A Check for Rational Inattention

Greg Howard*

September 14, 2021

Abstract

Who is rationally inattentive, and in what situations? Using millions of online chess moves, I test whether players are rationally inattentive, by comparing the marginal benefits (better moves) and marginal costs (less time for future moves) of attention. I cannot reject that skilled players equalize marginal benefit and cost. Unskilled players, when they have little time, have higher marginal cost because they spend too long on moves. A simple intervention improves players' attention allocation.

Keywords: attention allocation, deterministic games, cognitive costs

JEL Codes: D83, D91, C72

*University of Illinois, glhoward@illinois.edu. I thank Hassan Afrouzi, Vivek Bhattacharya, Shihan Xie, and the Lichess staff for valuable discussion, and participants at the Illinois Young Applied Faculty Lunch for their comments. All errors are my own.

> Most decisions should probably be made with somewhere around 70% of the information you wish you had. If you wait for 90%, in most cases, you're probably being too slow.
>
> Jeff Bezos, '2016 Letter to Shareholders"

Rational inattention is an increasingly-popular tool in economics. It allows for agents to regularly make mistakes while maintaining the ability to calculate welfare and do counterfactual analysis. The key assumption is that agents allocate attention optimally. I investigate when this assumption is appropriate. Does rational inattention come naturally, or is it a skill? More generally, which decision-makers and in what situations is attention allocated rationally?

I study rational inattention in online chess, where players face a trade-off between using time to make a better move and losing time for future moves.[1] By measuring the marginal benefit and marginal cost, I test whether agents optimally allocate their attention. For skilled players and players with sufficient time, I am unable to reject that the marginal benefit equals marginal cost. But for unskilled players when time is limited, I find evidence that marginal benefit is below marginal cost. In other words, by spending too long when they have little time, unskilled players do not adjust their attention as much as a rationally-inattentive agent.

Testing rational inattention is difficult. First, it is unusual to observe direct measures of attention. Second, economists do not often have a way to evaluate whether an agent made the "right" choice, and even more rarely observe how bad the choice was. Third, empirical tests often require an exogenous change in the attention cost. Finally, if rational inattention is a skill, economists would like a setting in which agents have seen similar problems repeatedly, but still regularly make mistakes.

Chess meets all these criteria. The game is played with clocks, so an economist can observe how long a player spent choosing their move. Computer engines are significantly better than humans and are commonly used to evaluate moves. Time controls vary, giving

[1] Online chess is fast compared to classical chess, which readers may be familiar with due to the Queen's Gambit on Netflix. This paper focuses on games lasting between 30 seconds and 30 minutes.

plausibly exogenous variation in attention.[2] Finally, online chess attracts both casual players and the best players in the world.

I use millions of moves played on Lichess, one of the leading online chess websites, to test whether players are rationally inattentive. Because the setting is so rich, I go beyond simple tests of whether attention is valuable or whether attention is strategically allocated. Rather, I directly test whether the marginal benefit of attention, making better moves, is the same as the marginal cost of losing time for future moves.

The test requires estimating the value function of the players. I use local linear regression to estimate the empirical probability of a victory based on the time remaining for each player and a strong computer engine's evaluation of the position. This procedure also calculates the marginal cost of spending additional time. Then, matching moves across time controls, I calculate the marginal benefit of attention using the time control as an instrumental variable.

In several of the faster time controls, I reject the hypothesis that the average marginal cost and average marginal benefit are equal. Players in these time controls have a lower marginal benefit than marginal cost. Digging into these results, I reject the hypothesis of equalized marginal benefit and marginal cost specifically for unskilled players. Unskilled players spend too long on moves, across a variety of time controls. For the best players, I am unable to reject the optimality condition in any time control.

To provide complementary evidence that some players are not optimizing their allocation of attention, I show that a simple intervention by Lichess when the clock gets low—a small beep and changing the color of the clock—lowers the amount of time spent on moves and raises the average mistake size. This result is consistent with players adjusting their strategy to better optimize attention. Unskilled players react more strongly.

One thing to note about the interpretation of attention costs is that I measure attention using the time spent on a move, which is appropriate in a fast online chess game where attention is likely undivided. Relatedly, in chess, the cost of attention is the opportunity

[2]A time control is a limit on how long each player has to make all their moves in the game. If they run out of time before the game ends on the board, they lose.

cost of time for future moves. In contrast, many rational inattention models have a fixed abstract attention cost. One interpretation of the cost is that attention has an opportunity cost, which corresponds to my framework. Another interpretation is that agents get direct disutility from having to process information, which makes less sense in a game people play for fun. I discuss further in Section 4.3.

Interpreting the cost as an opportunity cost does have application to important economic decisions. As the introductory quote illustrates, firms also face a choice of acquiring additional information with significant opportunity cost (Bezos, 2016).

Related Literature

Rational inattention was first used in economics by Sims (2003). Many recent papers have derived empirical predictions of rational inattention (Matějka, 2015; Fosgerau et al., 2020; Caplin et al., 2016, 2019). And many papers have now tried to test some of these predictions in the lab (Caplin and Dean, 2015; Dean and Neligh, 2019; Dewan and Neligh, 2020; Bronchetti et al., 2020). See Section 4.4 of Maćkowiak et al. (2020) for a review. Of note, Gabaix et al. (2006) investigates the trade-off of spending time on different decisions in a laboratory.

Increasingly, economists have looked outside the laboratory for evidence of rational inattention. There is a wide variety of settings: firm returns and mutual funds (Cohen and Frazzini, 2008; Menzly and Ozbas, 2010; Kacperczyk et al., 2016), baseball (Phillips, 2017; Bhattacharya and Howard, forthcoming; Archsmith et al., 2021), online finance behavior (Mondria et al., 2010; Sicherman et al., 2016; Olafsson et al., 2018), rental search (Bartoš et al., 2016), online shopping (Taubinsky and Rees-Jones, 2018; Morrison and Taubinsky, 2020), health insurance (Brown and Jeon, 2020), migration (Porcher, 2020), and forecasting (Coibion et al., 2018; Gaglianone et al., 2020; Xie, 2019). The general finding is that agents are allocating attention in strategic ways.[3]

[3]Olafsson et al. (2018) and Bronchetti et al. (2020) find interesting deviations from rational inattention, with Olafsson et al. (2018) showing that some types of information are directly utility-enhancing, and

For most of these papers, finding evidence of rational inattention is twofold: first, establish a role for attention, i.e. "when attention costs change, does it affect the quality of decisions?;" and second, show that attention is allocated on a rational basis, i.e. "when the stakes increase, do people pay more attention?"[4] Sometimes these are combined into "are better decisions made when the stakes are high?"

The test I propose in this paper is stronger than those tests. Unlike other applications, I empirically measure both the benefit and cost of attention. With that measurement, I test not whether people adjust attention, but whether they adjust optimally.

The findings regarding skill shed light on when rational inattention is an appropriate theory. Maćkowiak et al. (2020) defend rational inattention because agents that make repeated decisions figure out which information to pay attention to. They give the example of a driver being rational in processing traffic signals because they drive regularly, but they hypothesize the same driver would be less rationally-inattentive if the car were to spin out because that is a new situation. This paper confirms the chess analog of that hypothesis by showing that skilled chess players are better at rationally allocating attention.

The results on skill also speak to the literature that takes place in the laboratory. Rejecting rational inattention for a decision in which test subjects are inexperienced or have little time may not be informative for whether they would be rationally inattentive in situations where they have more experience or time.

This paper also contributes to a literature using chess to explore economic phenomena, from backward induction (Levitt et al., 2011) to gender's effect on strategic decisions (Gerdes and Gränsmark, 2010) to risk-taking (Dreber et al., 2013) to level-k thinking (Biswas and Regan, 2015). This paper is also similar to Romer (2006) who estimates a value function to judge decision-making in football.

Bronchetti et al. (2020) showing that agents undervalue information relative to the rational inattention prediction. These results are consistent with my findings regarding unskilled chess players, who do change their allocations of attention across time controls, but not as much as they would if they were optimizing their allocation.

[4]These are the tests I perform in Appendix A.

1 Setting and Data

Chess is a deterministic game, but the number of possibilities is extensive, and it has not been solved. Still, computers are much better than humans. In fact, professional chess players use computers extensively in their preparation for games against other humans. And computers are an important tool for humans to evaluate their play and identify mistakes.

The most ubiquitous computer engine is Stockfish, and it is amongst the strongest in the world. If Stockfish sees all the way to the end of the game, it might evaluate the position as "checkmate for black in 4 moves." More commonly, Stockfish evaluates the position using pawn equivalents, e.g. "white is winning by 0.92 pawns."[5]

Lichess is one of the most popular chess websites, and millions of games are played there each month, by novices and the world's best chess players (including Magnus Carlsen, the world champion). Lichess games are available for download (Duplessis, 2017). This includes each move, the time remaining after each move, the players, the time control, and the result. For some games, they include Stockfish's evaluation of each position.[6]

I use data from April 2017, which includes about 68 million moves with the computer evaluation. Throughout the paper, I focus on moves in seven of the most popular time controls: 15 second, 30 second, 1 minute, 3 minute, 5 minute, 10 minute, and 15 minute.[7] Together these make up a majority of the moves on Lichess. I only include games that are rated and not part of a tournament. Summary statistics are presented in Table 1.

The setting selects for people interested in chess. Beyond the obvious reason that no one has to play chess, it also selects for players who review their games. My empirical strategy further selects for players that play and analyze enough games in multiple time controls

[5] A rough way to evaluate a position–commonly used by chess players without access to an engine–is to sum the point value of the pieces on the board per side, where a pawn is worth 1, bishops are worth 3, etc.

[6] To have a Stockfish evaluation included in the dataset, someone must review the game. The most common reason is that one of the players wants to see where their mistakes were. Sometimes, people also look at games in which they were not involved. Occasionally, Lichess will look at the evaluation to detect cheaters.

[7] Some time controls, which I do not use, also include an increment in which players get extra time every move.

such that I can use the exogenous variation of the time control. None of these selections are inherently problematic, but they affect the interpretation and external validity of the results. When I find that the unskilled players do not optimally allocate attention, those players are nonetheless invested in chess.[8]

One reason chess is a good place to study rational inattention is that the players understand the importance of allocating time. In Appendix A, I show that attention is valuable by regressing the strength of a move on how long it takes to make it, instrumenting using the time control to get the causal effect. I also show that players are strategic by showing the ordinary least squares version of the same regression is biased because players take longer on harder moves.

The evidence in Appendix A is also important for what it cannot do: tell us whether players are allocating attention optimally. Such a limitation is typical of empirical tests in the literature because it is rare that economists observe both the benefits and the costs of attention. I aim to overcome this limitation in the rest of the paper.

2 Theory

In chess, attention is valuable and strategically allocated (see Appendix A). However, rational inattention models make a stronger assumption: agents optimize their attention. Often, the first order conditions of such an optimization determine behavior. In this section, I propose a test based on such a first-order condition.

To model chess, I consider four state variables: whose turn it is, the position of the board, and the clocks of the two players. Denote the value function as V. On white's turn,

$$V(w, s, t^w, t^b) = \max_{S, s', \tau} \mathbb{E}[V(b, s', t^w - \tau, t^b)|S]$$

[8] A less problematic selection is that draw offers and resignations are common, so situations in which the outcome is not obvious are more common.

such that

$$\mathbb{I}(S|Q^w(s)) = \tau$$

where the first argument of V represents whether it is white or black's turn, s is the position on the board, t is the time remaining for each player. S is the signal received by agent, τ is the time they take on the move, and s' is the position of the board after they make their move. $\mathbb{I}$ is cost of the signal S based on the prior information $Q^w(s)$. The cost comes in terms of the time it takes to make the move, τ.[9]

Implicit in this notation is the assumption that the prior information, Q^w is a function of the board position only. This assumptions still allows for the signal to affect future priors, but only if it is intermediated through the move the player makes.[10] I discuss whether this is likely to bias my estimates in Section 4.3.

Black faces a similar problem, with the difference that they minimize V:

$$V(b, s, t^w, t^b) = \min_{S, s', \tau} \mathbb{E}[V(w, s', t^w, t^b - \tau)|S]$$

such that

$$\mathbb{I}(S|Q^b(s)) = \tau$$

There exist some terminal states s such that the value function is 0, 1, or in relatively rare cases, $\frac{1}{2}$, i.e. white wins, black wins, or there is a draw.[11] Furthermore, if t^w reaches 0, the value function is 0, and if t^b reaches 0, the value function is 1.

Because of the symmetry, I consider only white's problem for the rest of this section. When I take it to data, I test the first-order conditions of the agents combined.

[9]For example, in many economic models, $\mathbb{I}$ is proportional to the reduction in entropy between the prior information and the posterior (Shannon, 1948). I do not impose that $\mathbb{I}$ is Shannon.

[10]This assumption does not rule out forward-looking behavior. However, this is a stronger assumption than that agents do not acquire information that will be useful for future decisions beyond its use for current decisions (see Afrouzi and Yang (2021) for a discussion of how this relates to the linearity of information). My assumption further imposes that the whether you spend 1 second to play s' or 10 seconds to play s', they provide the same information for subsequent moves.

[11]Draws are rare in online chess (2.9 percent), unlike in classical chess.

A standard result in the rational inattention literature is that the actions are a sufficient statistic for the signal the agent receives, as there would be no reason for an agent to want any extra information. Hence I assume that $S = s'$ without loss of generality. Define

$$\vec{\pi}_s(\tau) = \arg\max_{\vec{p}(s')}\{\vec{p}(s') \cdot V(b, s', t^w - \tau, t^b) | \mathbb{I}(s', Q^w(s)) \leq \tau\}$$

where $\vec{p}(s')$ is a vector of probabilities of getting the signal s'. $\vec{\pi}$ is the best achievable strategy within time τ. The problem the white player tries to solve is:

$$V(w, s, t^w, t^b) = \max_{\tau} \vec{\pi}_s(\tau) \cdot V(b, s', t^w - \tau, t^b)$$

If $\vec{\pi}(\tau)$ is continuous and differentiable, then the first-order condition is:

$$\vec{\pi}_{s'}(\tau) \cdot \frac{\partial V}{\partial t_w}(b, s', t^w - \tau, t^b) = \frac{d\vec{\pi}_{s'}}{d\tau} \cdot V(b, s', t^w - \tau, t^b)$$

Intuitively, the left-hand side measures the expected marginal cost of time, in terms of the future value function. The right-hand side measures the marginal benefit of spending more time to have a better position. Another way to write this is

$$\mathbb{E}\frac{\partial V}{\partial t^w}(b, s', t^w - \tau, t^b) = \sum_{s'} \frac{d\pi_{s'}}{d\tau} V(b, s', t^w - \tau, t^b) \tag{1}$$

Equation (1) is a necessary condition for maximization and is the focus of my empirical tests.

3 Empirical Strategy

3.1 Estimating the Value Function

Both sides of equation (1) require estimating the value function. There are estimated to be more than 5×10^{42} possible chess positions (Shannon, 1950). So a key challenge is reducing the dimensionality. I use the Stockfish evaluation as a stand-in for s.

If s becomes a single-dimensional argument, then I can approximate $V(b, s, t^w, t^b)$ using a local linear regression. In practice, I split the evaluation into 40 bins, and each clock into 12 bins per time control.[12] For each bin, I estimate a regression

$$\text{Result}_i = \beta_0 + \beta_1 \text{Stockfish Evaluation}_i + \beta_2 t_i^w + \beta_3 t_i^b + \epsilon_i$$

on the observations i that fall into the bin, as well as bins that are adjacent to it in all three categories. Result_i is a variable that is 1 for a win, $\frac{1}{2}$ for a draw, and otherwise 0. I use the predicted value for V, and $\hat{\beta}_2$ as the marginal cost of time for observations in that bin.

I run this separately for each time control and for five bins of player skill. I drop the estimate if the bin of interest has fewer than 10 total observations. I check that the value function estimation is able to match the data in Appendix B.

One concern is that some dimensions of the position may not be captured by Stockfish: for example, the difficulty for a human. In Section 4.3, I discuss in detail how this might bias the results. To preview the argument, I find such large deviations from theory that the other dimensions of the position would have to be implausibly large to explain the results.

3.2 Variation in the Time Spent on a Move

To get at the causal effect of time on the quality of move, I use a matching strategy combined with instrumental variables.

For each move, I attempt to match it to a move in another time control. I match the player, the number of turns into the game, and the approximate computer evaluation before

[12]The evaluation is split into 38 equally sized bins in which Stockfish evaluates the position in terms of pawns. I censor estimation at 20 pawns because there are a few extreme outliers that otherwise make local regression a bad approximation. Above 20 pawns, there is not much of an effect on the probability of winning. There is also a bin for an evaluation of "checkmate for white in [x] moves" and another bin for "checkmate for black in [x] moves." For the clocks, there is significantly more curvature in the win probability close to zero seconds remaining (see Figure B.1). So for each time control, I have a bin for the first 1/60th, 1/60th to 1/30th, 1/30th to 1/20th, 1/20th to 1/15th, 1/15th to 1/10th, 1/10th to 1/5th, and then every fifth remaining of the total time control.

the start of the move. The turn and player must match exactly. I split starting evaluations into 50 bins, and they must be in the same bin.[13] The matching is done without replacement, so each move is never used more than once. For the seven time controls, I match each with the next higher time control (e.g. 3 minutes to 5 minutes), and repeat the analysis matching it with the next lower time control (e.g. 3 minutes to 1 minute).

Denote the pair of moves using i. The original move is i, c and the paired move is i, c^* where c denotes the time control and c^* is the matched time control. For most time controls, I match several hundred thousand moves. Summary statistics for matched moves can be seen in Appendix Tables C.1 and C.2. The average Elo rating of players in the matched sample is slightly higher in some but not all cases, and the average change in Stockfish evaluation is less negative.

To measure how much the quality of the move improves under a different amount of time, I evaluate the value function using the time remaining from the primary move, but the engine evaluation from the matched move. Comparing the value function of the base move to the value function of the paired move differs only on the positions on the board, and not on the clocks in the other time control. In math,

$$V_{i,c^*} = V(b, s_{i,c^*}, t^w_{i,c}, t^b_{i,c})$$

Finally, I run the two-stage least-squares regression:

$$\begin{aligned} V_{ic} &= \beta\tau_{ic} + \alpha_i + \epsilon_{ic} \\ \tau_{ic} &= \gamma \text{Time Control Indicator}_c + \delta_i + \eta_{ic} \end{aligned}$$

[13]This resembles coarsened exact matching in that I throw out many data points that are not similar, unlike nearest-neighbor matching (Iacus et al., 2012). The difference in starting evaluation across matched observations averages less than two-hundredths of a pawn in each pair of time controls, and none are statistically significant.

In Appendix Figures C.4 and C.5, I show that the results are robust to having 100 bins of starting evaluations.

α_i and δ_i are fixed effects for each pair of observations. β measures an average of the marginal benefit of spending more time on the quality of the move.

3.3 Aggregation

Across many different positions, times, and skill-levels, the marginal benefit is not constant. $\hat{\beta}$ is a weighted average of marginal benefits:

$$\hat{\beta}_{MB} = \frac{\mathbb{E}[V_{i,c^*} - V_{i,c}]}{\mathbb{E}[\tau_{i,c^*} - \tau_{i,c}]} = \sum_i w_i \hat{\beta}_i$$

where w_i is proportional to $\tau_{i,c^*} - \tau_{i,c}$. Undoing the weights to recover the average treatment effect is impossible. But because I only care about comparing marginal benefit to marginal cost, I can calculate a weighted average of the marginal cost. I measure

$$\hat{\beta}_{MC} = \sum_i w_i \frac{\partial \hat{V}}{\partial t^w}$$

where the $\frac{\partial \hat{V}}{\partial t^w}$ is the coefficient on t^w in the estimation of the value function.

With this reweighting, if agents were everywhere equating marginal benefit with marginal cost, then $\hat{\beta}_{MB} = \hat{\beta}_{MC}$ because they are weighted the same. In practice, the reweighting does not have much effect on the estimates of average marginal cost.

There is one additional consideration. In general, $\hat{\beta}_{MB}$ is likely to be smaller or larger than the true marginal benefit because the returns to attention are highly concave—which I show in the next section—and the time controls are not that close together. Matching moves to a lower time control provides an upper bound for the marginal benefit, and vice versa. Therefore, each matching strategy provides a one-sided test to reject rational inattention, where the null hypothesis is that $\beta_{MB} \leq \beta_{MC}$ if the matching is to a higher time control and $\beta_{MB} \geq \beta_{MC}$ if the matching is to a lower time control.

4 Results

4.1 Comparing Marginal Benefit and Marginal Cost

Figure 1 plots the results of $\hat{\beta}_{MB}$ and $\hat{\beta}_{MC}$ for each time control. Both panels are presented in a log-log scaling. In panel (a), moves are paired with moves from a higher time control. In panel (b), moves are paired with moves from a lower time control. Hence, theory implies the marginal benefit estimate should be below the marginal cost estimate in panel (a) and vice versa in panel (b). Panel (a) conforms with the null hypothesis. For one point estimate, the order is reversed, but the marginal cost line is well within the marginal benefit's confidence interval.[14]

In Panel (b), there are three observations in the faster time controls for which the marginal cost is higher than the marginal benefit, in contrast to theory. In these time controls, players are spending too much time to decide on a move.

Note that the marginal benefit lines are decreasing as the time control increases, consistent with the idea of concave returns to attention.

4.2 Heterogeneity by Skill

Theory predicts that $\beta_{MC} = \beta_{MB}$ not only on the entire dataset but also on any subset of the data. So in this subsection, I split the data into categories to examine whether there are deviations from rational inattention.

I divide players into five groups per time control based on their Elo rating (which reflects the results of previous games). A higher Elo rating is better. I repeat the analysis from Figure 1 for each of the five groups within each of the time controls.

The results are presented in Figure 2. Because each time control is presented with its own y-axis, there is no log-scale. In the panels on the upper row, I match moves to the

[14]The tables with the precise estimates and standard errors are in Appendix C. Of particular note, the first-stage F-statistic for the IV-regressions range from 400 to 32,000.

next-shorter time control. In the the lower row, I match moves to the next-higher time control. That means that theory would predict that the marginal cost line should be above the marginal benefit line in the top row, and the opposite in the bottom row.

In general, the top row conforms with theory. There are several notable exceptions to the theory in the bottom row. In the panels corresponding to 30 second, 1 minute, 3 minute, and 5 minute chess, worse players all have a lower marginal benefit than marginal cost. In 30 second chess, all but the highest quintile have lower marginal benefit. For 1 minute and 5 minute chess, the two lowest quintiles exhibit the same pattern, and for 3 minute chess, the lowest quintile exhibits the pattern as well. For the least-skilled players in the fastest time controls, the marginal cost is three to six times higher than the point estimate of the marginal benefit and double the largest point in the 95 percent confidence interval.

The fact that marginal benefit is far below marginal cost for low-Elo players means that rationally allocating attention is associated with skill. Unskilled players are not able to equalize marginal benefit and marginal cost. Skilled players are at least close enough so as not to be detectable using my empirical strategy. This is not to say that unskilled players do not respond partially to changes in marginal cost. Across time controls, the marginal benefit of moves is still generally increasing as the time controls get faster (remember that the scale of the figures is changing across panels in Figure 2). It is just not increasing as fast as the marginal cost is.

High-Elo players differ from low-Elo players in a variety of ways. Unsurprisingly, they are better at chess, and are more likely to win, conditional on the opponent's skill.[15] However, they also play more. Looking at only the 3-minute time control, a player in the top quintile plays more than twice as many moves in a given month as a player in the bottom quintile. Unfortunately, the data does not allow me to disentangle which features of skilled players make them better at allocating attention.

[15] The fact that unskilled players are optimizing attention suboptimally is not why they are worse. A player in the bottom quintile in 15 minute games makes larger mistakes as a player in the top quintile in a 30 second game, as measured by the average change in Stockfish evaluation. So allocating time better within a 30 second game is not going to make a bottom quintile player as good as a top quintile player.

Appendix Figure 2 carries out a similar heterogeneity analysis by the amount of time on a player's clock, confirming the result in Figure 1 that the deviations are found when players have little time.

4.3 Discussion of Potential Bias

In this section, I consider whether players playing optimally would necessarily satisfy the tests I propose. It should be noted that for many of the results, the marginal cost was several times larger than the marginal benefit. While some of the following concerns could bias the estimates on the margin, they are too small to explain the large differences from theory.

Dimensions of the Position Not Captured by the Stockfish Evaluation

One potential threat is that the Stockfish evaluation is not a sufficient summary of the position on the board to effectively calculate win probabilities. The main challenge is that the complexity of the game is not something the engine considers, but does matter for who wins the game. For example, when a player has an advantage, simplifying the position by trading off pieces is generally considered advantageous. A computer evaluating the position might not consider this, and so it does not factor into my measure of the marginal benefit.

The analysis in Appendix Table B.1 shows that the gains in predictiveness of win probability when accounting for such a strategy is quite small compared to the predictiveness based on the Stockfish evaluation and the clocks. Including a dummy variable for a capture flexibly interacted with the clocks of both players and the computer evaluation only improves the R^2 of a regression of winning the game on my value function by a tiny fraction. If playing such a move only matters a tiny bit for the win probability, it shows that a player cannot gain much benefit from such a move, meaning that even if they are more likely to play moves that might help them along non-Stockfish dimensions, the amount of bias this leads to must be quite small compared to the size of the marginal benefit based on Stockfish.

Appendix B also looks at the idea of putting the king in check—which constrains the number of moves for the opponent—and makes a similar conclusion.

Corner Solutions

Another possible objection is that if the solution is not interior, the first-order condition may not be a necessary condition for optimality. If players are constrained to take a certain amount of time by the physical amount of time it takes to click,[16] then it could be that the marginal benefit is lower than the marginal cost.

However, the estimated coefficient is a weighted average of marginal benefits, based on how much faster you make the moves when time controls are shorter. So if a player was constrained, that move would get zero weight. Hence, the average is based only on moves in which the player is not constrained.

Having Already Thought Through the Move

Another objection is that the player could have already spent time thinking through the move. In a longer time control, there is potential for more unmeasured time spent considering the move. This is likely to be a small bias, as the player must usually think about many possible positions ahead of time. Moreover, this would mean that I am overestimating the marginal benefit, so this bias could not explain the deviation from rational inattention.

Thinking Ahead

A related story is that players think several moves into the future, and since this store of knowledge is not included in the state-space, the marginal benefit might be underestimated. This is the opposite concern from the previous one, and like that, I think it will be small. The first reason is theoretical: under rational inattention, it is suboptimal to spend time processing information that is not relevant for the current move when the costs are linear,

[16] Or if they take zero time because they "pre-moved," which means to have already clicked on a move before their opponent made their move.

as you can always think about future moves when they arrive.[17]

Nonetheless, a signal that a player should play a certain move might indicate that a subsequent move is good. However, at these time controls, players are unlikely to plan moves for far ahead that are different than their initial instincts in that position. For example, playing a move that forks the queen and the king is not going to bias the results because taking the queen is obvious.

Further, the heterogeneity suggests this is not the operative story. If planning ahead was an important confounder, it would occur more for skilled players and when players have more time. But it is the unskilled players and the players with little time that exhibit the deviations from rational inattention.

Learning

A final concern is agents sacrificing time to become better chess players, biasing the estimated marginal benefit down. However, players have unlimited time after the game for analysis. In fact, because someone has to analyze the game to be included in the dataset, the players in my data are more likely to use this option.

This argument extends to other costs and benefits of attention outside the model. For example, economists sometimes assume that cognitive effort is directly utility-lowering. Explaining my empirical results would require the opposite assumption: that cognitive effort is directly utility-enhancing. This assumption may be plausible for chess, which is volunatry and fun. However, it is hard to square with why players devote too much attention during the game rather than afterward.

5 Improving the Allocation of Attention

Lichess reminds players that their time is running low by playing a beeping noise and changing the color of the clock that the player sees on their screen from black to red. This reminder

[17] Linearity means that the cost of a signal about a future move is the same now or later.

occurs at a set time: in 1 minute chess, it is when the player's clock is at 10 seconds.

This section shows that players take shorter time and make larger mistakes after this intervention. Because players were previously taking too long when they had little time, this analysis suggests that the intervention improves their allocation of attention. It adds to the evidence that players are allocating attention sub-optimally because they make better choices when given a simple reminder.[18]

The nature of this intervention does not allow for a regression discontinuity design. That is because the beep occurs at 10 seconds in 1 minute games, and the time at which players make their move is endogenous. As expected, I observe significant bunching in the number of moves made at times just past the threshold rather than just before it (Figure 3a).[19] Consistent with previous results that marginal cost exceeds marginal benefit for unskilled players, I show the bunching is stronger for lower-rated players in Panels (d) and (e).[20]

Even though I cannot do a regression discontinuity analysis, the data still tells us about the players' behavior in response to the beep. The bunching itself reveals that the beep reminds players that the marginal cost of their time is high. Figure 3b shows the average amount of time spent on a move, binned by the number of seconds remaining when they make the move. Moves made with 9 seconds remaining are not that much shorter because some of those moves are made after a player has thought for awhile, and then the beep reminds them how high the marginal value of time is. But there is a large drop-off between 9 and 8 seconds, suggestive that players do move much faster after having heard the beep.

Figure 3c shows the change in the computer evaluation. Players make larger mistakes on average after they hear the beep.[21] The intervention is presumably beneficial to players

[18] Antioch (2020) also shows players blunder more after the beep. Antioch argues the beep is a bad thing, but my previous results argue that speeding players up is beneficial. A few things that differentiate our analysis is that Antioch defines blunders in a binary way and does not look at bunching or the time spent on the move.

[19] I only consider moves where the computer evaluation before the player's turn is less than a 5 pawns advantage. I show robustness for 30 seconds in Appendix Figure C.2.

[20] All five quintiles can be seen in Appendix Figure C.3.

[21] The coefficient is 0.37 at the discontinuity, and the robust standard error is 0.06. With player fixed effects, the coefficient and robust standard error are 0.34 and 0.06.

because it causes them to speed up. Based on the analysis from the previous section, the marginal benefit of making better moves is smaller than the marginal cost of time. So even though they make larger "mistakes," the additional speed helps their win probability.

6 Conclusion

When and for whom is rational inattention a good assumption? When there is sufficient time and for agents skilled at the task.

References

Afrouzi, Hassan and Choongryul Yang, "Dynamic Rational Inattention and the Phillips Curve," 2021.

Antioch, "chess-blunders," 2020. `https://github.com/Antiochian/chess-blunders`.

Archsmith, James E, Anthony Heyes, Matthew J Neidell, and Bhaven N Sampat, "The Dynamics of Inattention in the (Baseball) Field," Technical Report, National Bureau of Economic Research 2021.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka, "Attention Discrimination: Theory and Field Experiments with Monitoring Information Acquisition," *American Economic Review*, 2016, *106* (6), 1437–75.

Bezos, Jeff, "2016 Letter to Shareholders," 2016.

Bhattacharya, Vivek and Greg Howard, "Rational Inattention in the Infield," *American Economic Journal: Microeconomics*, forthcoming.

Biswas, Tamal and Kenneth Regan, "Measuring level-k reasoning, satisficing, and human error in game-play data," in "2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)" IEEE 2015, pp. 941–947.

Bronchetti, Erin T, Judd B Kessler, Ellen B Magenheim, Dmitry Taubinsky, and Eric Zwick, "Is Attention Produced Rationally?," Technical Report, National Bureau of Economic Research 2020.

Brown, Zach Y. and Jihye Jeon, "Endogenous Information Acquisition and Insurance Choice," Technical Report, University of Michigan 2020.

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

_ , John Leahy, and Filip Matějka, "Rational Inattention and Inference from Market Share Data," Technical Report, CERGE-EI 2016.

_ , Mark Dean, and John Leahy, "Rational Inattention, Optimal Consideration Sets, and Stochastic Choice," *Review of Economic Studies*, 2019, *86* (3), 1061–94.

ChessBase News, "Time is precious for the Ambassador of Blancpain," 2003.

Cohen, Lauren and Andrea Frazzini, "Economic Links and Predictable Returns," *Journal of Finance*, 2008, *63* (4), 1977–2011.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar, "How Do Firms Form Their Expectations? New Survey Evidence," *American Economic Review*, 2018, *108* (9), 2671–2713.

Dean, Mark and Nathaniel Neligh, "Experimental Tests of Rational Inattention," Technical Report, Columbia 2019.

Dewan, Ambuj and Nathaniel Neligh, "Estimating Information Cost Functions in Models of Rational Inattention," *Journal of Economic Theory*, 2020, *187*, 1–32.

Dreber, Anna, Christer Gerdes, and Patrik Gränsmark, "Beauty queens and battling knights: Risk taking and attractiveness in chess," *Journal of Economic Behavior & Organization*, 2013, *90*, 1–18.

Duplessis, Thibault, "lichess.org open database," 2017.

Fosgerau, Mogens, Emerson Melo, André de Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *International Economic Review*, 2020, *61*, 1569–89.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg, "Costly information acquisition: Experimental analysis of a boundedly rational model," *American Economic Review*, 2006, *96* (4), 1043–1068.

Gaglianone, Wagner Piazza, Raffaella Giacomini, João Victor Issler, and Vasiliki Skreta, "Incentive-Driven Inattention," *Journal of Econometrics*, 2020.

Gerdes, Christer and Patrik Gränsmark, "Strategic behavior across gender: A comparison of female and male expert chess players," *Labour Economics*, 2010, *17* (5), 766–775.

Iacus, Stefano M, Gary King, and Giuseppe Porro, "Causal inference without balance checking: Coarsened exact matching," *Political analysis*, 2012, pp. 1–24.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp, "A Rational Theory of Mutual Funds' Attention Allocation," *Econometrica*, 2016, *84* (2), 571–626.

Levitt, Steven D, John A List, and Sally E Sadoff, "Checkmate: Exploring backward induction among chess players," *American Economic Review*, 2011, *101* (2), 975–90.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt, "Rational Inattention: A Review," Technical Report, CERGE-EI 2020.

Matějka, Filip, "Rationally Inattentive Seller: Sales and Discrete Pricing," *Review of Economic Studies*, 2015, *83* (3), 1125–1155.

Menzly, Lior and Oguzhan Ozbas, "Market Segmentation and Cross-Predictability of Returns," *Journal of Finance*, 2010, *65* (4), 1555–1580.

Mondria, Jordi, Thomas Wu, and Yi Zhang, "The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data," *Journal of International Economics*, 2010, *82* (1), 85–95.

Morrison, William and Dmitry Taubinsky, "Rules of Thumb and Attention Elasticities: Evidence from Under- and Overreaction to Taxes," Technical Report, UC Berkeley 2020.

Olafsson, Arna, Michaela Pagel, Olivier Coibion, and Yuriy Gorodnichenko, "The ostrich in us: Selective attention to personal finances," *NBER working paper*, 2018, *23945.*

Phillips, David C, "Stopping on Nine: Evidence of Heuristic Managerial Decision-Making in Major League Baseball Pitcher Substitutions," *Southern Economic Journal*, 2017, *84* (2), 577–599.

Porcher, Charly, "Migration with Costly Information," Technical Report, Dartmouth University 2020.

Romer, David, "Do Firms Maximize? Evidence from Professional Football," *Journal of Political Economy*, 2006, *114* (2), 340–365.

Shannon, Claude E, "A mathematical theory of communication," *The Bell system technical journal*, 1948, *27* (3), 379–423.

_ , "XXII. Programming a computer for playing chess," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1950, *41* (314), 256–275.

Sicherman, Nachum, George Loewenstein, Duane J Seppi, and Stephen P Utkus, "Financial Attention," *Review of Financial Studies*, 2016, *29* (4), 863–897.

Sims, Christopher A, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50* (3), 665–690.

Taubinsky, Dmitry and Alex Rees-Jones, "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment," *Review of Economic Studies*, 2018, *85* (4), 2462–2496.

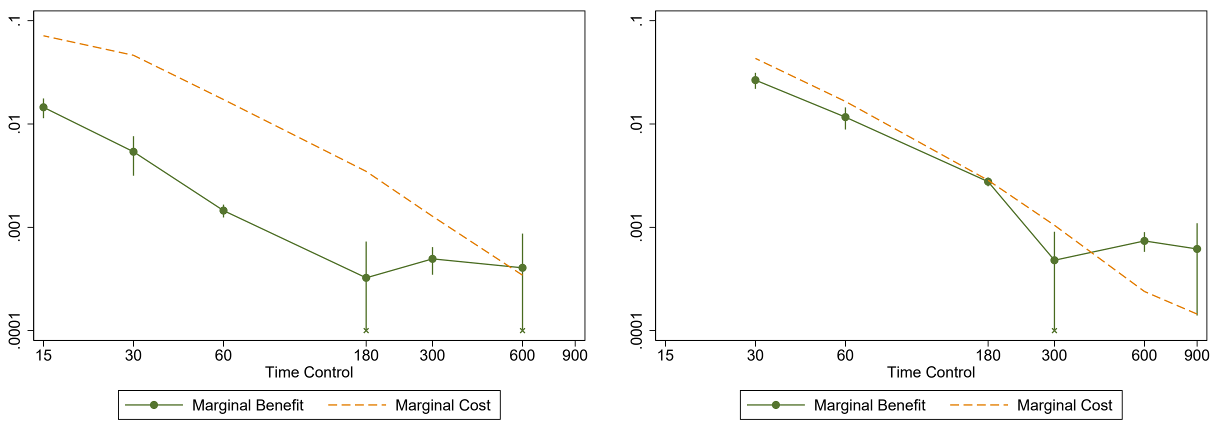
Xie, Shihan, "An Estimated Model of Household Inflation Expectations: Information Frictions and Implications," 2019.

Tables and Figures

Table 1: Summary Statistics

Time Control	(1) Number of Observations	(2) Unique Players	(3) Unique Games	(4) Mover's Elo Rating	(5) Time Spent (Seconds)	(6) Stockfish Evaluation (Pawns)	(7) Stockfish Eval. Change (Pawns)
15 seconds	446,760	3,098	8,529	1601	0.47	0.85	-1.75
				(216)	(0.62)	(8.51)	(3.37)
30 seconds	329,770	2,538	6,076	1836	0.87	0.57	-1.23
				(268)	(0.92)	(7.36)	(2.91)
1 minute	5,953,052	20,169	101,917	1592	1.60	0.54	-1.13
				(269)	(1.45)	(7.61)	(2.78)
3 minutes	4,210,627	24,771	67,774	1758	3.77	0.46	-0.92
				(320)	(4.64)	(7.02)	(2.52)
5 minutes	6,199,133	37,914	101,148	1563	6.08	0.48	-0.96
				(283)	(7.34)	(7.36)	(2.59)
10 minutes	6,853,468	42,954	112,429	1571	10.34	0.49	-1.01
				(268)	(12.94)	(7.77)	(2.65)
15 minutes	1,054,168	9,766	16,033	1664	14.13	0.44	-0.87
				(239)	(19.11)	(7.42)	(2.41)

Notes: Reported values in columns (4) to (7) are means, with standard deviations in parentheses. Stockfish evaluation is censored at $+20$ and -20 pawns. Stockfish Evaluation is at the start of the turn, from the perspective of the moving player. The change is how much the evaluation changes as a result of the turn. Source: Lichess



(a) Lower Bound of Marginal Benefit

(b) Upper Bound of Marginal Benefit

Figure 1: **The average marginal benefit and marginal cost of spending additional time on a move.** Intervals are 95 percent confidence intervals, censored at .0001. Censoring is denoted by an X at the bottom of the interval. Tables of these results, including precise estimates and standard errors can be found in Appendix Tables C.4 and C.6. Time control on the x-axis is measured in seconds, e.g. 60 corresponds to a 1-minute time control. The rational inattention prediction from Section 3 is that the marginal benefit should fall below the marginal cost in Panel (a) and should be above marginal cost in Panel (b).

Source: Lichess, author's calculations.

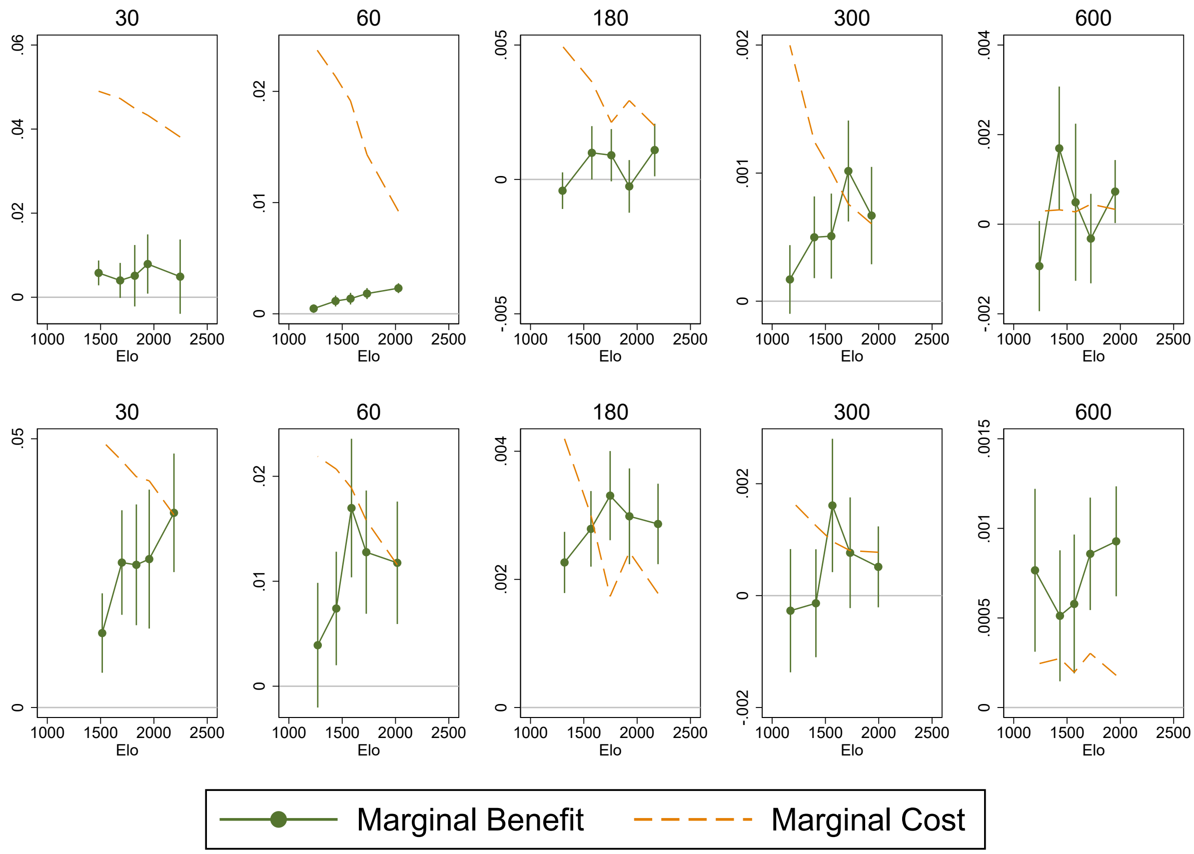
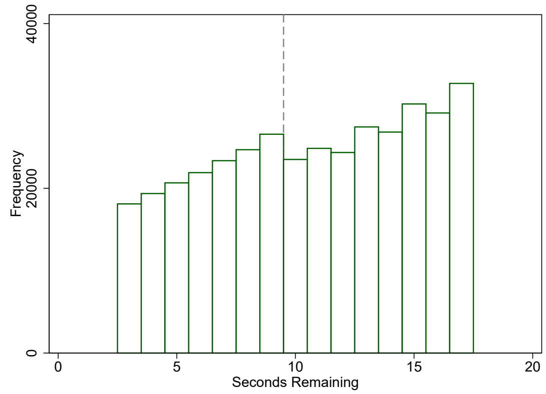
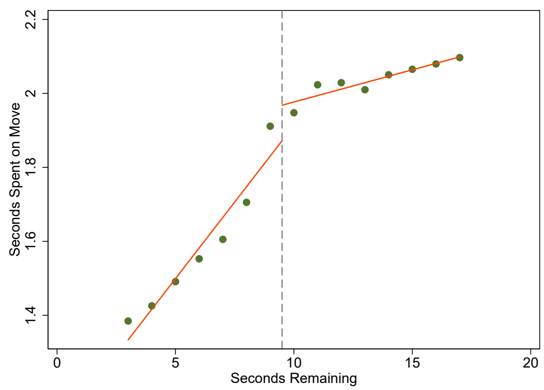


Figure 2: **Heterogeneity in Marginal Benefit and Marginal Cost by Elo rating.** The rational inattention prediction is that the marginal benefit should fall below the marginal cost in the top panels and vice versa in the bottom panels. The number above the panel indicates the seconds in the baseline time control.

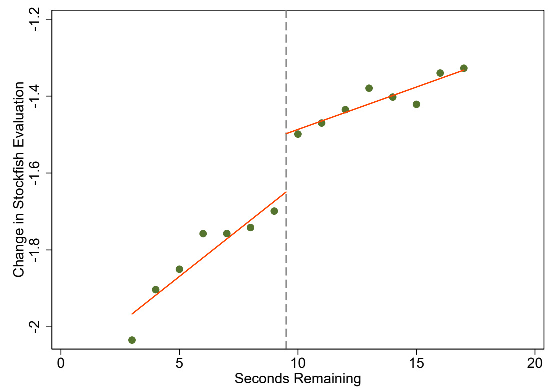
Source: Lichess, author's calculations.



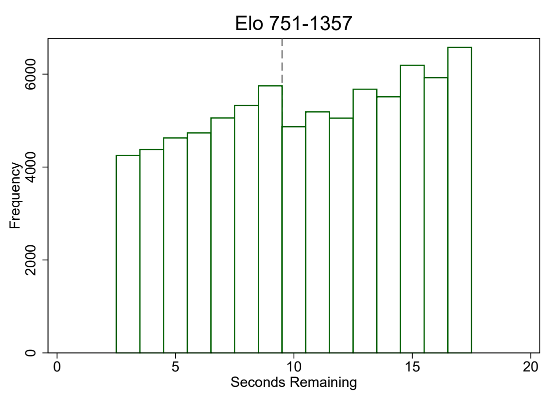
(a) The Distribution of Time Remaining After a Move



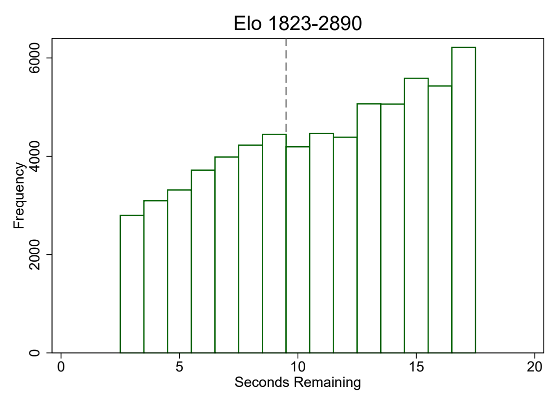
(b) The Amount of Time Spent



(c) The Change in Computer Evaluation



(d) The Distribution of Time Remaining After a Move, Unskilled Players



(e) The Distribution of Time Remaining After a Move, Skilled Players

Figure 3: **The Low-Time Reminder.** Panel (a) shows a histogram of the number of moves played by the number of seconds remaining in the 1 minute time control. The number of seconds remaining is measured discretely, and only 3 seconds to 17 seconds are plotted. The grey dashed line indicated the time at which the beep occurs. Panels (b) and (c) show the average amount of time spent and the change in computer evaluation, by the number of seconds remaining at the end of the move. Panels (d) and (e) show the same as Panel (a), but only for the lowest and highest quintile of the mover's Elo.

Source: Lichess, author's calculations.

A The Benefit and Allocation of Attention

To those experienced with chess, it is unsurprising that additional attention improves the quality of play or that players strategically allocate their time. In the words of former world champion Vladimir Kramnik, "time is precious when you don't have enough of it" (ChessBase News, 2003). Nonetheless, in this appendix, I present econometric evidence to demonstrate these two facts.

I consider the relationship between the time spent on and the quality of a move. As a naive measure of the quality of the move, I use the change in the Stockfish evaluation of the position from before and after the move. Computers anticipate the best move, so if the player plays that, the evaluation changes only slightly. Therefore, the vast majority of evaluation changes are negative. A big negative evaluation change would occur when the players makes a big mistake.

I run the regression

$$\text{Stockfish Evaluation Change}_i = \beta\tau_i + \text{Player-Turn Fixed Effects} + \epsilon_i$$

where τ_i is the time spent on the move. I run the regression separately on moves in the 3 minute time control and in the 5 minute time control. I only include moves where the computer, before the move is made, evaluates the position to be within 5 pawns of being even.

I also run this regression on the pooled 3- and 5-minute sample, instrumenting for τ_i using an indicator variable for the 5 minute time control.

The results of all three regressions are presented in Table A.1. The instrumental-variables regression has a positive and statistically significant effect.[22] These results would suggest that the causal effect of spending an extra second on your move is worth about 0.03 pawns.

[22] The F-statistic indicates weak instruments are not an issue. The first-stage regression shows that players spend 1.7 additional seconds on a move in the 5 min time control compared to the 3 minute time control. For comparison, the average time spent on a move in 3 minute time control is 3.7 seconds.

Table A.1: Motivating Facts

	(1) Evaluation Change	(2) Evaluation Change	(3) Evaluation Change
Seconds Spent	-0.0375	-0.0230	0.0306
	(0.000263)	(0.000146)	(0.00196)
Observations	3886279	4938789	9129274
F-statistic			30753.3
Time Control	3 min	5 min	3 and 5 min
Fixed Effects	Player-Turn	Player-Turn	Player-Turn
Instrument	–	–	Time Control

Notes: Evaluation change measures the difference in the Stockfish evaluation of the position at the end of the move, minus its evaluation at the start of the move. Seconds spent is measured using the difference in the clock from the start to the end of the move. In column (3), the seconds spent on the move is instrumented using an indicator variable for being in the 5-minute time control. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Lichess

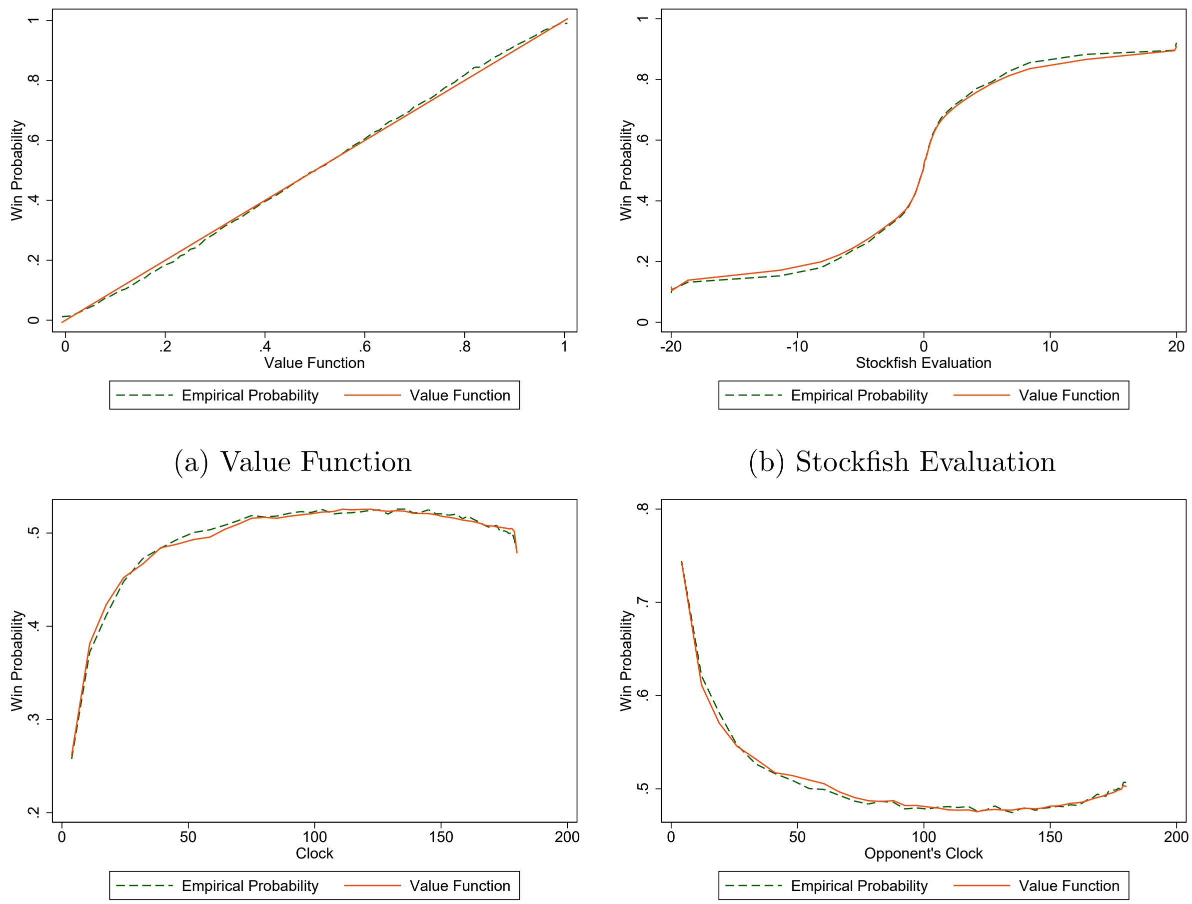
Nonetheless, there is a negative correlation between time spent and the engine evaluation: for moves on which a player spends an extra second, the resulting position is 0.02 or 0.04 pawns worse. The most likely reason that the OLS and the IV regressions give different answers is that there is an important omitted variable in the OLS that biases it away from the causal effect, namely the difficulty of the move. Players spend longer considering harder moves, and then end up playing worse moves because it is harder.

These results establish that chess is a good place to study how players allocate attention. First, it establishes that attention is valuable, by the positive sign of the IV-coefficient. Second, it establishes that players are making choices on how to allocate attention based on the difficulty of the move, by the negative OLS-coefficient.

B Checking the Fit of the Value Function Estimation

In this section, I check how well the value function estimation works. First, I look at the marginal probability of winning across each state variable. Second, I check to see if the distributions of the value function and the marginal value of time look as I expect them to.

Throughout this section, I show the estimated value function on the 3 minute time control, but it looks qualitatively similar for all time controls.



(a) Value Function

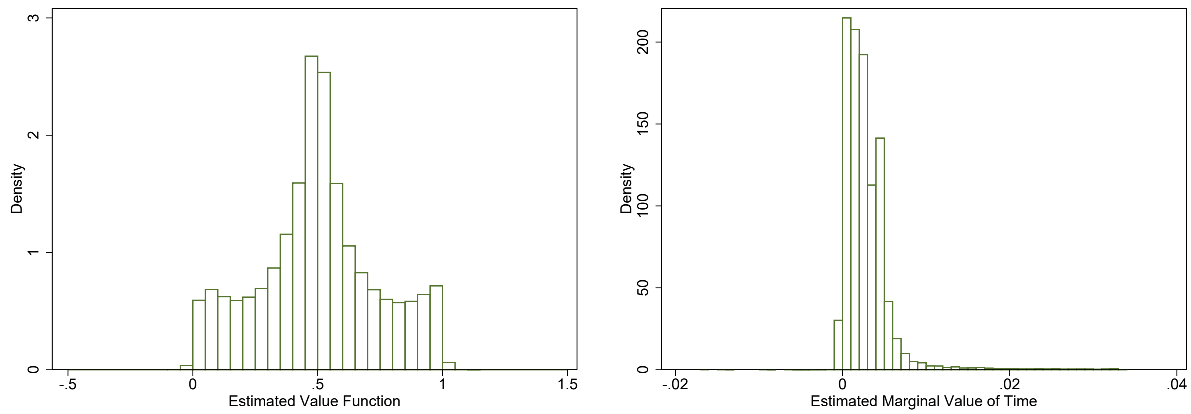
(b) Stockfish Evaluation

(c) Time

(d) Opponent's Time

Figure B.1: **Checking the Fit of the Value Function Estimation.**The green dashed line is the empirical relationship between the game result the x-axis variable. All observations are split into 50 equal-sized bin of the x-variable, and the average of the x-value is plotted against the average value of the game result in that bin, where wins are 1, losses are 0, and draws are 1/2. The orange solid line plots the same, but with the average value function instead of the average game result. Data is from the 3-minute time control. For the Stockfish Evaluation, evaluations of "checkmate in [x] moves" are not graphed.
Source: Lichess, author's calculations

In Figure B.1, I plot the average value of the game result alongside the average value of our estimated evaluation for 50 bins of the value function, 50 bins of the Stockfish evaluation, 50 bins of the player's clock, and 50 bins of the opponent's clock. They seem to match fairly well, suggesting that the local regression strategy is able to adequately capture the curvature



(a) Value Function

(b) Marginal Value of Time

Figure B.2: **Distributions of the Estimated Value Function.** Panel a presents a histogram of the estimate value function, with bins 0.05 across. Histogram is trimmed at -0.5 and 1.5 (a total of nine points are trimmed). Panel b presents a histogram of the estimate of the marginal value of time, with bins 0.001 across. Histogram is trimmed at the 99th percentile. Number of observations: 4,239,728.
Source: Lichess, author's calculations.

of the value function. Note that the downward sloping part of Panel (c) or the upward sloping part of Panel (d) does not imply that time has negative value, as time is correlated to the Stockfish Evaluation.[23]

In Figure B.2, I show the distribution of the estimated value function, and the estimated marginal value of time. Although nothing constrains it to be so, there are few values outside of the [0,1] range. The marginal value of time is almost everywhere positive, and it peaks at a number close to 0 and has a long right-tail. I attribute the facts that a small minority of value function estimates lie outside of $[0, 1]$ and that the marginal value is sometimes negative to measurement error. The tests I propose take averages over many data points, so the measurement error will wash out by the law of large numbers.

The last thing that I want to check is whether the Stockfish evaluation is an adequate proxy for the position. Specifically, because the chess position is many dimensional, it is not obvious that the Stockfish evaluation captures most of the information about the position that could be useful to evaluate who is likely to win.

[23]In Figure B.2, few of the observations are estimated to have a negative marginal value of time.

In Table B.1, I check whether a few of the observable things about the position add to the predictive power of our value function. In column (1), I regress the result of the game (0, $\frac{1}{2}$, or 1) on the value function. This explains about a quarter of the variance in the result of the game, and the coefficient is close to 1. In column (2), I add a regressor for whether the move played was a capture, and in column (3), I add a regressor for whether it was a check, i.e. threatened the opposing king. Both of these are potentially another dimension on which a player might seek an advantage: in the first case, the player is simplifying the game, as fewer pieces typically make the game easier to evaluate; and in the second case, the player is limiting the other player's possible moves, as when a player is in check, they must play a move that takes them out of check.[24]

In both of these columns, the additional variable is statistically significant, which is not surprising with over 4 million observations. However, they do not explain a significant amount of the variation. The R^2 increases from 0.24526 to 0.24537 or 0.24532, so about one-hundredth of a percent of the variation in outcomes.

In column (4), I add a dummy for each possible move notation,[25] which includes which type of piece moved, which square they moved to, whether there was a piece captured, and whether the move was made with check. This has over 4000 coefficients, so I do not report them, but even then, the R^2 increases by only one-tenth of a percent.

In columns (5), (6), and (7), I interact the regressors in Columns (2), (3), and (4) with the same bins of the player's clock, the opponent's clock, the Stockfish evaluation, and the Elo rating of the player.[26] This could be important if putting the opponent in check was valuable but only in certain situation, such as when the opponent was low on time. In columns (5)

[24] At some levels of skill, there may also be a psychological factor in which the opposing player finds being placed in check to be intimidating. With little time remaining, it is also sometimes a strategy to place the other player in check when they do not anticipate it because it will take them longer to respond than if they are able to play any move.

[25] An example would be "Qxg3+", where the "Q" represents that the queen moved, the "x" represents that a piece was captured, "g3" is the square the piece moved to, and "+" indicates that the move was with check.

[26] Recall that there are 12 bins for each clock, 40 bins for the evaluation, and 5 bins for the Elo rating. Some of these bins are sparsely populated.

Table B.1: Sufficiency of the Value Function

	(1) Game Result	(2) Game Result	(3) Game Result	(4) Game Result	(5) Game Result	(6) Game Result	(7) Game Result
Value Function	1.029	1.032	1.032	1.033	1.039	1.038	1.013
	(0.000877)	(0.000882)	(0.000890)	(0.000913)	(0.00559)	(0.00559)	(0.00840)
Capture		-0.0119					
		(0.000486)					
Check			-0.0156				
			(0.000844)				
Constant	-0.0146	-0.0129	-0.0150				
	(0.000485)	(0.000491)	(0.000486)				
Observations	4239397	4239397	4239397	4238214	4239075	4238273	3350848
R^2	0.24526	0.24537	0.24532	0.24636	0.24872	0.24868	0.31237
Adjusted R^2	0.24526	0.24536	0.24532	0.24567	0.24512	0.24532	0.18136
Fixed Effects				Move Notation	Capture × Bins	Check × Bins	Notation × Bins
Num of Non-Dropped Fixed Effects				3874	20210	18851	536274

Notes: Game result is 1 for a win, 0 for a loss, and 1/2 for a draw. "Value Function" is the value function described in Section 3. "Capture" and "Check" are dummy variables for capturing an opponent's piece and putting the opponent's king in check, respectively. Move notation includes indicator variables for the way the move is denoted which includes information on which type of piece is moved, the square it is moved to, whether a piece is captured, and whether the king is checked. In columns (5)-(7), those variables are interacted with the bins of computer evaluation, clock, opponent's clock, and Elo rating, as described in the text. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Source: Lichess, author's calculations.

and (6), I am estimating over 20,000 additional coefficients, but the R^2 goes up only about three-tenths of one percent.[27] Finally, in column (7), where about a fifth of the sample is dropped because there is only one observation within the bin, and where there are 637,000 fixed effects where the observations are not dropped, the R^2 finally increased significantly, by about 7 percent. But much of this is likely because there are simply so many fixed effects: the adjusted R^2 has actually decreased by 6 percent.

This analysis confirms that the computer evaluation and the clocks are capturing the vast majority of the variation in who is likely to win the game. Of course, there are additional dimensions of the position that can help determine the win probability, but based on the fact that several of the most likely components had negligible impacts on the predictive power of the value function, I am confident that the computer evaluation and the clocks are capturing a large majority of the variation.

[27] Note that the adjusted R^2 goes up less.

C Appendix Figures and Tables

Tables C.1 and C.2 report the same summary statistics as Table 1, but only on the subsample of the data that is matched. As expected, the number of observations, unique players, and unique games are substantially lower. In some cases, the average Elo rating is higher, though not universally. Time spent is similar to the full sample. Of note, the change in Stockfish evaluation is less negative in the matched samples (column 7). Players playing and analyzing games in multiple time controls are probably slightly higher skilled. It is also more likely to match moves near the start of the game, when evaluations tend to change less.

Table C.1: Summary Statistics for Matched Data (Lower Time Control)

Time Control	(1) Number of Observations	(2) Unique Players	(3) Unique Games	(4) Mover's Elo Rating	(5) Time Spent (Seconds)	(6) Stockfish Evaluation (Pawns)	(7) Stockfish Eval. Change (Pawns)
30 seconds	21,140	825	3,750	1632	0.76	0.35	-0.87
				(219)	(1.05)	(6.59)	(2.22)
1 minute	30,125	936	7,147	1707	1.36	0.59	-0.88
				(257)	(1.34)	(6.11)	(2.75)
3 minutes	128,628	5,140	22,127	1638	3.52	0.31	-0.68
				(310)	(4.54)	(6.06)	(2.06)
5 minutes	71,422	4,645	16,926	1644	5.51	0.15	-0.55
				(296)	(7.11)	(4.92)	(1.72)
10 minutes	136,626	7,225	30,062	1494	9.46	0.24	-0.50
				(269)	(11.89)	(4.66)	(1.65)
15 minutes	27,781	2,356	5,749	1692	12.89	0.14	-0.43
				(246)	(17.95)	(4.00)	(1.28)

Notes: Reported values in columns (4) to (7) are means, with standard deviations in parentheses. Stockfish evaluation is censored at $+20$ and -20 pawns. Stockfish Evaluation is at the start of the turn, from the perspective of the moving player. The change is how much the evaluation changes as a result of the turn. Source: Lichess

Table C.2: Summary Statistics for Matched Data (Higher Time Control)

Time Control	(1) Number of Observations	(2) Unique Players	(3) Unique Games	(4) Mover's Elo Rating	(5) Time Spent (Seconds)	(6) Stockfish Evaluation (Pawns)	(7) Stockfish Eval. Change (Pawns)
15 seconds	21,000	825	4,314	1812	0.46	0.41	-1.00
				(252)	(0.62)	(6.05)	(2.27)
30 seconds	25,665	936	3,559	1729	0.93	0.44	-0.97
				(266)	(0.86)	(5.95)	(2.23)
1 minute	157,925	5,140	35,586	1634	1.47	0.25	-0.69
				(309)	(1.40)	(5.93)	(2.13)
3 minutes	73,665	4,645	16,556	1661	3.80	0.18	-0.51
				(297)	(4.59)	(4.62)	(1.72)
5 minutes	143,274	7,225	28,990	1587	5.97	0.24	-0.59
				(261)	(7.72)	(5.68)	(1.76)
10 minutes	30,458	2,356	7,960	1696	10.54	-0.14	-0.51
				(243)	(14.25)	(5.74)	(1.64)

Notes: Reported values in columns (4) to (7) are means, with standard deviations in parentheses. Stockfish evaluation is censored at $+20$ and -20 pawns. Stockfish Evaluation is at the start of the turn, from the perspective of the moving player. The change is how much the evaluation changes as a result of the turn. Source: Lichess

Tables C.3 to C.6 show the first-stage estimates and the precise numbers behind Figure 1. The first two tables correspond to Panel (a) and the second two tables to Panel (b).

Table C.3: First-Stage, Figure 1a

	(1) 15 second	(2) 30 second	(3) 1 min	(4) 3 min	(5) 5 min	(6) 10 min
Time Control Indicator	0.295	0.431	2.099	1.727	3.509	2.395
	(0.00931)	(0.0106)	(0.0139)	(0.0323)	(0.0387)	(0.133)
Observations	34142	41010	224162	121828	229318	49936

Notes: First-stage of regression estimating the marginal benefit of attention. Each column represents the difference in time taken between the time spent in the time control indicated by the column title, and the next highest time control (i.e. 1 min for 30 seconds). Each regression includes a fixed effect for the matched variable, i.e. the player who is moving, the turn of the game, and the approximate computer evaluation. The outcome is measured in seconds. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Lichess.

Table C.4: IV Regression, Figure 1a

	(1) 15 second	(2) 30 second	(3) 1 min	(4) 3 min	(5) 5 min	(6) 10 min
Seconds Spent	0.0145	0.00539	0.00145	0.000324	0.000495	0.000406
	(0.00161)	(0.00114)	(0.000106)	(0.000206)	(0.0000753)	(0.000237)
Observations	34142	41010	224162	121828	229318	49936
F-statistic	1005.5	1658.2	22684.3	2850.6	8230.3	325.3
Marginal Cost	.0715	.0463	.0173	.00348	.00128	.000344

Notes: Instrument variables regression estimating the marginal benefit of attention. Each column represents the regression of the probability of winning based on the position and the clocks in the designated time control on the time spent on the move, using matched moves from the next highest time control and the instrumental variables strategy described in the text. Each regression includes a fixed effect for the matched variable, i.e. the player who is moving, the turn of the game, and the approximate computer evaluation. The coefficient can be interpreted as the marginal benefit in the probability of winning based on the resulting position per second spent on the move. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Lichess.

Table C.5: First-stage, Figure 1b

	(1) 30 second	(2) 1 min	(3) 3 min	(4) 5 min	(5) 10 min	(6) 15 min
Time Control Indicator	-0.302	-0.440	-2.103	-1.721	-3.533	-2.363
	(0.00788)	(0.0103)	(0.0132)	(0.0319)	(0.0403)	(0.134)
Observations	33506	43516	225082	122336	229570	49644

Notes: First-stage of regression estimating the marginal benefit of attention. Each column represents the difference in time taken between the time spent in the time control indicated by the column title, and the next lowest time control (i.e. 30 seconds for 1 minute). Each regression includes a fixed effect for the matched variable, i.e. the player who is moving, the turn of the game, and the approximate computer evaluation. The outcome is measured in seconds. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Lichess.

Table C.6: IV Regression, Figure 1b

	(1) 30 second	(2) 1 min	(3) 3 min	(4) 5 min	(5) 10 min	(6) 15 min
Seconds Spent	0.0266	0.0117	0.00277	0.000480	0.000738	0.000617
	(0.00240)	(0.00143)	(0.000139)	(0.000217)	(0.0000813)	(0.000244)
Observations	33506	43516	225082	122336	229570	49644
F-statistic	1463.9	1827.1	25469.9	2906.7	7695.8	310.4
Marginal Cost	.0432	.0165	.00285	.00105	.000239	.000145

Notes: Instrument variables regression estimating the marginal benefit of attention. Each column represents the regression of the probability of winning based on the position and the clocks in the designated time control on the time spent on the move, using matched moves from the next lowest time control and the instrumental variables strategy described in the text. Each regression includes a fixed effect for the matched variable, i.e. the player who is moving, the turn of the game, and the approximate computer evaluation. The coefficient can be interpreted as the marginal benefit in the probability of winning based on the resulting position per second spent on the move. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Source: Lichess.

Figure C.1 shows the same analysis as presented in Figure 2, except that instead of looking for heterogeneity amongst skill level, the heterogeneity is based on the number of seconds remaining on the player's clock at the start of their turn. Consistent with the previous result that most of the deviations happened in the faster time controls, it is also true that the deviations from rational inattention in Figure C.1 happen when there is less time on the clock. This can be seen in the bottom row (where moves are matched to a faster time control) in every time control. Theory would predict marginal benefit ought to be higher, but it is lower in the last parts of the game.

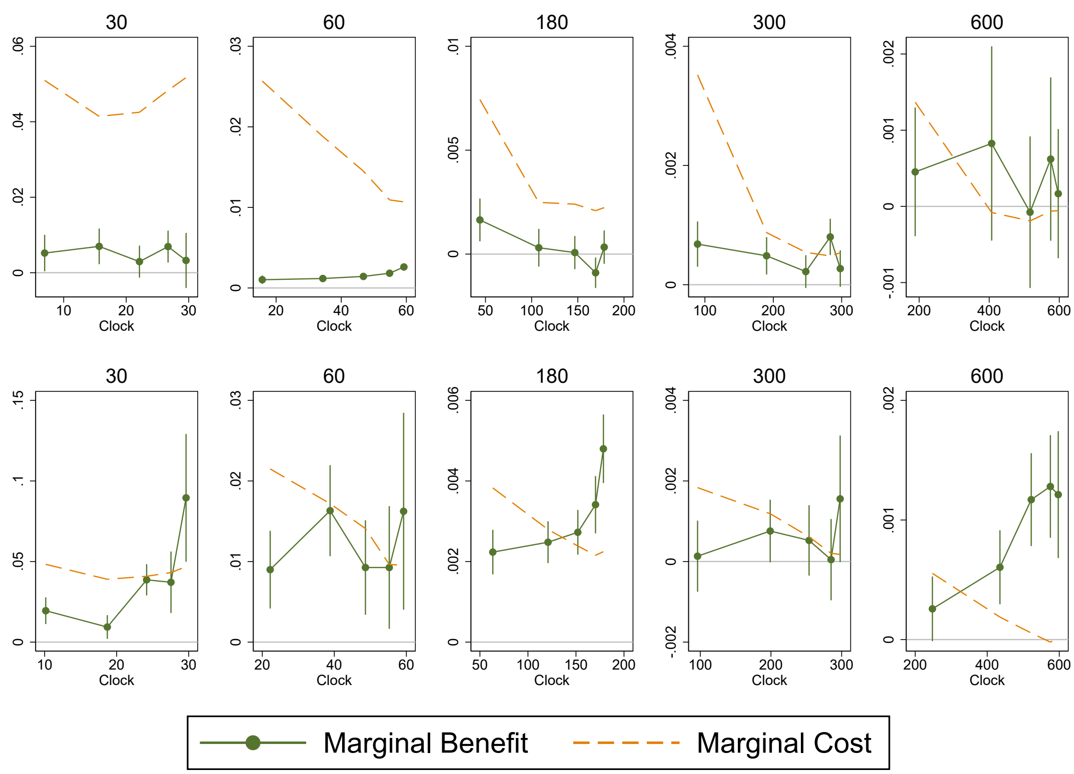
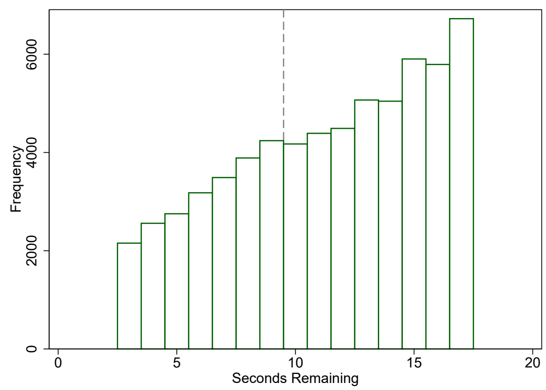
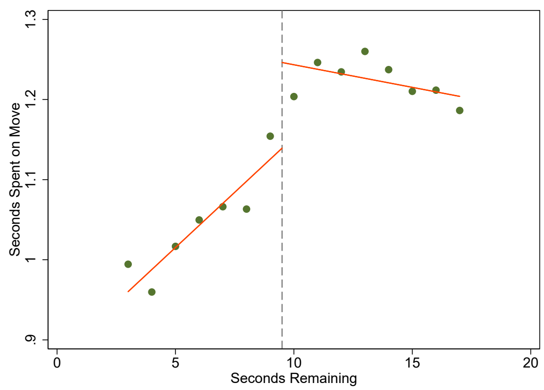


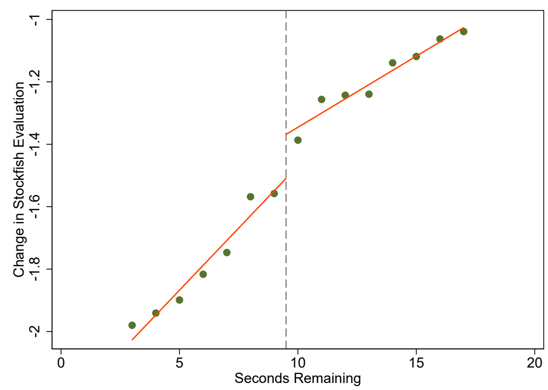
Figure C.1: **Heterogeneity in Marginal Benefit and Marginal Cost by Time Remaining at the Beginning of the Turn.** The rational inattention prediction is that the marginal benefit should fall below the marginal cost in the top panels and vice versa in the bottom panels. The number above the panel indicates the seconds in the baseline time control.
Source: Lichess, author's calculations.



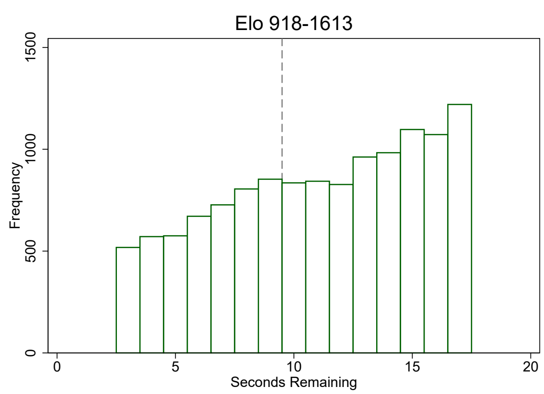
(a) The Distribution of Time Remaining After a Move



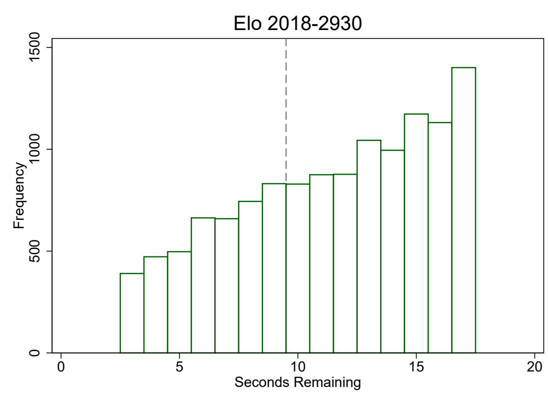
(b) The Amount of Time Spent



(c) The Change in Computer Evaluation



(d) The Distribution of Time Remaining After a Move, Low-Elo



(e) The Distribution of Time Remaining After a Move, High-Elo

Figure C.2: **The Role of the Low-Time Reminder, 30 second time control.** Panel (a) shows a histogram of the number of moves played by the number of seconds remaining. The number of seconds remaining is measured discretely. The beep occurs when there are 10 seconds remaining. Panels (b) and (c) show the average amount of time spend and the change in computer evaluation, by the number of seconds remaining at the end of the move. Panels (d) and (e) show the same as Panel (a), but only for the lowest and highest quintile of the mover's Elo.

Source: Lichess, author's calculations.

Figure C.3 shows all five quintiles of Elo for the distribution of time remaining after a move. There is significantly more bunching for the lower-rated players.

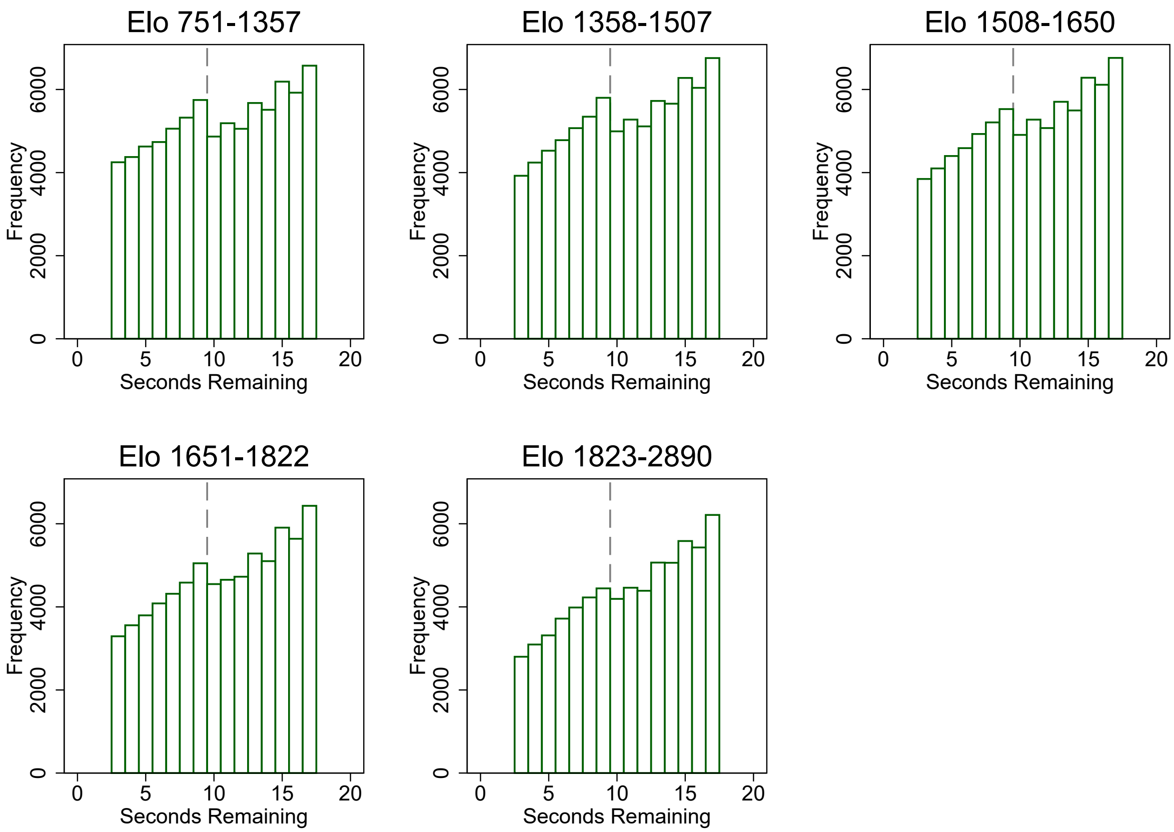
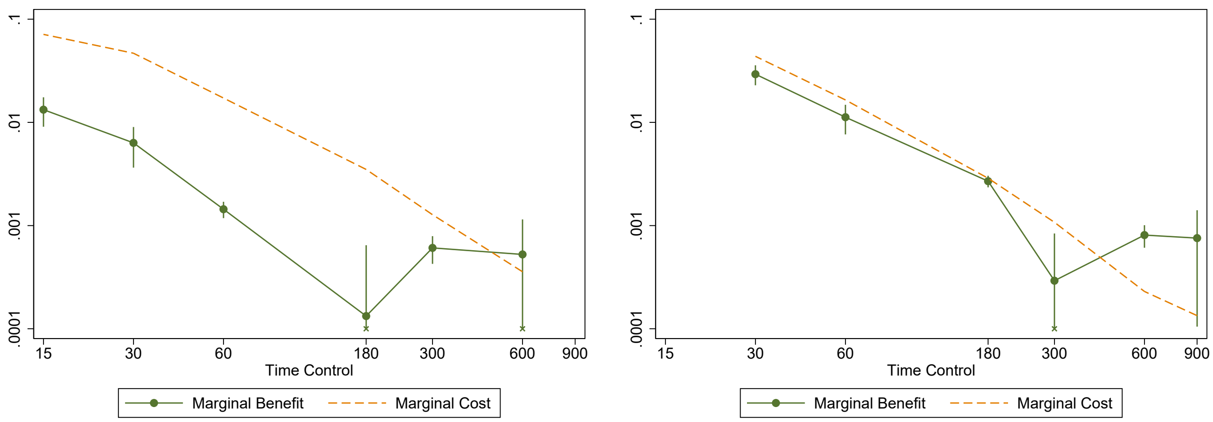


Figure C.3: **Distribution of the Time Remaining, Split over Elo-rating bins.** The figures shows a histogram of the number of moves played by the number of seconds remaining in 1 minute games. The number of seconds remaining is measured discretely. The beep occurs when there are 10 seconds remaining.

Source: Lichess, author's calculations.

Figures C.4 and C.5 show the robustness of Figures 1 and 2 to matching based on 100 bins of the initial Stockfish evaluation, instead of 50 bins.



(a) Lower Bound of Marginal Benefit

(b) Upper Bound of Marginal Benefit

Figure C.4: **The average marginal benefit and marginal cost of spending additional time on a move. Robustness to using 100 bins.** Intervals are 95 percent confidence intervals, censored at .0001. Censoring is denoted by an X at the bottom of the interval. Tables of these results, including precise estimates and standard errors can be found in Appendix Tables C.4 and C.6. Time control on the x-axis is measured in seconds, e.g. 60 corresponds to a 1-minute time control. The rational inattention prediction from Section 3 is that the marginal benefit should fall below the marginal cost in Panel (a) and should be above marginal cost in Panel (b).
Source: Lichess, author's calculations.

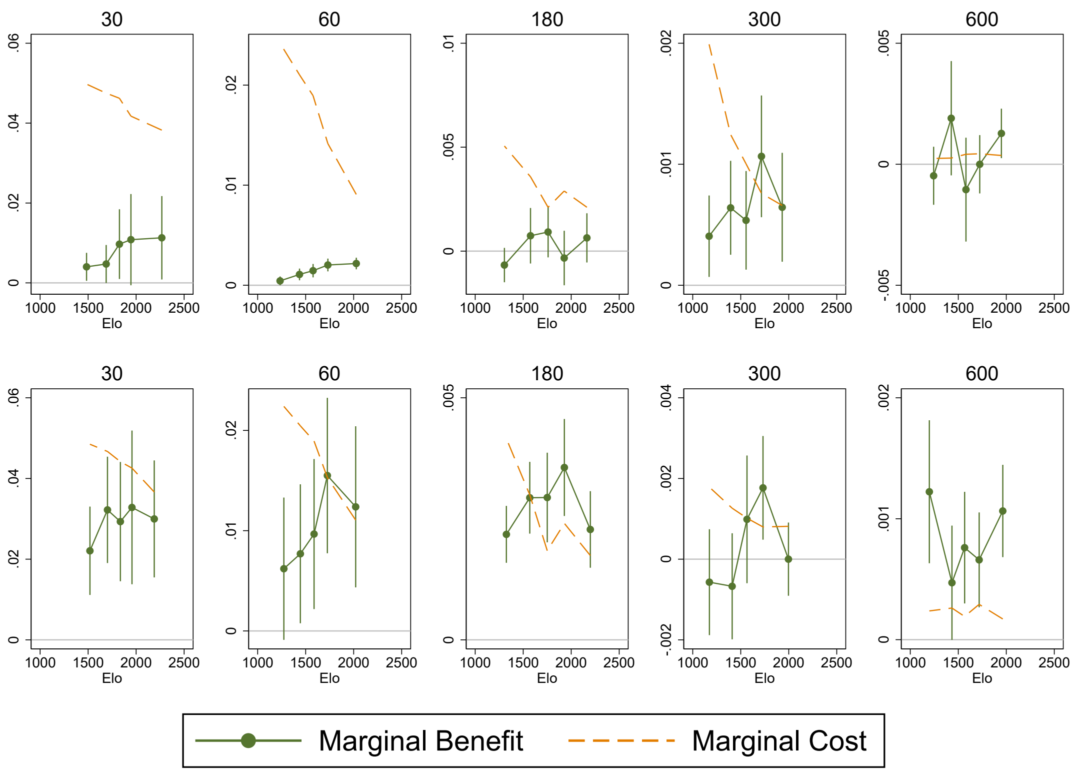


Figure C.5: **Heterogeneity in Marginal Benefit and Marginal Cost by Elo rating. Robustness to using 100 bins.** The rational inattention prediction is that the marginal benefit should fall below the marginal cost in the top panels and vice versa in the bottom panels. The number above the panel indicates the seconds in the baseline time control.

Source: Lichess, author's calculations.