



Personal bests as reference points

Ashton Anderson^{a,b,1,2} and Etan A. Green^{b,c,3}

^aDepartment of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3G8; ^bMicrosoft Research New York City, New York, NY 10011; and ^cDepartment of Operations, Information and Decisions, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104

Edited by Jose A. Scheinkman, Columbia University, New York, NY, and approved January 9, 2018 (received for review April 19, 2017)

Personal bests act as reference points. Examining 133 million chess games, we find that players exert effort to set new personal best ratings and quit once they have done so. Although specific and difficult goals have been shown to inspire greater motivation than vague pronouncements to “do your best,” doing one’s best can be a specific and difficult goal—and, as we show, motivates in a manner predicted by loss aversion.

reference points | personal bests | loss aversion | motivation | goals

There is nothing noble in being superior to your fellow man; true nobility is being superior to your former self.

Attributed to Ernest Hemingway

A long line of research suggests that small differences in outcomes are felt disproportionately when they bridge a reference point separating psychological losses from psychological gains (1–3). This phenomenon of loss aversion explains a number of empirical puzzles: aversions to gambles in which losses are possible (4), aversions to parting with randomly endowed objects (5), and aversions to selling investments at a loss (6, 7). Reference points provoke aversions to losses, thereby distorting important decisions. But where do reference points come from?

One source of reference points is externally generated goals (8), such as round numbers. For instance, baseball players, students, and marathon runners exert effort to outperform round-numbered batting averages, standardized test scores, and race times, respectively (9, 10). However, reference points can also be internally generated, as when they correspond to expectations (11, 12) or sunk costs (13). In this paper, we propose that the internally generated goal of one’s personal best, or past peak performance, acts as a reference point. For example, real estate agents may try to beat their biggest sales, auctioneers may try to beat their highest bids, and teachers may try to beat their best evaluations.

We study personal bests in the context of chess ratings. We hypothesize that players will stop playing once they set a new personal best rating, out of an aversion to falling behind, and that they will play longer and try harder when a personal best is in reach, hoping to eclipse it. We ground these hypotheses in a simple utility model, which we detail in *Materials and Methods*. In our model, players choose whether to play and how much effort to exert if they do. The loss-averse player experiences a positive shock when her rating eclipses the reference point (10). In *SI Appendix*, we model how reference-dependent risk preferences affect opponent selection, but we find no support for the theorized relationship in the data.

A principal difficulty in testing these hypotheses is that individuals are typically far from their best, and hence behavior near personal bests is rarely observed. We overcome this difficulty by using a massive dataset comprising 133 million online chess games played by 70,000 players, in which we observe 284,000 instances of new personal bests being set.

We find that a player’s best rating acts as a reference point. First, win rates increase as players approach their personal best ratings, suggesting that players exert effort to set new personal bests. Second, players quit at discontinuously higher rates after setting new personal best ratings, consistent with an aversion to falling back into the domain of losses. For comparison, we con-

duct comparable tests for round-numbered ratings. Whereas personal bests influence both decisions over whether to play and how much effort to exert during games, round numbers only influence decisions over whether to play.

The literature on goal setting concludes that specific and appropriately difficult goals inspire greater motivation than vague pronouncements to “do your best” (14, 15). Yet, when performance is quantifiable, doing one’s best is a specific goal. It is also calibrated to be appropriately difficult (cf. ref. 16)—rarely impossible, and, if too easy, quickly surpassed and reset. We show that people exert effort to do their best and quit once they have done so, consistent with loss aversion around personal best reference points.

Results

To test our theoretical predictions, we conduct an empirical analysis of behavior around personal bests in the context of chess ratings on the Free Internet Chess Server (FICS). A chess player is assigned a rating, updated after every game she plays, that estimates her skill level. The FICS rating system is simple: when a player wins, her rating goes up, and when she loses, it goes down. How many rating points each player would gain with a win or lose with a loss depends on the difference in the players’ ratings (see *Materials and Methods* for further details on the rating system). Ratings fluctuate around a player’s true skill, and, when these fluctuations reach a new peak, the player sets a new

Significance

Research in psychology, economics, and neuroscience suggests that small differences in outcomes are felt disproportionately when they bridge a reference point separating psychological losses from psychological gains. However, knowledge of where reference points come from is limited. We propose that one’s personal best, or past peak performance, acts as a reference point by inducing effort when current performance would otherwise fall short. Analyzing a massive dataset of online chess games, we find that players exert effort to set new personal best ratings and quit once they have done so. In education, fitness, and other domains, technology is making performance quantifiable. Our results suggest that these advances will motivate individuals to compete with their past selves.

Author contributions: A.A. and E.A.G. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The data reported in this paper have been deposited in the Open Science Framework (https://osf.io/tsfgv/?view_only=e9e1c483aa3d488797f2e70de86abea1).

¹To whom correspondence should be addressed. Email: ashton@cs.toronto.edu.

²Present address: Department of Computer Science, University of Toronto, Toronto, ON, Canada M5S 3G8.

³Present address: Department of Operations, Information and Decisions, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1706530115/-DCSupplemental.

rating	RD	win	loss	draw	total	best
1464	41.3	5638	7092	747	13477	1573 (12–Nov–2016)

Fig. 1. Example user information display. RD measures the variance of a player's rating.

personal best. The rest of the time, her current rating trails her personal best.

Chess ratings are highly visible. Fig. 1 shows an example user's information page, which publicly and prominently displays the player's current rating and best past rating. Current ratings are also shown beside players' names when playing a game, and players receive text and sound notifications when they reach a new personal best.

We construct our dataset from the complete set of blitz games (which are expected to last between 6 min and 30 min) played on FICS between 2000 and 2015. Our unit of analysis is the "player-game," of which there are two per game: one for the white pieces and one for the black pieces. The complete set of blitz games comprises 313 million player-games across 156.5 million games. To produce the dataset for our main analyses, we carry out a series of filtering steps. For example, we filter out player-games before a player's 200th game, to allow players to establish a meaningful personal best; we filter out player-games where the player's rating is too uncertain; and we filter out player-games where the player has achieved a personal best in the last 20 games, so as to consider instances in which beating a personal best is a meaningful goal (see *Materials and Methods* for more details). After filtering, our dataset comprises 212 million player-games across 133 million games. Our results do not depend on these filtering restrictions; as described in *SI Appendix*, we replicate our empirical results with different filtering parameters and obtain meaningfully unchanged results.

Fig. 2 shows a histogram of games at each value of the difference between a player's current rating and her personal best before her last game. Players are typically far from their personal best ratings—the median difference is -118 points, and only 3.7% of games are played within 30 points of the player's personal best. For reference, in most cases, a win against an equally rated player adds 8 points, a loss to an equally rated player subtracts 8 points, and the greatest possible rating change from a single game is 16 points. In the histogram, values to the right of 0 represent instances of players setting new personal bests. Although

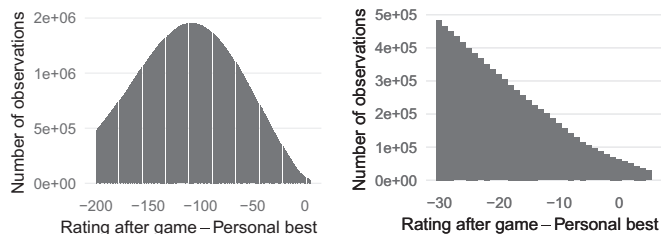


Fig. 2. Histogram of the difference between current ratings and personal best ratings (zoomed in near the reference point on the right).

only 1 in 750 player-games results in a new personal best, we still observe 284,000 instances of personal bests being set.

We observe how behavior changes as players approach and surpass their personal best ratings. In the main text, we report average outcomes for each rating point difference between a player's current rating and her personal best rating. Where comparisons are made between players whose ratings are just shy of their personal bests and those who just set a new personal best by winning the previous game, we restrict the sample to the 101.5 million player-games that follow a win. This restriction ensures that observations on either side of the reference point are comparable.

To observe how behavior changes as players approach their personal bests, we compare behavior when players' ratings are close to their personal bests with behavior when players' ratings are farther away. One concern with this approach, however, is that certain types of players may be close to their personal bests more often than other types of players. Thus, differences in behavior may be confounded by differences in player attributes. To address this concern, we run comparable regressions with player fixed effects, which we report in *SI Appendix*. These estimates reflect only within-player differences in behavior, rather than differences between players.

Quitting. What happens when players set new personal bests? Fig. 3A shows how the probability of quitting varies with the distance between a player's current rating and her personal best rating from before her last game. We define quitting as not playing another game within 1 h of finishing the most recent game; in *SI Appendix*, Fig. S8, we show qualitatively identical results for a 24-h threshold. The probability of quitting jumps across the

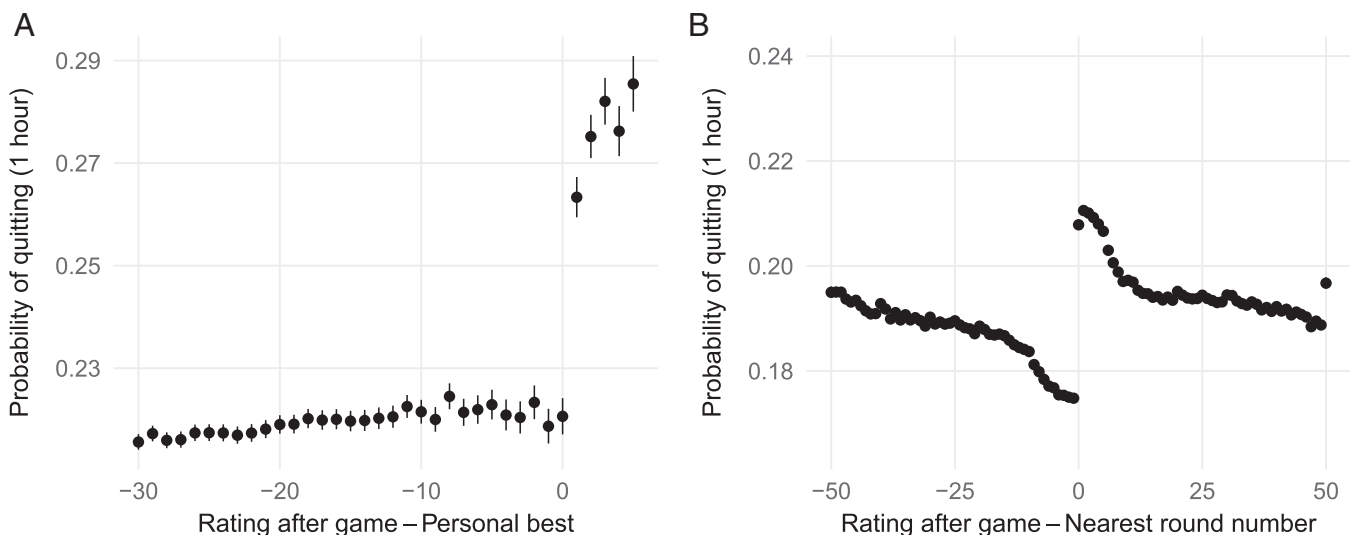


Fig. 3. Probability of quitting for at least 1 h around personal bests (A) and round numbers (B), with 95% confidence intervals.

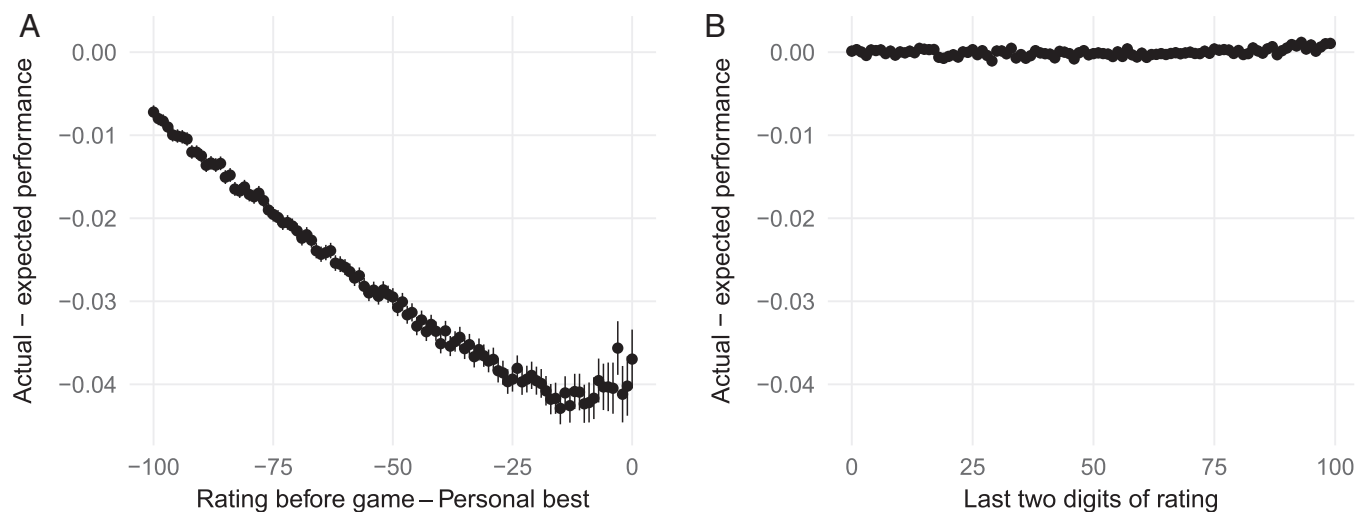


Fig. 4. Performance short of personal bests (A) and round numbers (B), with 95% confidence intervals.

reference point—a 4.5 percentage point, or 20%, increase. As predicted, players are discontinuously more likely to quit after setting a new personal best.

This effect is more pronounced for more-frequent players and for long-standing personal bests. Among the half of players whose median time between games is less than 10 min, the probability of quitting jumps 29% (a 4.6 percentage point increase from a baseline of 15.7 percentage points), compared with a jump of 14% (a 4.1 percentage point increase from a baseline of 30.5 percentage points) among less frequent players. Breaking a personal best that is fewer than 20 games old is associated with a 9% jump in the probability of quitting (a 2.0 percentage point increase on a baseline of 22.2 percentage points), compared with a 20% jump in the probability of quitting (a 4.5 percentage point increase on a baseline of 21.9 percentage points) for personal bests that have stood for at least 20 games.

Achieving a personal best precipitates not only a higher rate of quitting but also longer quitting spells. Among those who quit with ratings one point short of their personal bests, the median duration between games is 752 min. Among those who quit after eclipsing their personal bests by one rating point, the median duration between games is 816 min.

For comparison, we measure quitting near round numbers, which have been shown to act as reference points in other domains (9, 10). Fig. 3B shows how the probability of quitting varies with the distance to the nearest multiple-of-100 rating (where all ratings ending in 51 to 99 are to the left of 0, and all ratings ending in 01 to 50 are to the right of 0). As with personal bests, players are discontinuously more likely to quit after breaking a century marker—players with ratings ending in 01 quit 3.5 percentage points more often than players with ratings ending in 99. This relative increase of 20% is the same as the corresponding relative increase around personal bests. (There is also a smaller discontinuous jump around the round number of 50—players with ratings ending in 51 quit 0.6 percentage points, or 3.3%, more often than players with ratings ending in 49.) By this comparison, personal bests motivate as powerfully as round numbers.

We find evidence of a goal gradient, or the increase in intensity often observed when a goal is imminent (17–19), for round-numbered ratings but not for personal bests. The probability of quitting decreases as ratings approach a multiple of 100 but stays flat as ratings approach personal bests. We suspect that this disparity follows from differential awareness of the two reference points, rather than differential motivation. Players receive notifications of their personal best ratings only after they eclipse their

previous best. Current ratings, by contrast, are shown beside player names during every game they play. Hence, players whose ratings trail their personal best ratings are likely more aware of their proximity to round-numbered reference points than to their personal bests.

Effort. Do players try harder when a personal best is within reach? Effort is difficult to observe directly, so we measure effort indirectly as performance relative to expectations. Specifically, we compare observed win rates to predicted win rates, where the predicted win rate is the empirically observed probability of a win for a given difference in ratings between the player and her opponent. (We treat a draw as half a win.) In our data, players win 50% of games against equally rated opponents, they win 62% of games against opponents whom they outrate by 100 points, and they win 73% of games against opponents whom they outrate by 200 points. Do players win more often than these expectations when they are close to their best ratings?

If effort enhances performance, and if players try harder when a personal best is in reach, then win rates will outperform expectations when current ratings are just short of personal bests. However, ratings fluctuate around a player's true skill, implying that higher ratings overestimate ability. Hence, regression to the mean predicts that win rates will underperform expectations as current ratings approach personal bests. Jointly, these effects predict that regression to the mean will subside, and may even reverse, near personal bests.

Fig. 4A shows the difference between observed and predicted win rates as a function of a player's rating distance from her personal best. Away from the reference point, performance declines as ratings increase, in line with regression to the mean. However, the trend abates about 10 rating points from the reference point—approximately the distance at which a win could realistically set a new personal best rating. At the reference point, performance is ~ 1 percentage point higher than if the prevailing trend had continued unabated. This suggests that players try harder when near their personal best—so much so as to reverse the regression to the mean. Although we cannot identify the mechanism by which performance improves (whether by heightened concentration, computer assistance, selection of overrated opponents, or other means), the improvement implies that players find some way to exceed expectations when a personal best is within reach.

Does effort subside just after setting a new personal best rating? In *SI Appendix*, Fig. S9, we estimate the same performance

measure for players who have just set a new personal best. As in the quitting analysis, we restrict the data to player-games following a win, to ensure that observations on either side of the boundary are comparable. We find that performance is similar across the boundary. Players continue exerting effort after setting a new personal best, likely in pursuit of another.

As in *Quitting*, we perform a comparable analysis for round-numbered ratings. We do not expect a regression to the mean for round numbers, since there is no analogous selection effect on the last two digits of the rating. Fig. 4B shows the difference between observed and predicted win rates as a function of a player's rating distance from round numbers. Actual performance is almost identical to expected performance throughout the entire range, implying that, when just shy of round-numbered ratings, players do not increase their effort enough to improve performance. By this comparison, personal bests inspire more effort than round numbers.

Discussion

Quantitative measures of performance are ubiquitous, and peak performance along these measures is often salient. Many students care about their highest test scores (20), and many athletes care about their fastest times (21). Moreover, quantitative measures of performance are proliferating. Recent educational programs in the United States expanded the use of test scores to evaluate schools and teachers (22), and new devices quantify performance along dimensions hitherto ignored. For instance, the proliferation of accelerometers on wrists and in pockets has created a sudden awareness of, and competitiveness over, the most steps one has taken in a day (23). When performance can be tracked, peak past performance becomes a salient benchmark for comparison.

Previous research shows that peak events factor disproportionately in experienced utility (24) and self-perceptions (25). We show that peak performance acts as a reference point. Individuals exert effort to achieve new personal bests and quit once they've done so.

Materials and Methods

The Rating System. FICS assigns a rating to every player at every point in time using the Glicko algorithm, which is an extension of the popular Elo rating system used by official chess federations (26). The algorithm is Bayesian and models a player's rating as a Gaussian belief distribution characterized by a mean and a variance, with an initial mean of 1,720 points and an initial variance of 350 points.

The mean is the player's rating and is updated from game results according to the ratings of the players. The amount the player gains from winning a game is a logistic function of the rating difference between the player and the opponent. For a rating difference $D = \text{rating}_{\text{player}} - \text{rating}_{\text{opponent}}$, the victory reward is $\Delta = k(1 - 1/(1 + 10^{-cD}))$ rating points, where $c \approx 1/400$, k is the maximum victory reward (usually 16), and the penalty for losing is $k - \Delta$ points. The constant c is calibrated such that the expected rating change is always zero. Hence, a victory reward of Δ is associated with a win probability of $1 - \Delta/k$. For instance, a player who chooses $\Delta = 3k/4$ —i.e., who gains $3k/4$ rating points with a win and loses $k/4$ rating points with a loss—is expected to win $1 - \Delta/k = 25\%$ of the time.

The player's rating variance decreases as she plays more games and increases as time elapses since her last game. This variance is used in two ways: (i) to determine whether a high rating counts as a personal best—ratings only count as personal bests when the variance is below 80—and (ii) to scale the maximum victory reward k . Ratings are designed to fluctuate more when they have high variance, so the maximum victory reward is an increasing function of rating variance.

The rating system is well calibrated. *SI Appendix, Fig. S10* shows a calibration plot comparing predicted and actual win rates at each victory reward when $k = 16$. Predicted and actual win rates match closely at every victory reward. All of the details of the exact FICS implementation of the Glicko rating system can be found online (www.freechess.org/Help/HelpFiles/glicko.html).

Data Preparation and Descriptives. Our data comprise the complete set of blitz games—i.e., with a maximum duration between 6 min and 30 min—

played on the FICS between 2000 and 2015. Each observation is a player-game: For each game, there are two player-games, one for either side. We exclude player-games for which the player (i) joined FICS before January 1, 2000, (ii) is a computer account—either a bot or a human who uses chess program assistance during her games, (iii) plays against computer accounts in $>25\%$ of games, (iv) has played fewer than 200 games, (v) has a rating variance greater than 80, or (vi) set a personal best rating in the previous 20 games.

The first restriction ensures that we have complete data for every player in our dataset, which starts on January 1, 2000. The second limits our attention to human players. The third excludes players who may be abusing the rating system; players who play too many games against bots may be repeatedly exploiting known bot weaknesses to unfairly gain rating points. The fourth affords players the opportunity to set meaningful personal bests. The fifth restricts our attention to player-games that could potentially count as personal bests; a rating can only count as a personal best if the rating variance is less than 80. The final restriction limits our analyses to instances in which personal bests are likely to be meaningful goals. Absent this restriction, our analyses are complicated by the fact that players close to their personal bests are a mix of two different populations: those who just set a personal best and then lost and those who have not set a personal best recently.

This filtering leaves a dataset of 212 million player-games, for which we observe the identity of the player, her rating, and her best rating; the identity of the opponent and her rating; the game result; and a timestamp for when the game began.

Effort Model. Consider a player with rating r contemplating a game with a victory reward $\Delta \in [0, k]$. If she wins, her rating rises to $r + \Delta$; if she loses, her rating drops to $r + \Delta - k$. The probability of winning is increasing in $e \in [0, \infty)$, the effort she exerts. We represent this relationship with a cumulative distribution function, $F(e; \Delta)$; for brevity, we write this function as $F(e)$. Infinite effort guarantees victory; zero effort guarantees a loss. We assume that the first unit of effort increases the probability of winning the most, or, more generally, that marginal gains from effort are decreasing—i.e., $F''(e) < 0$. Effort is costly, however, with a cost, $c(e)$, such that $c(0) = 0$ and $c'(e) > 0$. We further assume that the first unit of effort is the least costly, or, more generally, that the marginal cost of effort is increasing—i.e., $c''(e) > 0$.

We are interested in how proximity to a reference point influences a loss-averse player's willingness to play and her exertion when she does. Following ref. 10, we assume a piecewise-linear value function that jumps at a reference point θ ,

$$v(x) = \begin{cases} x + \epsilon & x > \theta \\ x & x \leq \theta \end{cases}$$

The jump at θ implies loss aversion—i.e., the player experiences the greatest loss when a fixed decrease in r shifts her rating from the domain of gains to the domain of losses.

Effort. We first evaluate the player's effort conditional on choosing to play. In particular, we consider two ratings: one such that $r + \Delta \leq \theta$ —i.e., the player cannot reach the reference point even if she wins—and another such that $r + \Delta - k \leq \theta < r + \Delta$ —i.e., a win puts the player above the reference point, and a loss puts her below it.

The player's expected utility sums over four components: a positive utility shock from playing, which we denote α , the valuation of her rating were she to win; the valuation of her rating were she to lose; and the cost of effort. For $r + \Delta \leq \theta$,

$$\mathbb{E}[U(e)] = \alpha + (r + \Delta)F(e) + (r + \Delta - k)(1 - F(e)) - c(e). \quad [1]$$

The optimal, or utility-maximizing, effort level satisfies the first-order condition

$$\underbrace{k \cdot F'(e)}_{\text{Marginal expected gain}} = \underbrace{c'(e)}_{\text{Marginal cost}}. \quad [2]$$

The player exerts effort until the marginal expected gain is equal to the marginal cost. If the first unit of effort is more beneficial than costly—i.e., if $k \cdot F'(0) > c'(0)$ —then there exists a unique utility-maximizing effort level. Both existence and uniqueness follow from the assumptions of strictly declining marginal gains from effort ($F''(e) < 0$) and strictly increasing marginal costs of effort ($c''(e) > 0$).

Now consider the expected utility of a player for whom $r + \Delta - k \leq \theta < r + \Delta$,

$$\mathbb{E}[U(e)] = \alpha + (r + \Delta + \epsilon)F(e) + (r + \Delta - k)(1 - F(e)) - c(e). \quad [3]$$

The optimal effort level satisfies the first-order condition

$$\underbrace{(k + \epsilon) \cdot F'(e)}_{\text{Marginal expected gain}} = \underbrace{c'(e)}_{\text{Marginal cost}}. \quad [4]$$

When the outcome of the game determines whether the player ends up above or below the reference point, her marginal expected gain is greater than when the outcome could not change her position relative to the reference point.

To determine how the optimal effort level changes between Eqs. 2 and 4, we derive a relationship between the optimal effort level, e^* , and the coefficient on the marginal gain. Consider a generalized form of the first-order condition, $a \cdot F'(e^*) = c'(e^*)$, where $a > 0$. Then, by the Implicit Function Theorem,

$$\frac{\partial e^*(a)}{\partial a} = \frac{-F'(e^*)}{aF''(e^*) - c''(e^*)} > 0, \quad [5]$$

where the inequality follows from our assumptions about winning probabilities ($F'(e) > 0, F''(e) < 0$) and costs ($c'(e) > 0$). An increase in a , or the marginal gain from effort, increases the optimal effort. Hence, the reference-dependent player exerts more effort when the outcome determines which side of the reference point she falls on.

Quitting. We now consider how the player's proximity to the reference point influences her willingness to play. A player is willing to play if her expected utility from playing is greater than her utility from not playing, which is simply the value of her rating, $v(r)$. Let Δ be such that $r + \Delta - k \leq \theta < r + \Delta$ —i.e., the player will end above the reference point if she wins and below the reference point if she loses. Let e^* be the optimal effort when close to the reference point—i.e., the effort level that solves the first-order condition in Eq. 4. Then she plays if and only if $v(r) < \mathbb{E}[U(e^*)]$, or

$$v(r) < \alpha + r + \Delta - k + (k + \epsilon) \cdot F(e^*) - c(e^*). \quad [6]$$

We are interested in how the player's willingness to play changes as her rating moves across the reference point. When the player's rating is below the reference point, then $v(r) = r$, and she plays if

$$\alpha > k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*). \quad [7]$$

When the player's rating is above the reference point, then $v(r) = r + \epsilon$, and she plays if

$$\alpha > k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*) + \epsilon. \quad [8]$$

Hence, the threshold for playing is higher (by ϵ) when the player's rating is above the reference point than when her rating is below the reference point. In other words, the player needs to gain more utility from playing—i.e., she needs to have a higher α —to absorb the risk of falling below the reference point.

Further, assume that players enjoy playing to different degrees—i.e., that for player i , α_i is drawn from a distribution G_α . Thus, the probability that a player just below the reference point chooses to play is $1 - G_\alpha(\gamma)$, where $\gamma \equiv k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*)$; the probability that a player just above the reference point chooses to play is $1 - G_\alpha(\gamma + \epsilon)$. This implies that the probability of playing drops—or that the probability of quitting jumps—at the reference point by an amount equal to $G_\alpha(\gamma + \epsilon) - G_\alpha(\gamma)$.

Goal Gradient. Our model also implies a goal gradient, or that a reference-dependent player will be more willing to play as her rating approaches the reference point. To see this, compare two players with ratings short of the reference point. For the first player, $r_1 + \Delta > \theta$ —i.e., a win would push her rating past the reference point. For the second player, $r_2 + \Delta \leq \theta$ —i.e., a win would not push her rating past the reference point. The model predicts a goal gradient if player 1's threshold for playing is lower than player 2's.

Player 1 plays if $\alpha > k - \Delta + c(e_1^*) - (k + \epsilon) \cdot F(e_1^*)$, as in Eq. 7, with e_1^* solving the first-order condition in Eq. 4. Player 2 plays if $\alpha > k - \Delta + c(e_2^*) - k \cdot F(e_2^*)$, with e_2^* solving the first-order condition in Eq. 2. Hence, player 1's threshold for playing is lower if $(k + \epsilon) \cdot F(e_1^*) - c(e_1^*) > k \cdot F(e_2^*) - c(e_2^*)$, i.e., if the expected net gain from winning (in utiles) is greater for player 1 than for player 2. This condition always holds. The player exerts effort until marginal gains equal marginal costs. Since $e_1^* > e_2^*$, player 1 exerts more effort than player 2, implying that net gains are larger for player 1.

ACKNOWLEDGMENTS. We thank participants at the 2017 Summer Decision Making Symposium, International Conference on Computational Social Science 2017, and the 2017 Meeting of the Society for Judgment and Decision Making, as well as Dan Goldstein, Dorothy Kronick, Nicolas Lambert, Alex Markle, Alex Rees-Jones, Maurice Schweitzer, Joe Simmons, Uri Simonsohn, and Glen Weyl for helpful discussions and comments on previous drafts. We are also indebted to Jake Hofman, without whom this paper would not exist. We are grateful for generous financial assistance from Microsoft Research, The Wharton School, and the University of Toronto. Part of this work was conducted while A.A. and E.A.G. were at Microsoft Research New York City.

1. Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–292.
2. Tom SM, Fox CR, Trepel C, Poldrack RA (2007) The neural basis of loss aversion in decision-making under risk. *Science* 315:515–518.
3. De Martino B, Camerer CF, Adolphs R (2010) Amygdala damage eliminates monetary loss aversion. *Proc Natl Acad Sci USA* 107:3788–3792.
4. Rabin M (2000) Risk aversion and expected-utility theory: A calibration theorem. *Econometrica* 68:1281–1292.
5. Kahneman D, Knetsch JL, Thaler RH (1990) Experimental tests of the endowment effect and the Coase theorem. *J Pol Econ* 98:1325–1348.
6. Odean T (1998) Are investors reluctant to realize their losses? *J Finance* 53:1775–1798.
7. Genesove D, Mayer C (2001) Loss aversion and seller behavior: Evidence from the housing market. *Q J Econ* 116:1233–1260.
8. Heath C, Larrick RP, Wu G (1999) Goals as reference points. *Cognit Psychol* 38: 79–109.
9. Pope D, Simonsohn U (2011) Round numbers as goals evidence from baseball, SAT takers, and the lab. *Psychol Sci* 22:71–79.
10. Allen EJ, Dechow PM, Pope DG, Wu G (2016) Reference-dependent preferences: Evidence from marathon runners. *Manage Sci* 63:1657–1672.
11. Mellers B, Schwartz A, Ho K, Ritov I (1997) Decision affect theory: Emotional reactions to the outcomes of risky options. *Psychol Sci* 8:423–429.
12. Köszegi B, Rabin M (2006) A model of reference-dependent preferences. *Q J Econ* 121:1133–1165.
13. Thaler RH, Johnson EJ (1990) Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Manage Sci* 36:643–660.
14. Locke EA, Shaw KN, Saari LM, Latham GP (1981) Goal setting and task performance: 1969–1980. *Psychol Bull* 90:125–152.
15. Locke EA, Latham GP (2002) Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *Am Psychol* 57:705–717.
16. Ordóñez LD, Schweitzer ME, Galinsky AD, Bazerman MH (2009) Goals gone wild: The systematic side effects of overprescribing goal setting. *Acad Manage Perspect* 23: 6–16.
17. Hull CL (1932) The goal-gradient hypothesis and maze learning. *Psychol Rev* 39: 25–43.
18. Kivetz R, Urminsky O, Zheng Y (2006) The goal-gradient hypothesis resurrected: Purchase acceleration, illusory goal progress, and customer retention. *J Marketing Res* 43:39–58.
19. Anderson A, Huttenlocher D, Kleinberg J, Leskovec J (2013) Steering user behavior with badges. *Proceedings of the 22nd International Conference on World Wide Web (Assoc Comput Machinery, New York)*, pp 95–106.
20. Martin AJ (2006) Personal bests (PBs): A proposed multidimensional model and empirical analysis. *Br J Educ Psychol* 76:803–825.
21. Stoeber J, Uphill MA, Hotham S (2009) Predicting race performance in triathlon: The role of perfectionism, achievement goals, and personal goal setting. *J Sport Exerc Psychol* 31:211–245.
22. Wong A (December 9, 2015) The bloated rhetoric of No Child Left Behind's demise. *The Atlantic*. Available at <https://www.theatlantic.com/education/archive/2015/12/the-bloated-rhetoric-of-no-child-left-behinds-demise/419688/>. Accessed January 26, 2018.
23. Sedaris D (June 30, 2014) Stepping out: Living the Fitbit life. *The New Yorker*. Available at <https://www.newyorker.com/magazine/2014/06/30/stepping-out-3>. Accessed January 26, 2018.
24. Kahneman D, Wakker PP, Sarin R (1997) Back to Bentham? Explorations of experienced utility. *Q J Econ* 112:375–405.
25. Williams EF, Gilovich T (2012) The better-than-my-average effect: The relative impact of peak and average performances in assessments of the self and others. *J Exp Social Psychol* 48:556–561.
26. Glickman ME (1999) Parameter estimation in large dynamic paired comparison experiments. *Appl Stat* 48:377–394.