hours, the bottom of the range would be 235 minutes (3 hours and 55 minutes); for runners within 5 minutes of 4 hours and 30 minutes time it would be 265 minutes (4 hours and 25 minutes).[7] This ensures that if we sufficiently control for individual characteristics, the comparison is between similar runners. The differences in time improvement from one race to the next can then be plausibly assigned to a motivation from one of the reference points. Third, we use a specification simply measuring whether a runner improved her time from the last race. The dependent variable $Improved_{ijt}$ is an indicator that takes the value of one if the time in race j is faster than race $j-1$. We use this specification to consider the possibility that the improvement we are observing is biased by the results above the reference point providing more room for improvement. With this dependent variable the threshold for success is only improving, a much lower bar. The specification is similar and of the general form:

$$
\begin{aligned}
Improved_{ijt} = {} & \alpha_0 + \alpha_1 Above\_Round\_Number_{ij-1} + \alpha_2 NewPR_{it-1} + \alpha_3 Gender_i + \alpha_4 Age_{it} \\
& + \alpha_5 Age\_Sq_{it} + \alpha_6 Experience_{it} + \alpha_7 Experience\_Sq_{it} + \alpha_8 Races\_in\_Year_i \\
& + \alpha_9 TotalPR_{it} + \alpha_{10} LN(Previoustime - Min)_{it} + \alpha_{11} Marathons_i \\
& + \alpha_{12} MonthSince_i + \alpha_{13} \sum Month_{ij} + \epsilon_{ijt}
\end{aligned}
$$

This specification is estimated with a conditional random effects logit, and the explanatory variables have the same interpretation as above. Coefficients in this specification explain a change in the probability of improvement from a change in the independent variable.

[7] We use a log specification to maintain a constant percentage effect, as any time difference in improvement is very unlikely to be linear. The 4 hour 3 minute finish time (243 minutes) would provide a value of LN(8) for this variable, as we would take LN(243-235).

4. RESULTS

We focus first on the pure impact of round numbers. In Table 3 we use a sample of runners who finished a marathon within 5 minutes of the 3.5, 4, and 4.5- hour mark. The dependent variable is the difference in time from the previous to current race. In columns one through three we exclude any PR effects, and in columns four through six we include an indicator for whether a runner achieved a new PR in her previous race. A negative coefficient indicates a variable correlated with an improvement from the previous race. In all regressions month indicators are included, but excluded from the tables for brevity. The results indicate a runner who finishes five minutes above the 3.5, 4, and 4.5-hour marks improves more in the next race than a runner in the same time frame below that mark. The approximately 2.5 minute improvement is consistent and significant across all specifications, and shows a large effect of a nearby round number all else equal.

Columns four to six provide the results for the analysis using achievement of a new PR in the previous race. The idea is that a runner may lose some motivating effect if the round number is swamped by a recent personal accomplishment. We find that reaching a new personal record seems to make improvement more likely at the three and a half hour mark, less likely at the four hour mark, and insignificant at the four and a half hour marks. Perhaps, more serious runners that can achieve a three-and-a-half-hour race are less likely to lose motivation by achieving a new PR. Most importantly, the effect of a round number is not lessened by the inclusion of the *NewPR* variable, showing the same strong effect in improvement as was shown in the difference of time.

The results for the control variables are as expected and are a similar sign, magnitude and statistical significance as in Burdina, et. al. (2017). The variable *Ln(Previoustime-Min)* is negative and significant, indicating the predicted outcome of improvement being larger with a slower time. Those with a slower finish time have more room to improve and improve more than those with a faster finish times, and thus we are assured that the effect of the '*Above RN*' variables comes from desire to achieve the goal, and not the ease we could attribute to a cushion. The effect of experience is positive, and this result is most likely due to the fact that less serious runners tend to run more shorter races (e.g. 5Ks), thus running more races over one's career may indicate the runner is not a serious marathon runner. The square of the experience term is negative which indicates that at some threshold finish time improves with experience.

Races in a year is our proxy for training and the effect is negative, indicating that the more races one runs in the year of the marathon in question, the faster their time. This is expected, more races equate to more training which leads to faster times. *Total PR* is the fastest career finish for a runner, a measure we use for the potential of a runner. The slower the overall PR the less the improvement of a runner. Age is negative, indicating that older runners tend to improve more than younger runners do, but the positive square term indicates that this effect lessens quickly.

TABLE 3

Runners within 5 minutes of a round number (time difference)

	(1)	(2)	(3)	(4)	(5)	(6)
	3.5 hours	4 hours	4.5 hours	3.5 hours	4 hours	4.5 hours
Above RN by 5 minutes	-2.662***	-2.503***	-2.604***	-2.737***	-2.484***	-2.588***
	(0.597)	(0.538)	(0.681)	(0.598)	(0.539)	(0.681)
Ln(Prev-Min)	-1.169***	-1.064***	-0.702**	-1.219***	-1.036***	-0.686**
	(0.314)	(0.26)	(0.342)	(0.315)	(0.26)	(0.343)
New PR				-1.268**	0.951*	1.109
				(0.552)	(0.554)	(0.898)
Total PR	0.854***	0.578***	0.477***	0.882***	0.571***	0.474***
	(0.0303)	(0.0155)	(0.0142)	(0.0326)	(0.0161)	(0.0145)

Month since last marathon	0.136***	0.0557***	-0.0368**	0.140***	0.0578***	-0.0368**
	(0.0135)	(0.012)	(0.0163)	(0.0134)	(0.0119)	(0.016)
Experience	0.161***	0.0961***	0.0378*	0.153***	0.0762***	0.0335*
	(0.0201)	(0.0176)	(0.0211)	(0.0163)	(0.0143)	(0.0172)
Experience squared	-0.00047***	-0.00034***	-0.00023***	-0.00045***	-0.00029***	-0.00022***
	(0.000086)	(0.0000771)	(0.0000889)	(0.0000819)	(0.0000724)	(0.000084)
Races in a year	-0.0344	-0.0995**	-0.0269	-0.018	-0.110**	-0.0293
	(0.0565)	(0.0497)	(0.0608)	(0.0567)	(0.0498)	(0.0609)
Age	-0.829***	-0.994***	-0.854***	-0.836***	-0.981***	-0.851***
	(0.206)	(0.161)	(0.193)	(0.206)	(0.161)	(0.193)
Age squared	0.0102***	0.0129***	0.0124***	0.0102***	0.0129***	0.0124***
	(0.00237)	(0.0018)	(0.0021)	(0.00237)	(0.0018)	(0.0021)
Number of Marathons	0.1000**	0.244***	0.325***	0.0974**	0.263***	0.331***
	(0.0498)	(0.0357)	(0.042)	(0.0487)	(0.035)	(0.0413)
Gender (=1 if female)	-1.447*	0.358	2.201***	-1.434*	0.362	2.174***
	(0.764)	(0.554)	(0.678)	(0.763)	(0.554)	(0.678)
Constant	-151.2***	-111.9***	-112.3***	-155.9***	-111.4***	-112.0***
	(7.247)	(4.68)	(5.256)	(7.508)	(4.708)	(5.257)
N	8344	9663	6952	8344	9663	6952
adj. R^2	0.1491	0.1807	0.2219	0.1491	0.1812	0.2224

Standard errors in parentheses. Sample is selected only using runners with a prior race 5 minutes above or 5 minutes below a round number. The dependent variable is the change in a runner's marathon result from the last race. The specification is estimated using OLS with fixed effects for time and runner. * $p < .10$, ** $p < 0.05$, *** $p < .01$

For each specification we also run a series of falsification tests to ensure the results are not driven by something other than a reference point. In these falsification tests we try the same specifications, but with reference points replaced by nearby random times. For example, we try a regression discontinuity comparing finishing a race between 3:52-3:57 to 3:47-3:52 in place of the 4-hour mark. Aside from the false round number variable the exact same specifications were used in these robustness tests.[8] The results for all round number reference points (using both time difference and improvement) were never significant and the signs on the coefficients of

[8] We exclude the results from the paper for space, but tables are available upon request.

interest switched often. The lack of significance for these false round numbers encourages interpretation of the reference point effect we find in our round number results.

Table 4 provides the same specifications with the simple improvement indicator as the dependent variable (=1 if the runner improved during the current race, 0 otherwise). We find that a runner with a previous race above a round number is substantially more likely to improve, all else equal. This result is consistent across the timeframes and highly significant in each instance.[9] This effect is in addition to the room for improvement variable, which also has a positive effect on the likelihood of improvement. This simpler specification shows a clear, consistent effect of round numbers on improvement, all else equal. Taken with falsification tests that do not show a similar effect for nearby times, there is substantial evidence for a round number as reference point effect in improvement. The control variables all show expected signs and plausible magnitudes. The higher the overall PR a runner, the less likely improvement. The more months since a marathon, a variable we think plausibly substitutes for seriousness of training, the less likely improvement. Increased experience makes improvement less likely, but that effect is mitigated by the squared term as the number of races increases. Increased age makes improvement more likely, but again to a point. There is no significant difference between men and women in likelihood of improvement from race to race.

Similar to the findings presented in Table 3, reaching a new personal record seems to make improvement less likely at the four and four and a half hour marks, but not at three and a half. The effect of a round number remains with the inclusion of the *NewPR* variable, showing the same strong effect in improvement as was shown in the previous specification.

[9] As stated above, this is in contrast to the falsification tests where the coefficient of interest is attenuated, changes sign, and is never significant.

TABLE 4

Improvement of runners within 5 minutes of a round number

	3.5 hours	4 hours	4.5 hours	3.5 hours	4 hours	4.5 hours
	(1)	(2)	(3)	(4)	(5)	(6)
Above RN by 5	0.267***	0.206***	0.223***	0.270***	0.202***	0.222***
	(0.0719)	(0.0652)	(0.0776)	(0.0719)	(0.0653)	(0.0776)
Ln(Prev-Min)	0.126***	0.0961***	0.0409	0.128***	0.0925***	0.0376
	(0.0378)	(0.0316)	(0.0388)	(0.0378)	(0.0317)	(0.0388)
New PR				0.102	-0.263***	-0.231**
				(0.065)	(0.0661)	(0.0994)
Total PR	-0.0989***	-0.0494***	-0.0294***	-0.101***	-0.0475***	-0.0287***
	(0.00357)	(0.00195)	(0.00156)	(0.00375)	(0.00199)	(0.00159)
Month since last marathon	-0.0131***	-0.00452***	-0.00315*	-0.0134***	-0.00389***	-0.00284
	(0.00181)	(0.00143)	(0.00183)	(0.00183)	(0.00144)	(0.00183)
Experience	-0.00831***	-0.0105***	-0.00646***	-0.00810***	-0.0110***	-0.00686***
	(0.00222)	(0.00204)	(0.00225)	(0.00222)	(0.00205)	(0.00226)
Experience squared	2.52E-05***	3.76E-05***	2.88 E-05***	2.46 E-05**	3.98 E-05***	3.03 E-05***
	(9.6E-06)	(9.14E-06)	(9.84E-06)	(9.59E-06)	(9.19E-06)	(9.87E-06)
Races in a year	0.00301	0.0141**	-0.00569	0.00204	0.0159***	-0.0047
	(0.00636)	(0.00567)	(0.00649)	(0.00638)	(0.0057)	(0.00651)
Age	0.110***	0.0719***	0.0361*	0.111***	0.0683***	0.0358*
	(0.0223)	(0.0179)	(0.0201)	(0.0223)	(0.018)	(0.0201)
Age squared	-0.00146***	-0.000998***	(0.000215)	-0.00146***	-0.000972***	-0.000620***
	(0.000259)	(0.000199)		(0.000259)	(0.0002)	(0.000215)
Number of Marathons	-0.0317***	-0.0201***	-0.0164***	-0.0311***	-0.0209***	-0.0168***
	(0.00516)	(0.00387)	(0.00404)	(0.00516)	(0.00389)	(0.00404)
				(0.00516)		
Gender (=1 if female)	0.121	-0.0213	-0.0664	0.121	-0.0185	-0.0582
	(0.0738)	(0.0573)	(0.0656)	(0.0737)	(0.0574)	(0.0658)
Constant	17.98***	10.25***	7.519***	18.26***	10.01***	7.409***
	(0.805)	(0.548)	(0.564)	(0.826)	(0.551)	(0.566)
N	8327	9649	6942	8327	9649	6942

Standard errors in parentheses. Sample is selected only using runners with a prior race 5 minutes above or 5 minutes below a round number. The dependent variable is an indicator for whether a runner improved her result from the last race. The regression is estimated with a conditional random effects logit. Coefficients represent the marginal difference in probability of improving associated with a one unit change in the independent variable. R^2 for the logit model. * $p < .10$, ** $p < 0.05$, *** $p < .01$

Finally, we use a pooled sample of all runners within five minutes of a round number, regardless of which of the results the previous finish was closest to. We do not use this as a primary method of estimation as it merges dissimilar runners, but we find it to be a useful check on the broader results using a larger sample size. Estimates are generally similar with the exception of the variable controlling for log of improvement, which while similar in value to the previous results, is never significant. The estimates of improvement and time difference are attenuated from the more specific regressions, but still present a time difference of 1.5 minutes, and an increase in the likelihood of improvement of nine percent. Additionally, we ran the same specification on runners within two and ten minutes of round numbers. Results for the two minutes grouping were not significant for the improvement indicator, but showed a reduction in time of 1.7 minutes which was significant at the one percent level. The ten minutes grouping was significant for both the improvement and time difference variables, and showed a slightly more substantial impact for each than those in Table 5. All specifications passed the same robustness tests used in the models above.

TABLE 5

Pooled results for runners within 5 minutes of a round number

	Time Difference	Time Difference	Improvement	Improvement
Above RN by 5	-1.584***	-1.540***	0.0886**	0.0912**
	(0.403)	(0.397)	(0.0357)	(0.0363)
Ln(Prev-Min)	-0.0621	0.0171	0.0121	0.00747
	(0.203)	(0.200)	(0.0180)	(0.0183)
New PR		10.55***		-0.752***
		(0.371)		(0.0351)
Total PR	0.139***	0.118***	-0.0132***	-0.0121***
	(0.00801)	(0.00810)	(0.000696)	(0.000719)
Months since last marathon	0.0630***	0.0388***	-0.00583***	-0.00423***
	(0.00899)	(0.00891)	(0.000814)	(0.000821)

Experience	0.0673***	0.0922***	-0.00563***	-0.00743***
	(0.0119)	(0.0119)	(0.00104)	(0.00107)
Experience squared	-0.000272***	-0.000359***	0.0000226***	0.0000289***
	(5.24 e+05)	(0.0000523)	(0.00000461)	(0.00000473)
Races in a year	-0.123***	-0.185***	0.0104***	0.0148***
	(0.0345)	(0.0344)	(0.00300)	(0.00306)
Age	0.0181	0.0284	0.00372	0.00128
	(0.106)	(0.106)	(0.00924)	(0.00947)
Age squared	0.00127	0.00169	-0.000196*	-0.000209**
	(0.00117)	(0.00117)	(0.000102)	(0.000105)
Number of Marathons	0.0513**	0.107***	-0.00509***	-0.00891***
	(0.0228)	(0.0234)	(0.00187)	(0.00197)
Gender (=1 if females	-0.601*	-0.720**	0.0396	0.0477
	(0.359)	(0.362)	(0.0307)	(0.0314)
Constant	-36.49***	-36.09***	3.511***	3.616***
	(2.863)	(2.871)	(0.251)	(0.257)
N	24952	24952	24914	24914

Standard errors in parentheses. Sample is selected only using runners with a prior race 5 minutes above or 5 minutes below a round number. The dependent variable is an indicator for a time difference in columns 1 and 2, and whether a runner improved her result from the last race in columns 3 and 4. Coefficients represent the marginal difference in probability of improving associated with a one unit change in the independent variable. R^2 for the logit model. * $p <$ .10, ** $p < 0.05$, *** $p < .01$

In addition to round numbers, we also perform a similar test on the effect of proximity to PR. Table 6 includes any runner whose finish time was near (within 5 minutes) her PR in the previous marathon. This analysis now includes PR times from 3 to 5 hours instead of limiting the PR dummy variables to cases where the PR was near the 'round number' finish time in the previous tables. The specifications are similar to those already shown, with the exception of the PR variables. Time difference from previous to current race is the dependent variable in columns 1 and 2, where coefficients have the same interpretation as Table 3 of a negative valuable as improvement in time. An improvement indicator is the dependent variable for columns 3 and 4, where coefficients have the same interpretation as Table 4 of a positive value increasing the probability of improvement.

Results indicate those runners with times above their PR in the previous race tend to both improve more and, unsurprisingly, are more likely to improve than those who bested their PR in the previous marathon, all else equal. The effect is smaller than the round number effect, but still large relative to time, additionally they are consistent and significant. The sample is constructed around the PR variable, but some results still fall near a round number effect. The round number effect for this sample remains significant in all but one specification. In specifications (1) and (3) we add a variable for those that are not within five minutes of any round number, in order to compare the effect for those who achieved the round number goal in the previous race and those who did not achieve that goal. The results indicate that those within 5 minutes of their PR improved more if they have not achieved a 'round number' finish time in the previous race compared to those who achieved that goal. This effect is only significant in the specification with 'time difference' as a dependent variable and is insignificant for the 'improvement' specification. In specifications (2) and (4), we include an interaction variable equal to the product of Above 5 PR and Above5 RN variables. This variable shows a similar effect to specifications (1) and (3) with the additional of a round number for those who did not achieve a PR in the previous race. The results indicate that those who finished five minutes above a round number were more likely to improve, but that effect lessens for those who also were able to achieve PR during the previous race. This effect is significant for both specifications. The PR effect is not eliminated by its inclusion, indicating that both reference points are likely important motivators for improvement. A direct comparison of importance is not possible, but the round number effect seems to be more consistent.

To control for the room to improve we use a different variable from the round number effect. Here, the natural log of the previous improvement (time difference between previous marathon finish time and PR at that time) serves to control for how much additional improvement is possible. Because of sample construction previous improvement can range from -5 to 5, not inclusive. In order to take the log of the value, we normalize this range from 0 to 10. We use this variable to allow for a constant percentage effect on the improvement of results. A one percent increase in this variable does not have a large impact on time difference or probability of improvement, but is consistently significant, and given the results of falsification tests performs well at controlling for improvement.

TABLE 6

Runners within 5 minutes of PR in previous marathon

	Time difference (1)	Time difference (2)	Improvement (3)	Improvement (4)
Above 5 min of PR	-1.507*** (0.500)	-1.768*** (0.516)	0.173*** (0.0583)	0.219*** (0.0602)
Above 5 min of RN	-1.842*** (0.695)	-3.481*** (0.836)	0.0525 (0.0764)	0.320*** (0.0910)
Above5 PR*Above5 RN		2.283** (1.125)		-0.399*** (0.124)
Not within 5 min of RN	0.550 (0.519)		-0.0637 (0.0552)	
Ln(Previous PR + 5)	-0.710*** (0.271)	-0.713*** (0.271)	0.0821*** (0.0318)	0.0819*** (0.0318)
Total PR	-0.00847 (0.0105)	-0.00938 (0.0105)	-0.00123 (0.000944)	-0.00112 (0.000938)
Months since last marathon	0.0710*** (0.0108)	0.0712*** (0.0108)	-0.00535*** (0.00124)	-0.00538*** (0.00124)
Age	0.165 (0.174)	0.165 (0.174)	0.0227 (0.0172)	0.0226 (0.0172)
Age squared	-0.00102 (0.00197)	-0.00104 (0.00197)	-0.000460** (0.000197)	-0.000457** (0.000197)
Marathons	0.804*** (0.0562)	0.805*** (0.0562)	-0.113*** (0.00750)	-0.112*** (0.00749)
Experience	0.129*** (0.0185)	0.129*** (0.0185)	-0.00761*** (0.00199)	-0.00762*** (0.00199)

Experience squared	-0.000391***	-0.000390***	0.0000266**	0.0000264**
	(0.0000901)	(0.0000901)	(0.0000104)	(0.0000104)
Races in a year	-0.212***	-0.212***	0.0304***	0.0305***
	(0.0477)	(0.0477)	(0.00512)	(0.00511)
Gender (=1 if female)	-2.988***	-2.979***	0.110**	0.108**
	(0.526)	(0.525)	(0.0466)	(0.0465)
Constant	-2.551	-1.772	0.456	0.352
	(4.440)	(4.400)	(0.426)	(0.420)
N	12453	12453	12438	12438
adj. R^2	0.0667	0.0662		

Standard errors in parentheses. Sample is selected using runners with a prior race 5 minutes above or 5 minutes below their previous Personal Record. The dependent variable is the change in a runner's result from the last race in the first two columns, and an indicator for improvement in the last two columns. * $p < .10$, ** $p < 0.05$, *** $p < .01$

The control variables are similar to past specifications with a few differences. *TotalPR* is not a significant predictor of improvement. In this sample, which is composed of runners at or near their personal best, women are more likely to improve than men and can expect a larger improvement.

We perform falsification tests for our PR specifications. We use a comparison of different distances from the previous PR. For example, we compare those runners who were ten to fifteen minutes above their PR to those five to ten minutes above. Using otherwise identical specifications to those below, we find no significance for our false personal record as reference point variables.

5. CONCLUSIONS AND DISCUSSION

This paper examines the effect of round numbers and personal bests as reference points on changes in future performance. Using the data from numerous different marathons and analyzing the differences in performance of runners over the course of many years, we find support for the hypothesis that personal bests along with round numbers are regarded as reference points for seasoned runners. Additionally, we find that a personal record acts as a motivator after controlling for individual characteristics. These results provide evidence that improvement is influenced by the proximity to a clear goal, which the reference points in this paper certainly are given the regression discontinuity approach.

The marathon setting also allows for a clear and unequivocal measure of improvement. Runners can set a goal with a clear idea of how to achieve it and easily measure success. While some environments may not provide as clear a goal, the marathon results can likely be extended to many important settings. For example, aiming for a target sales point or an investment goal can be viewed in the same light. While the path to those goals may be different, the end result is likely similarly influenced by round numbers and personal bests. A plausible common goal in saving for retirement may be to reserve $1,000 per month to retirement accounts. Proximity to the target amount could lead to a positive change in behavior to reach the round number, while a greater distance from the round number could serve to discourage increased savings. Similarly, a manager may not want to penalize workers who fell just short of a set goal as such situation may promote increased effort during the next task. While the evidence in this paper is specific to sport, we can see many options for improvement from situation tied to clear objectives and obvious reference points.

While our paper clearly illustrates that both round numbers and personal records serve as reference points, we are not yet able to state which of those goals dominates. Examination of marathon finish times suggests round numbers are reference points, but many things could go wrong during the course of a marathon and thus just looking at the finish time may not paint the whole picture. Looking at the pace of a runner at different stages of marathon may help us to determine if the initial goal is to beat the round number mark or if it is to set a personal record, and how that may change during the course of the event. Further work in this environment could also tie the within-race improvement examined in other papers to the between-race improvement we use in ours. Testing whether the round number motivation changes with the ability to prepare could provide additional insight into the behavior of using convenient reference points.

6. REFERENCES

Allen, E., Dechow, P., Pope, D., and Wu, G. "Reference-dependent preferences: Evidence from marathon runners". *Management Science.* 63(6), 2015, 1657-72

Anderson, A., and Green, E. A. "Personal bests as reference points." *Proceedings of the National Academy of Sciences of the United States of America*, *115*(8), 2018, 1772-1776.

Baker, M., Pan, X., and Wurgler, J. "The effect of reference point prices on mergers and acquisitions." *Journal of Financial Economics,* 106(1), 2012, 49–71.

Bar-Eli, M., Tenenbaum, G., Pie, J. S., Btesh, Y., and Almog, A. "Effect of goal difficulty, goal specificity and duration of practice time intervals on muscular endurance performance." *Journal of Sports Sciences*, *15(2)*, 1997, 125-135.

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

Barber, B., Lee, Y., Liu Y., and Odean, T. "Is the Aggregate Investor Reluctant to Realize Losses? Evidence from Taiwan." *European Financial Management*, 13(3), 2007, 423–447. doi: 10.1111/j.1468-036X.2007.00367.x

Boles T. and Messick, D. “A reverse outcome bias: the influence of multiple reference points on the evaluation of outcomes and decisions.” *Organizational Behavior and Human Decision Process*, 61 (3), 1995, 262-275.

Burdina, M., Hiller, R. S., & Metz, N. E. “Goal Attainability and Performance: Evidence from Boston Marathon Qualifying Standards.” *Journal of Economic Psychology*, 58, 2017, 77-88.

Card, D., Mas, A., Moretti, E., Saez, E. “Inequality at work: The effect of peer salaries on job satisfaction.” *American Economic Review*. 102(6), 2012, 2981–3003.

Camerer, C., Babcock L., Loewenstein G., and Thaler R. “Labor Supply of New York City Cabdrivers: One Day at a Time.” *Quarterly Journal of Economics*, 112(2), 1997, 407-41.

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

DellaVigna, S. “Psychology and economics: Evidence from the field.” *Journal of Economic Literature*, 47(2), 2009, 315–372.

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

Grant, D. “The essential economics of threshold-based incentives: Theory, estimation, and evidence from the Western States 100.” *Journal of Economic Behavior & Organization*, 130, 2016, 180-197.

Heath, C., Larrick, R., and Wu, G. “Goals as Reference Points.” *Cognitive Psychology*, 38, 1999, 79-109.

Huang, W. and Tseng, L. “How multiple reference points influence manager’s post-decisional regret.” *Social Behavior and Personality*, 35(4), 2007, 487- 498.

Ivkovic, Z., Poterba, J., and Weisbenner. S. "Tax-Motivated Trading by Individual Investors." *American Economic Review*, 95(5), 2005, 1605-30.

Kahneman, D. and Tversky A. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica,* 47(2), 1979, 263-291.

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

Koszegi, B. and Rabin M. "A Model of Reference-Dependent Preferences." *Quarterly Journal of Economics*, 121(4), 2006, 1133-65.

Koszegi, B. and Rabin, M. "Reference-dependent risk attitudes." *American Economic Review*, 97(4), 2007, 1047-73.

Markle, A., Wu, G., White, R., and Sackett, R. "Goals as Reference Points in Marathon Running: A Novel Test of Reference-Dependence." *Journal of Risk and Uncertainty*, 56(1), 2018, 19-50.

Odean, T. "Are Investors Reluctant to Realize Their Losses?" 1998, *Journal of Finance*, 53(5): 1775-98.

Odean, T. "Do Investors Trade Too Much?" *American Economic Review*, 89(5), 1999, 1279-98.

Pope, D. and Simonsohn, U. "Round numbers as goals". Psychology Science, 22(1), 2011, 71–79.

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

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