Personal Bests as Reference Points*

Ashton Anderson[1] and Etan A. Green[2]

[1]Microsoft Research
[2]University of Pennsylvania

April 15, 2017

Abstract

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. Though 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.

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

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 (Tversky and Kahneman, 1991). This phenomenon of loss aversion explains a number of empirical puzzles: aversions to gambles in which losses are possible (Kahneman and Tversky, 1979), aversions to parting with randomly endowed objects (Kahneman, Knetsch and Thaler, 1990), and aversions to selling investments at a loss (Odean, 1998; Genesove and Mayer, 2001). Reference points provoke aversions to losses, thereby distorting important decisions. But where do reference points come from?

One source of reference points are externally generated goals (Heath, Larrick and Wu, 1999), 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 (Pope and Simonsohn, 2011; Allen, Dechow, Pope and Wu, 2016). However, reference points can also be internally generated, as when they correspond to expectations (Mellers, Schwartz, Ho and Ritov, 1997) or sunk costs (Thaler and Johnson, 1990). 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,

1

and that they will play more often and try harder when a personal best is in reach, hoping to eclipse it.[1]

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 by employing 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. Win rates increase as players approach their personal-best ratings, suggesting that players exert extraordinary effort to set new personal bests. And the probability of quitting jumps discontinuously after players set new personal-best ratings, implying that players stop exerting effort after achieving new personal bests.

We conduct comparable tests for round-numbered ratings, and find that personal bests inspire greater motivation than round numbers. The discontinuous jump in the probability of quitting is 41% larger for personal-best ratings than for multiple-of-100 ratings. We also find no change in performance near round-numbered ratings, whereas there is a significant change near personal bests. However, we find that players play more often as their ratings approach multiples of 100, whereas there is no such increase in intensity as ratings approach personal bests.

The literature on goal setting concludes that specific and appropriately difficult goals inspire greater motivation than vague pronouncements to "do your best" (Locke, Shaw, Saari and Latham, 1981; Locke and Latham, 2002). Yet when performance is quantifiable, doing one's best is a specific goal. It is also calibrated to be appropriately difficult (cf. Ordóñez, Schweitzer, Galinsky and Bazerman, 2009)—rarely impossible, and if too easy,

---

[1]We ground these hypotheses in a simple utility model, which we detail in Appendix A. 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, as assumed by Allen et al. (2016) and evidenced by Markle, Wu, White and Sackett (2015).

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.

Personal Bests in Chess

To test our theoretical predictions, we conduct an empirical analysis of 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. The rating system is also transparent: exactly how much each player will gain or lose with a win or a loss is known in advance (Appendix B provides further detail on the chess rating system). Ratings fluctuate around a player's true skill. When those fluctuations reach a new peak, the player sets a new personal best. The rest of the time, her current rating trails her personal best.

Figure 1: Example user information display

```
Last disconnected: Wed Dec  7, 23:14 EST 2016

          rating    RD     win    loss    draw   total   best
Blitz      1464    41.3   5638    7092     747   13477   1573 (12-Nov-2016)
Standard   1723   193.8    142     116      19     277   1740 (16-Aug-2012)
Lightning  1469    77.1     17      66       4      87   1484 (07-Dec-2016)
```

Note: We analyze Blitz games, which last approximately 6–30 minutes. RD measures the variance of a player's current rating.

Chess ratings are highly visible. Figure 1 shows an example user's information page, which publicly and prominently displays the player's current rating and best past rating. We construct our dataset from the complete set of blitz games (which are expected to last between 6 and 30 minutes) 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

3

black pieces. The complete set of blitz games comprises 313M player-games across 156.5M games. To produce our final dataset, 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 it is more likely that beating a personal best would be a meaningful goal (see Appendix C for more details). After filtering, our dataset comprises 212M player-games across 133M games. Our results do not depend on any of these filtering restrictions; in Appendix D, we replicate our empirical results with different filtering parameters and obtain qualitatively identical results.
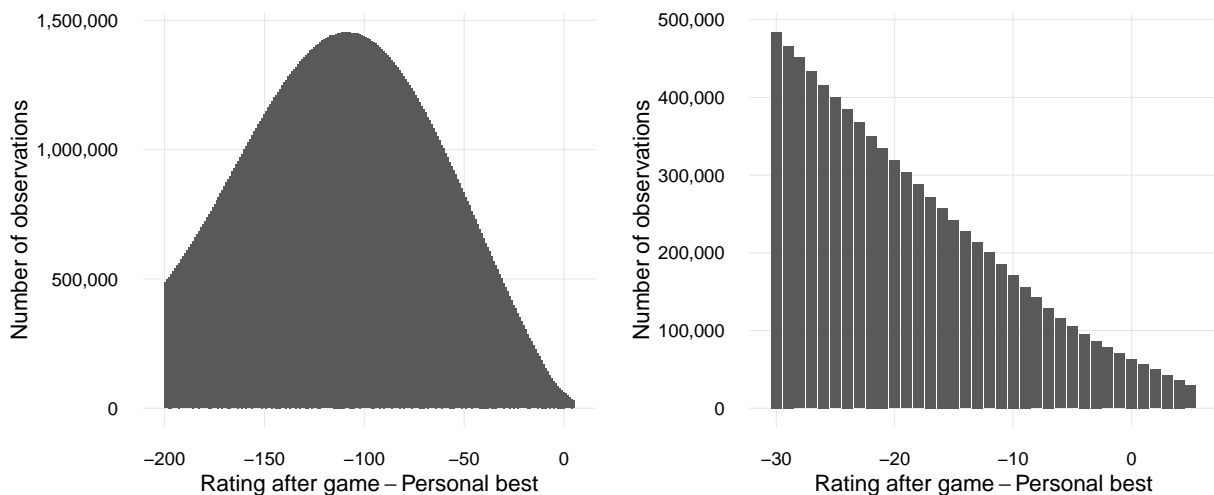

Figure 2: Histogram of differences between current ratings and personal bests.

Figure 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. Values to right of 0 represent instances of players setting new personal bests. Although only 1 in 750 player-games result in a new personal best, we still observe 284,000 instances of personal bests being set.

4

Quitting

What happens when players set new personal bests? Figure 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.[2] We define quitting as not playing another game within one hour of finishing the most recent game; in Appendix D, we show similar results for longer thresholds. Across the reference point, the probability of quitting jumps discontinuously—a 4.3 percentage point, or 20%, increase. As predicted, players are more likely to quit after setting a new personal best.
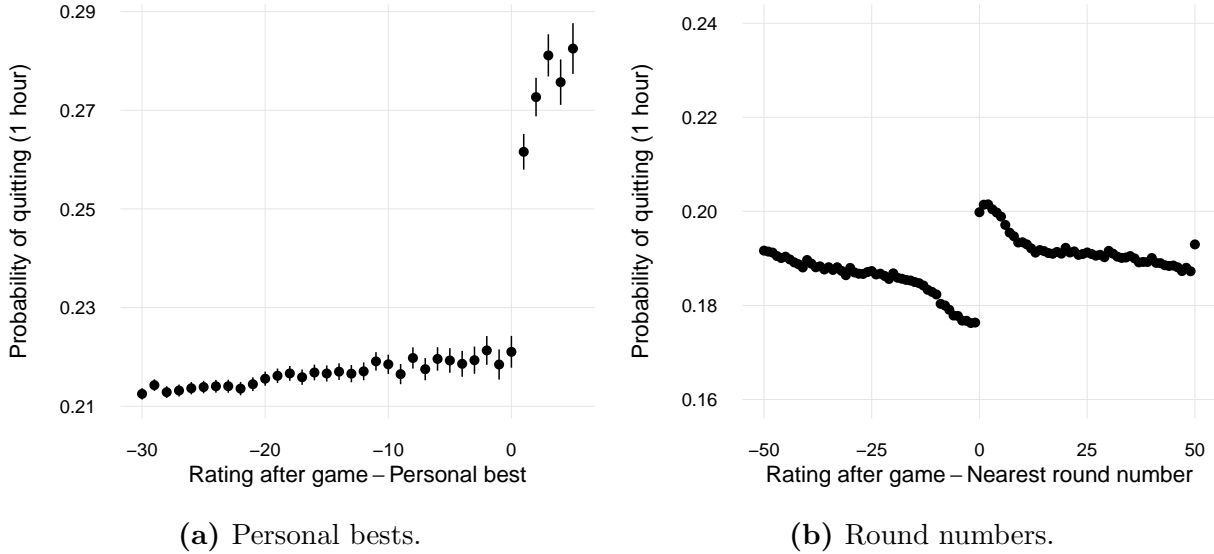

(a) Personal bests.   **(b)** Round numbers.

Figure 3: Probability of quitting for at least 1 hour, with 95% confidence intervals.

We do not find evidence of a goal gradient, or the increase in intensity often observed when a goal is imminent (Hull, 1932; Kivetz, Urminsky and Zheng, 2006; Anderson, Huttenlocher, Kleinberg and Leskovec, 2013). In fact, the probability of quitting rises slightly as ratings approach personal bests. We speculate that this is due to selection: players just short of

---

[2]Since values above the reference point can only be achieved with a win in the previous game, we restrict the sample to observations that follow a win to ensure that ratings on either side of the reference point are comparable.

their personal bests tend to have played fewer games and play shorter sessions (and hence quit more frequently) than players further from their personal bests.

For comparison, we measure quitting near round numbers, which have been shown to act as reference points in related domains (Pope and Simonsohn, 2011; Allen et al., 2016). Figure 3b shows how the probability of quitting varies with the distance to the nearest multiple-of-100 rating (where all ratings ending in 51–99 are to the left of 0, and all ratings ending in 01–50 are to the right). As with personal bests, players are discontinuously more likely to quit after breaking a century marker—players with ratings ending in 01 quit 2.5 percentage points more often than players with ratings ending in 99. This relative increase of 14.2% is statistically smaller than the corresponding relative increase around personal bests ($p < 10^{-6}$). There is also a smaller discontinuous jump around the round number of 50—players with ratings ending in 51 quit 0.4 percentage points more often than players with ratings ending in 49, a relative increase of 2.2%. By these comparisons, personal bests motivate more strongly than round numbers.

Notably, behavior around round numbers follows a goal gradient, in contrast to personal bests, with the probability of quitting decreasing as ratings approach round numbers.

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 empirical probability of a win for a given difference in ratings between the player and her opponent.[3] For instance, players win 50% of games against equally rated opponents, they win 62% of games against opponents whom they out-rate by 100 points, and they win 73%

---

[3]We treat a draw as half a win.

of games against opponents whom they out-rate by 200 points. Do players win more often than expected as they approach their personal-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 *under*perform 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.
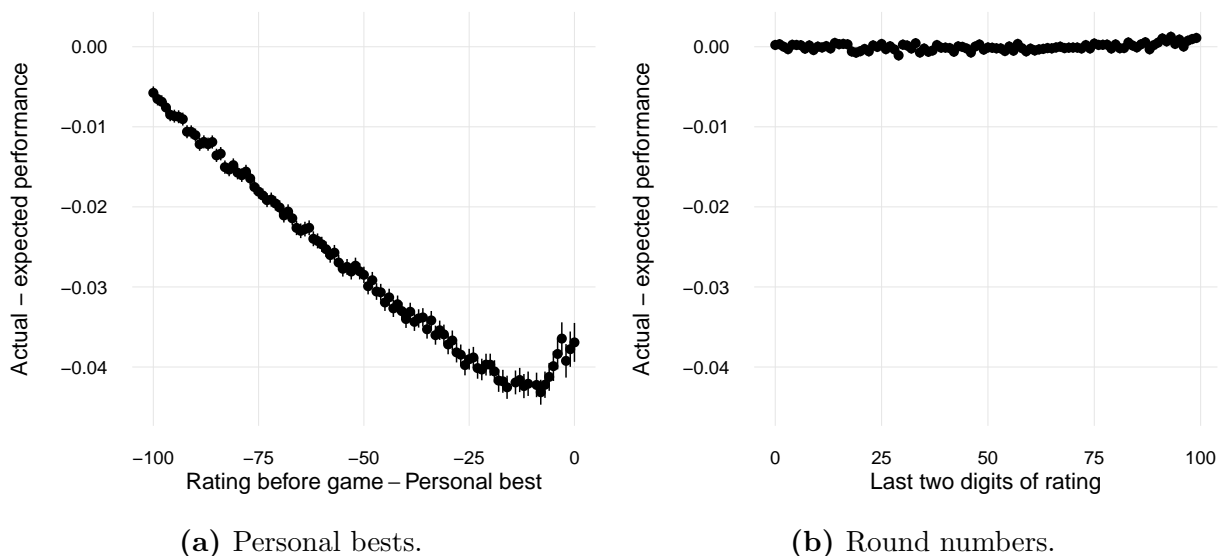

(a) Personal bests.

(b) Round numbers.

Figure 4: Performance short of personal bests and round numbers, with 95% confidence intervals.

Figure 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 reverses about 10 rating points from the reference point—approximately the distance at which a win could realistically set a new personal-best rating—and increases until the reference point. At the reference point, performance is approximately 1 percentage point higher than it would have been if the prevailing trend continued unabated until the reference point. This suggests

that players try harder when near their personal best—so much so as to reverse the regression to the mean. Though 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.

As in the previous section, 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. Figure 4b shows the difference between observed and predicted win rates as a function of a player's rating distance from round numbers (the y-axis is fixed to be the same as in Figure 4a). 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, round numbers inspire less effort than personal bests.

Discussion

Quantitative measures of performance are ubiquitous, and peak performance along these measures is often salient. Many students care about their highest test scores (Martin, 2006), and many athletes care about their fastest times (Stoeber et al., 2009). Moreover, quantitative measures of performance are proliferating. Recent educational programs in the US expanded the use of test scores to evaluate schools and teachers (Wong, 2015). 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 (e.g., Sedaris, 2014). 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

(Kahneman, Wakker and Sarin, 1997) and self perceptions (Williams and Gilovich, 2012). 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.

We predict similar behavior in other settings with caution. In many settings, performance is measured by outcomes, as with marathon times or test scores. In chess, every game offers an outcome—a win, a loss, or a draw—but performance is measured by ratings, which summarize an entire history of outcomes. Hence, ratings offer a more reliable measure of ability than outcomes, which may be intermittent and noisy. The behaviors we document occur when an individual is measurably close to her best. Hence, we suspect that these behaviors will be muted when it is unclear how far one is from her best.

References

Allen, Eric J, Patricia M Dechow, Devin G Pope, and George Wu (2016) "Reference-dependent preferences: Evidence from marathon runners," *Management Science*.

Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec (2013) "Steering user behavior with badges," in *Proceedings of the 22nd international conference on World Wide Web*, pp. 95–106, ACM.

Genesove, David and Christopher Mayer (2001) "Loss aversion and seller behavior: Evidence from the housing market," *The Quarterly Journal of Economics*, Vol. 116, pp. 1233–1260.

Glickman, Mark E (1999) "Parameter estimation in large dynamic paired comparison experiments," *Applied Statistics*, pp. 377–394.

Heath, Chip, Richard P Larrick, and George Wu (1999) "Goals as reference points," *Cognitive psychology*, Vol. 38, pp. 79–109.

Hull, Clark L (1932) "The goal-gradient hypothesis and maze learning.," *Psychological Review*, Vol. 39, p. 25.

Kahneman, Daniel, Jack L Knetsch, and Richard H Thaler (1990) "Experimental tests of the endowment effect and the Coase theorem," *Journal of political Economy*, Vol. 98, pp. 1325–1348.

Kahneman, Daniel and Amos Tversky (1979) "Prospect theory: An analysis of decision under risk," *Econometrica: Journal of the econometric society*, pp. 263–291.

Kahneman, Daniel, Peter P Wakker, and Rakesh Sarin (1997) "Back to Bentham? Explorations of experienced utility," *The quarterly journal of economics*, pp. 375–405.

Kivetz, Ran, Oleg Urminsky, and Yuhuang Zheng (2006) "The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention," *Journal of Marketing Research*, Vol. 43, pp. 39–58.

Locke, Edwin A and Gary P Latham (2002) "Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.," *American psychologist*, Vol. 57, p. 705.

Locke, Edwin A, Karyll N Shaw, Lise M Saari, and Gary P Latham (1981) "Goal setting and task performance: 1969–1980.," *Psychological bulletin*, Vol. 90, p. 125.

Markle, Alex, George Wu, Rebecca J White, and Aaron M Sackett (2015) "Goals as reference points in marathon running: A novel test of reference dependence," *Available at SSRN 2523510*.

Martin, Andrew J (2006) "Personal bests (PBs): A proposed multidimensional model and empirical analysis," *British Journal of Educational Psychology*, Vol. 76, pp. 803–825.

Mellers, Barbara A, Alan Schwartz, Katty Ho, and Ilana Ritov (1997) "Decision affect theory: Emotional reactions to the outcomes of risky options," *Psychological Science*, Vol. 8, pp. 423–429.

Odean, Terrance (1998) "Are investors reluctant to realize their losses?" *The Journal of finance*, Vol. 53, pp. 1775–1798.

Ordóñez, Lisa D, Maurice E Schweitzer, Adam D Galinsky, and Max H Bazerman (2009) "Goals gone wild: The systematic side effects of overprescribing goal setting," *The Academy of Management Perspectives*, Vol. 23, pp. 6–16.

Pope, Devin and Uri Simonsohn (2011) "Round numbers as goals evidence from baseball, SAT takers, and the lab," *Psychological science*, Vol. 22, pp. 71–79.

Sedaris, David (2014) "Stepping Out: Living the Fitbit Life," *The New Yorker*.

Stoeber, Joachim, Mark A Uphill, and Sarah Hotham (2009) "Predicting race performance in triathlon: The role of perfectionism, achievement goals, and personal goal setting," *Journal of Sport & Exercise Psychology*, Vol. 31, p. 211.

Thaler, Richard H and Eric J Johnson (1990) "Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice," *Management science*, Vol. 36, pp. 643–660.

Tversky, Amos and Daniel Kahneman (1991) "Loss aversion in riskless choice: A reference-dependent model," *The quarterly journal of economics*, pp. 1039–1061.

Williams, Elanor F and Thomas Gilovich (2012) "The better-than-my-average effect: the relative impact of peak and average performances in assessments of the self and others," *Journal of Experimental Social Psychology*, Vol. 48, pp. 556–561.

Wong, Alia (2015) "The Bloated Rhetoric of No Child Left Behind's Demise," *The Atlantic*.

A 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 exertion when she does. Following Allen et al. (2016), we assume a piecewise-linear value function that jumps discontinuously at a reference point θ:

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

The jump at the reference point 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.

A.1 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 $r + \Delta > \theta$ and $r + \Delta - k \leq \theta$—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 $r + \Delta \leq \theta$,

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

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

$$\underbrace{k \cdot F'(e)}_{\substack{\text{Marginal} \\ \text{expected gain}}} = \underbrace{c'(e)}_{\substack{\text{Marginal} \\ \text{cost}}} \tag{2}$$

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 $r + \Delta \geq \theta$ but $r + \Delta - k \leq \theta$:

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

The optimal effort level satisfies the first-order condition:

$$\underbrace{(k + \epsilon) \cdot F'(e)}_{\substack{\text{Marginal} \\ \text{expected gain}}} = \underbrace{c'(e)}_{\substack{\text{Marginal} \\ \text{cost}}} \tag{4}$$

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 (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 first-order condition, $a \cdot F'(e^*) = c'(e^*)$, where $a > 0$. Then by the Implicit Function Theorem,

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

where the inequality follows from our assumptions about winning probabilities—i.e., increasing with effort at a decreasing rate—and costs—i.e., increasing with effort at an increasing rate. 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 (i.e., when $a = k + \epsilon$, rather than $a = k$).

A.2   Quitting

Now we 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 $r + \Delta > \theta$ and $r + \Delta - k \leq \theta$—i.e., the player will end above the reference point if she wins and below the reference point if she loses. And let e^* be the optimal effort when close to the reference point—i.e., the effort level that solves the first-order condition in Equation 4. Then she plays if and only if:

$$
\begin{aligned}
v(r) &< \mathbb{E}[U(e^*)] \\
&< \alpha + F(e^*) \cdot (r + \Delta + \epsilon) + (1 - F(e^*)) \cdot (r + \Delta - k) - c(e^*) \\
&< \alpha + r + \Delta - k + (k + \epsilon) \cdot F(e^*) - c(e^*) \tag{6}
\end{aligned}
$$

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:

$$\alpha > k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*) \tag{7}$$

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

$$\alpha > k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*) + \epsilon \tag{8}$$

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—in order 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. Hence, the probability that a player just below the reference point chooses to play is $1 - G_\alpha(\gamma)$, where $\gamma \equiv k - \Delta + c(e^*) - (k + \epsilon) \cdot F(e^*)$; and 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 discontinuously at the reference point by an amount equal to $G_\alpha(\gamma + \epsilon) - G_\alpha(\gamma)$.

A.3   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:

$$\alpha > k - \Delta + c(e_1^*) - (k + \epsilon) \cdot F(e_1^*), \tag{9}$$

as in (7), with e_1^* solving the first-order condition in (4). Player 2 plays if:

$$\alpha > k - \Delta + c(e_2^*) - k \cdot F(e_2^*), \tag{10}$$

with e_2^* solving the first-order condition in (2). Hence, player 1's threshold for playing is lower if:

$$(k + \epsilon) \cdot F(e_1^*) - c(e_1^*) > k \cdot F(e_2^*) - c(e_2^*), \tag{11}$$

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 than for player 2.

B   The Rating System

FICS uses the Glicko algorithm to assign a rating to every player at every point in time. Glicko was invented by Mark Glickman as an extension of the popular Elo rating system used by official chess federations (Glickman, 1999). The system 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 1720 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 $\Delta = k\left(1 - 1/\left(1 + 10^{-cD}\right)\right)$ rating points, where $c \approx 1/400$ is a constant, 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 $\Delta = \frac{3k}{4}$—i.e., who gains $\frac{3k}{4}$ rating points with a win and loses $\frac{k}{4}$ rating points with a loss—is expected to win $1 - \frac{\Delta}{k} = 25\%$ of the time.

The player's rating variance decreases as the player plays more games and increases as time elapses since her last game. This variance is used in two ways on the server: 1) to determine whether a high rating counts as a personal best (ratings only count as personal bests when the variance does not exceed 80), and 2) 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. Figure 5 shows a calibration plot, comparing predicted and actual win rates at each victory reward.[4] 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.[5]

C Data Preparation and Descriptives

Our data comprise the complete set of *blitz* games—i.e., with an expected duration between 6 and 30 minutes—played on the Free Internet Chess Server (FICS) between 2000 and 2015.

---

[4]We restrict the sample to the X million player-games with a maximum victory reward of 16, which is the value of k when the variance is less than X and applies to the majority of observations.

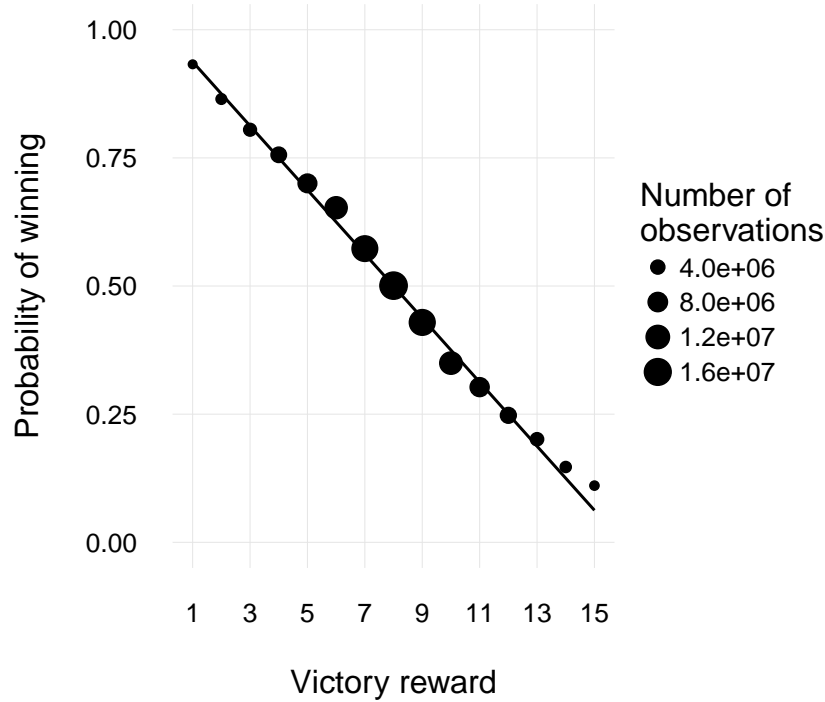[5] http://www.freechess.org/Help/HelpFiles/glicko.html

Figure 5: Predicted (line) and actual (dots) win rates at each victory reward.

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:

1. joined FICS prior to January 1, 2000;

2. is a "computer" account, which corresponds to either a bot or a player who uses chess program assistance during her games;

3. plays computer accounts as their opponent in more than 25% of their games;

4. has played fewer than 200 games so far;

5. has a rating variance greater than 80;

6. set a personal best rating in the previous 20 games.

The first restriction is to ensure that we have complete data for every player in our dataset, which starts on January 1, 2000. The second is to limit our attention to human

players. The third is to exclude players who may be abusing the rating system—players who play too many games against bots may be repeatedly exploiting known bot weaknesses in order to unfairly gain rating points. The fourth is to allow players the opportunity to set meaningful personal bests (for example, a personal best may not yet be meaningful if a player has only played two games so far). The fifth restriction is to limit our attention to player-games that could potentially count as personal bests (recall from Section B that a rating can only count as a personal best if the associated rating variance is less than 80). When a player's rating variance is above 80, the system treats her rating as too uncertain to count towards permanent records. The sixth restriction is to concentrate our attention on instances where personal bests are more likely to be meaningful goals. If a player has just recently broken her personal best, it may not be as motivating to try to break it again right away compared to when a player is approaching a personal best that she set a long time ago. Absent this restriction, our analyses are complicated by the fact that players very close to their personal best are a mix of two different populations: those who have very recently broken their personal best then lost a game, and those who haven't broken their 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.

D   Robustness Checks

D.1   Personal best recency restriction

Throughout the paper, we analyze a dataset in which we excluded player-games where the player had set a new personal best in their last 20 games. Here, we reproduce our two main empirical figures without this restriction (so that a player could have set a new personal
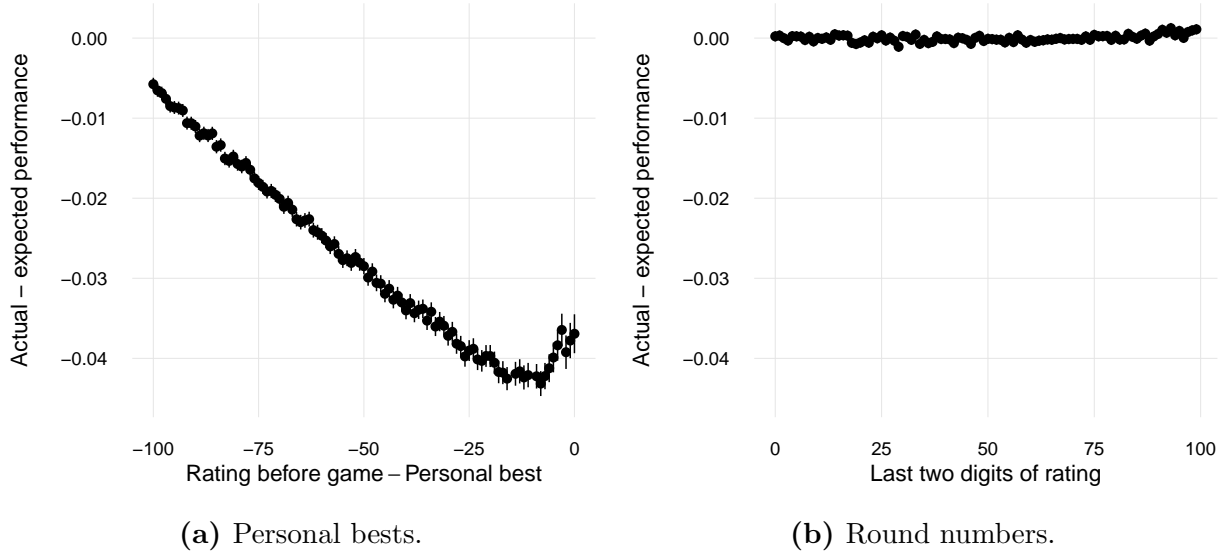
19

(a) Personal bests.　　　　　　**(b)** Round numbers.

Figure 6: Performance short of personal bests and round numbers, with 95% confidence intervals.

best at any time in the past). This expands our dataset to 220M player-games across 142M unique games. As can be seen in Figures 6 and 7, all of the results are qualitatively identical to those presented in the main text.

D.2   Definition of quitting

In Section C, we defined quitting as not playing a blitz game for at least one hour. However, the results are qualitatively the same regardless of what arbitrary threshold we choose as the definition of quitting. In Figure 8 we show how the probability of quitting varies with the distance between a player's current rating and her personal best rating prior to the previous game if we define quitting as not playing a blitz game for 24 hours. Across the personal-best reference point, the probability of quitting again jumps discontinuously—here it is a 1.5 percentage point, or 27.7%, increase. Across the round-number reference point, the discontinuous jump is a 0.6 percentage point, or 15.5% increase.
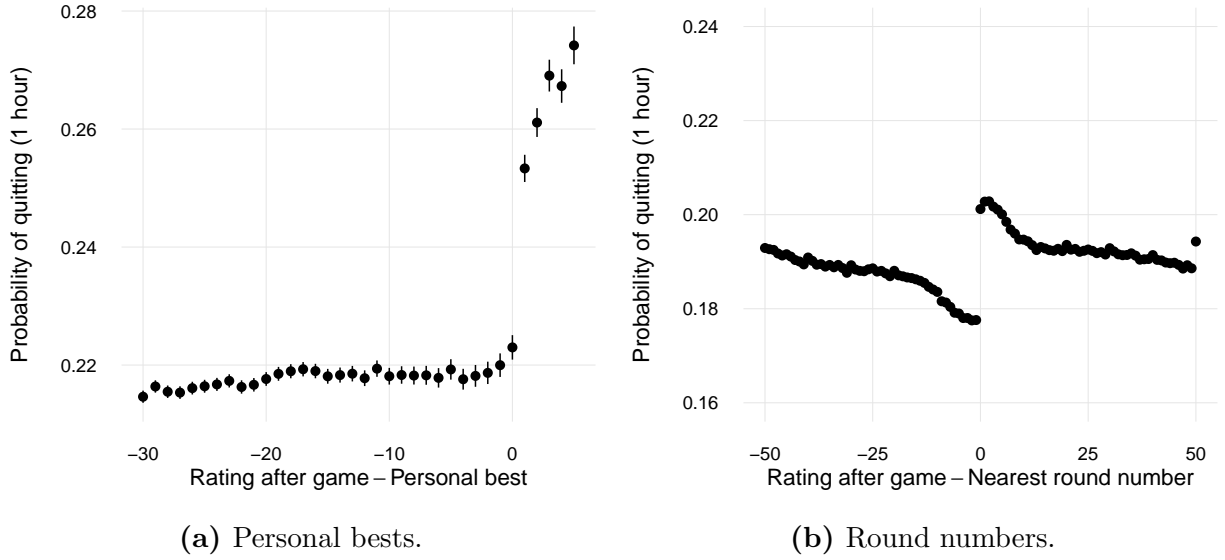
20

(a) Personal bests.

(b) Round numbers.

Figure 7: Probability of quitting for at least 1 hour, absent the personal best recency filter, with 95% confidence intervals.

(a) Personal bests.
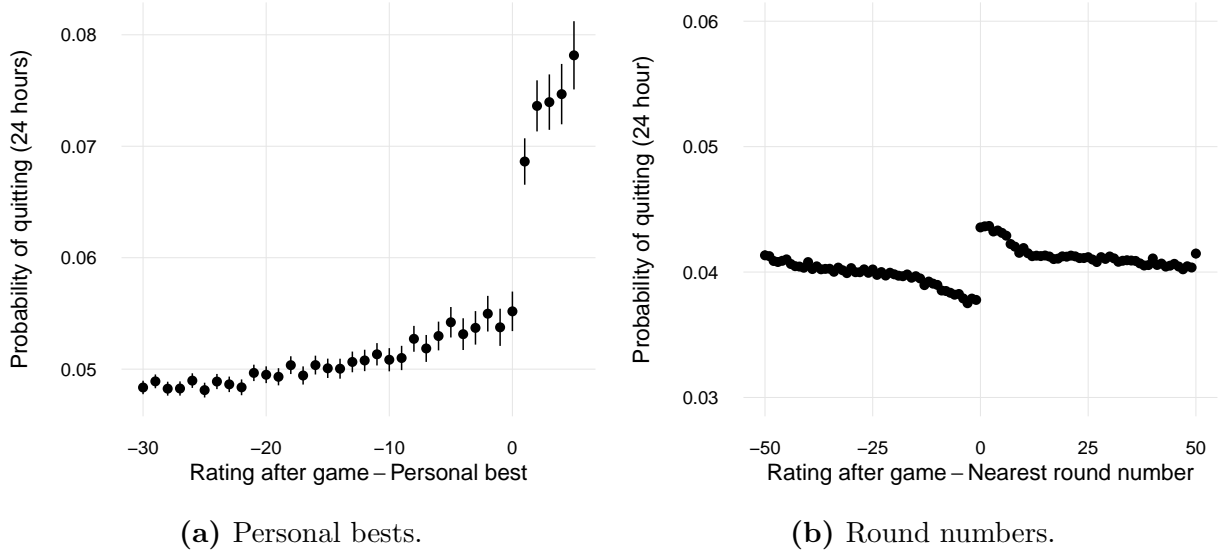
(b) Round numbers.

Figure 8: Probability of quitting for at least 24 hours, with 95% confidence intervals.