A check for rational inattention

Howard, Greg

University of Illinois

2021

A Check for Rational Inattention

Greg Howard*

June 9, 2021

Abstract

Models of rational inattention allow agents to make mistakes in their actions while assuming they do not make mistakes in attention allocation. I test this assumption by comparing attention's marginal benefit (better actions) and marginal cost (less time for future decisions) using millions of online chess moves. I cannot reject that skilled players equalize marginal benefit and marginal cost across different time controls. Bad players, when they have little time, under-adjust their attention allocation, leading them to have higher marginal cost. 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 economic models. It allows for agents to regularly make mistakes, while maintaining the ability to calculate welfare and do counterfactual analysis. The key assumption for these models is that agents are able to allocate attention optimally: they can make mistakes in their actions, but they do not make mistakes allocating attention. Yet the empirical evidence for rational inattention has been unable to test whether attention is optimally allocated. Rather, the tests have been typically weaker, focusing on whether attention is beneficial and whether it responds to changes in incentives.

I study rational inattention in online chess, where players face a trade-off between spending time in order to make a better move and losing time that may be valuable later.[1] By measuring the marginal benefit and marginal cost, I can test whether agents are able to optimally allocate their attention. For good players and for players with sufficient time, I am unable to reject that the marginal benefit and marginal cost are equal. But for worse 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, bad players do not adjust their attention as much as a rationally-inattentive agent.[2]

Testing rational inattention is difficult for a variety of reasons. First, it is typically hard to observe direct measures of attention. Second, economists do not tend to have a way to evaluate whether an agent made the "right" choice, and even more rarely how bad the choice was compared to alternatives. Third, empirical tests often require an exogenous change in the cost of attention, which is often hard to find. Finally, if rational inattention is an acquired

[1]Online chess is quite fast compared to classical chess, which readers may be more familiar with due to the recent popularity of the Queen's Gambit on Netflix. This paper focuses on games in which players have between a total of 15 seconds and 15 minutes to make all of their moves.

[2]Many papers in the rational inattention literature show that people adjust their actions less than would a fully-rational agent. There is a parallel in this empirical test, in that these chess players adjust their attention less than would a rationally-inattentive agent.

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 directly observe how long a player spent choosing their move. Computer engines are significantly better than any human and are commonly used to evaluate a human's moves. The time controls—the amount of time each player has to make all of their moves throughout the game—vary, giving plausibly exogenous variation in attention. And good chess players have devoted years to getting better at the game.

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 am able to 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. Given that models of rational inattention typically assume not just that players change the way they allocate attention, but that they actually are doing so optimally, my paper contributes by testing this stronger condition.

To do this test, I have to estimate the probability of victory for each move. 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. While doing this estimation, I also calculate the marginal cost of spending additional time.

Then, matching moves made in similar situations across time controls, I calculate the marginal benefit of making a better move using the time control as an instrumental variable. In the data, the returns to attention are concave, so matching to a shorter time control gives an overestimate of the marginal benefit, and matching to a longer time control gives an underestimate. This provides two one-sided tests: that the marginal cost should be below the first measure of marginal benefit, and that the marginal cost should be above the second measure.

In several of the faster time controls, I am able to reject the hypothesis that the average marginal cost and average marginal benefit are equalized. The players in these time controls have a lower marginal benefit than the marginal cost. This suggests players are spending too long on making moves when they have little time.

Digging into these results, I can reject the hypothesis of equalized marginal benefit and marginal cost for bad players, as defined by the Elo ratings. Low Elo rated 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. This shows that rational inattention is a skill that the best players have. While the bad players are still reallocating attention, they are not optimizing as well as the best players.

I also investigate heterogeneity by the stage of the game. Consistent with finding that players are not playing optimally in short time controls, I find that this suboptimality is driven by moves where time is more scarce.

To provide complementary evidence that some players are not optimizing their allocation of attention, I show that a simple intervention by Lichess when the player is low on time—a small beep and changing the color of the clock—lowers the amount of time spent on moves and also raises the average size of the mistake. This result is consistent with players becoming aware of how high the cost of their time is, and adjusting their strategy to better optimize attention. Bad players react more strongly to this intervention, consistent with the previous result that they are the ones that are not allocating attention optimally.

Related Literature

Rational inattention was first used in economics by Sims (2003). Since then, it has become more common in many different settings, including empirical ones. 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). See

Section 4.4 of Maćkowiak et al. (2020) for a review.

Increasingly, economists have looked outside the laboratory to look 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), online finance behavior (Mondria et al., 2010; Sicherman et al., 2016), 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).

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?"[3] Sometimes these are combined into "are better decisions made when the stakes are high?"

The test I propose in this paper is considerably stronger than those tests. Unlike any of the other applications, I can empirically measure both the benefit and cost of attention. With that measurement, I can test not only whether people react to changes in the costs of attention or the stakes of the situation, but whether they actually equalize their marginal benefit and marginal cost.

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 suggest the example of a driver being close to rational in processing traffic signals because they do that regularly, but they hypothesize the same driver might 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, who likely have much more experience, are better at

[3]These are the tests I perform in Appendix A, to show that chess is a good setting to study rational inattention.

rationally allocating attention.

The results on skill also speak to the literature that takes place in the laboratory. For example, rejecting rational inattention for a decision in which test subjects are inexperienced or have very 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). In the sense that I am also estimating a value function in a competition to judge the performance of decision-makers, this paper also bears some similarity to Romer (2006) who uses a similar strategy in football.

1 Setting and Data

Chess is one of the most popular board games in the world. Two players, one with control of the sixteen white pieces and one with control of the sixteen black pieces, take turns moving a piece around an eight-by-eight grid that makes up a chess board. Of the sixteen pieces, there are six types of pieces, which are each governed by their own rules. The weakest is a pawn, the strongest is the queen, and the most important one is the king. To win a chess game, a player must checkmate her opponent's king, which is when one of the player's pieces could capture the opponent's king the next turn and the opponent is unable to make a move that prevents that. There is no randomness in chess, but the number of possible chess games is extensive, and unlike many other deterministic games, it has not been solved.

Lichess is one of the most popular websites to play chess online, and millions of games are played there each month, by novices and by the world's best chess players.[4] In most competitive games, each player is given a certain number of minutes to make all of their moves, called a time control. When it is the player's turn, their clock counts down, and if it

[4]Magnus Carlsen, the world champion, has played more than 6000 games on the website.

ever reaches zero, they lose the game.

Since 1997, when IBM's Deep Blue defeated Gary Kasparov (Campbell et al., 2002), the best computers have been able to beat the best human players. Today, the gap is large enough that games between humans and computers are not competitive. Nonetheless, professional chess players use computers extensively in their preparation for games against other humans.[5] More importantly, computers are an important tool in order for humans to evaluate their play after the game, identifying on which moves they (or their opponents) made mistakes. Any player on Lichess can review their own games with the help of a strong chess engine.

The most ubiquitous computer engine today is called Stockfish, and it is amongst the strongest in the world. Stockfish, like many chess engines, has two parts to it. One part has rules to evaluate a position: e.g. more and stronger pieces are better. The other part is a searching algorithm that looks at how the position will develop. Sometimes, the computer can see all the way to the end of the game, in which case they might evaluate the position as "checkmate for black in 4 moves." More commonly, Stockfish evaluates the position using a measure called pawn equivalents, e.g. "white is winning by 0.92 pawns."[6]

Lichess makes the information detailing games played on their site available for download (Duplessis, 2017). This includes all the moves made in the game, the time remaining when the move was made, the players, the time control, and the result. Since April 2017, it has also included the time on each players' clock at the end of each move. And for about six percent of games, they also include Stockfish's evaluation of each position reached in the game.[7]

[5] In part, they memorize openings that computers suggest. They are also able to experiment with new openings, and a computer is able to tell them quickly whether their opening might lead to any weaknesses that their opponent might exploit.

[6] 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 and knights are worth 3, rooks are 5, and queens are 9. Computer engines have adopted this language in evaluating positions.

[7] To have a Stockfish evaluation included in the dataset, someone must review the game on the website using the engine. The most common reason this would occur is that one of the players wants to see where their mistakes are, and Lichess makes this easy to do. Sometimes, people also look at games in which they were not involved. Occasionally, Lichess will also look at the Stockfish evaluation for cheat detection

I use data from this first month, April 2017. This includes about 68 million moves for which there is data on the computer evaluation. Throughout the paper, I focus on moves in seven of the most popular time controls. These are 15 second, 30 second, 1 minute, 3 minute, 5 minute, 10 minute, and 15 minute chess.[8] Together these make up a majority of the moves played on Lichess. I also only include games that are rated and that are not part of a tournament.[9] Summary statistics are presented in Table 1.

One reason that chess is a good place to study rational inattention is that the players understand the importance of allocating time. As former world champion Vladimir Kramnik said, "Time is precious when you don't have enough of it" (ChessBase News, 2003). 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 the ordinary least squares version of the same regression is biased because players take longer on harder moves. This exercise further establishes chess as a good setting to study rational inattention.

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, the models of rational inattention used by economists make a stronger assumption: that agents

purposes.

[8]Some time controls also include an increment in which the player gets an extra few seconds per move. I do not include those time controls in my dataset.

[9]In a rated game, the player's Elo rating will change based on whether they win or lose, so it is less likely to have games in which they are not trying to win. In a tournament, there may be times where the relative value of a win and a draw is quite different than normal. Further, Lichess runs tournaments in which players can play with half of their initial time in order to score more tournament points. It is for these reasons that I drop tournament games.

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 Evalution is at the start of the turn, from the perspective of the moving player. The change is how much the evalution changes as a result of the turn.
Source: Lichess

are optimizing their allocation of attention. Often, the first order conditions of such an optimization are used to establish how economic agents behave. In this section, I propose a test based on such a first-order condition to see whether players are optimizing their attention.

To model chess, I consider four state variables: whose turn it is, the position of the board, and the clocks of the two players.[10] Denote the value function as V. On white's turn,

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

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 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.[11]

Implicit in this notation is the assumption that the prior information, Q^w or Q^b 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.[12] I discuss whether this is likely to bias my estimates in Section 4.4.

Black faces a similar problem, with the notable difference that they are attempting to minimize V:

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

[10]In principle, the position of the board also includes the previous positions of the board, because it affects such things as whether pawns are able to capture en pessant, whether one of the players has given up the right to castle, and whether there is a three-fold repetition, which would lead to a draw. In the empirical section, where I assume the board position is captured by the computer evaluation, the computer does take these things into account.

[11]For example, in many economic models, $\mathbb{I}$ might be proportional to the reduction in entropy between the prior information Q and the posterior once the agent sees S. This is the classic case of Shannon mutual information (Shannon, 1948). In this paper, I do not impose that $\mathbb{I}$ is Shannon.

[12]This assumption does not mean that players are not forward-looking, as evaluating possible positions is inherently a forward-looking endeavor. However, this is a stronger assumption than that agents do not acquire information that will be useful for future moves beyond its use for current moves (see Afrouzi and Yang (2021) for a discussion of how this relies on the linearity of information). My assumption implies that condition, but in addition, it also 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.

such that

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

There exist some terminal states s such that the value function at those states is 0, 1, or in relatively rare cases, $\frac{1}{2}$, i.e. white wins, black wins, or there is a draw.[13] Furthermore, if t^w reaches 0, the value function is 0, and if t^b reaches 0, the value function is 1.[14]

Since white's and black's problems are symmetric, I consider only white's problem for the rest of this section. When I take it to data, I flip the problem around and test the first-order conditions of the agents combined.

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 we can 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. Assume $\vec{\pi}(\tau)$ is continuous and differentiable. In that case, the problem the white player tries to solve is:

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

The first-order condition of this problem 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

[13]Draws are rare in online chess, unlike in classical chess, where they are the most common outcome. In my dataset, they occur 2.9 percent of the time.

[14]In some rare cases, the position s may be such that the value function is $\frac{1}{2}$ when the clock hits zero, if the opponent does not have enough pieces remaining to checkmate the king.

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 the maximization, and it is the focus of my empirical tests. A similar equation holds for black.

3 Empirical Strategy

3.1 Estimating the Value Function

Both sides of equation (1) require estimating the value function. Note that in our setup, V is equal to the expected score for white, when valued at 1 for a win, $\frac{1}{2}$ for a draw, and 0 for a loss. The dataset is quite large, so I estimate the value function by local regresssion over similar game situations.

The main challenge is measuring the position. There are estimated to be more than 5×10^{42} possible chess positions (Shannon, 1950), which is significantly larger than the number of games ever played.[15] So a key challenge is reducing the dimensionality of the game. I propose using the Stockfish evaluation as a stand-in for s, making the value function much easier to approximate. This is a good stand-in because it reduces the dimensionality of the game greatly, while also being highly correlated to the outcome. Further, "better" moves are judged by the computer evaluation throughout the chess community.

If s becomes a single-dimensional argument, then I can approximate $V(b, s, t^w, t^b)$ using a local linear regression. This has the advantage of simultaneously estimating the first derivative with respect to t^w. In practice, I split the evaluation into 40 bins, and each clock

[15] Ten billion people (an overestimate of the population) playing a million games a year (a game every 30 seconds) for a 10 thousand years (chess has arguably been around for about 5000 years) would be 10^{20}.

into 12 bins per time control, for a total of 5760 bins per time control.[16] 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 if the player ends up winning the game, $\frac{1}{2}$ if they draw, and 0 if they lose. I then use the predicted value as V within the bin of interest, and β_2 as the marginal cost of time for observations in that bin ($\frac{\partial V}{\partial t^w}$).

I run this separately for each time control and for five bins of player quality. In total this requires running 201,600 regressions. I drop the estimate if the bin of interest has fewer than 10 total observations.[17]

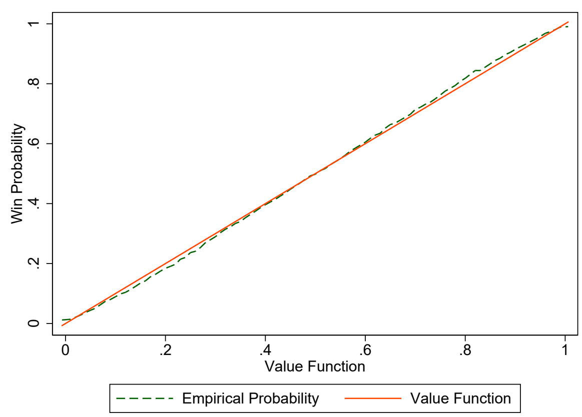
3.2 Checking the Fit of the Value Function Estimation

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.

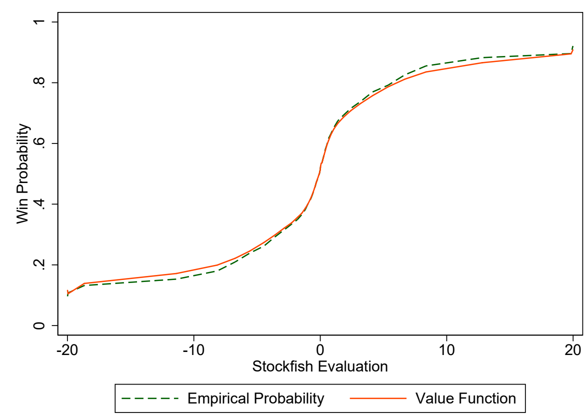
In Figure 1, I plot the average value of the game result alongside the average value of our estimated evaluation for 50 bins of the value function, 50 bins of the Stockfish evaluation, 50 bins of the player's clock, and 50 bins of the opponent's clock. They seem to match fairly

[16] 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 (roughly the queen, two rooks and an additional pawn advantage for one side) because there are a few extreme outliers that otherwise make local regression a bad approximation. Above 20 pawns, there does not seem to be much of an effect of additional advantage 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 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.

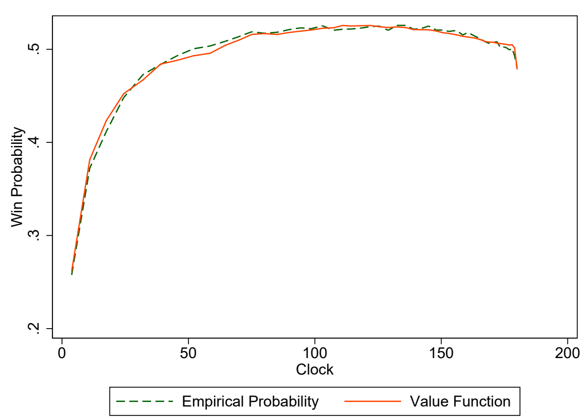
[17] For example, it is rare to have one player with very little time and one with lots of time.



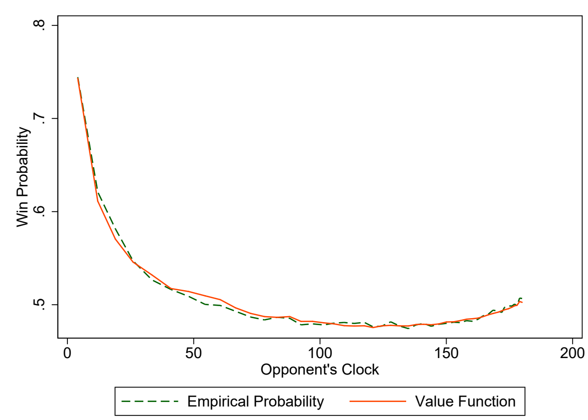
(a) Value Function



(b) Stockfish Evaluation

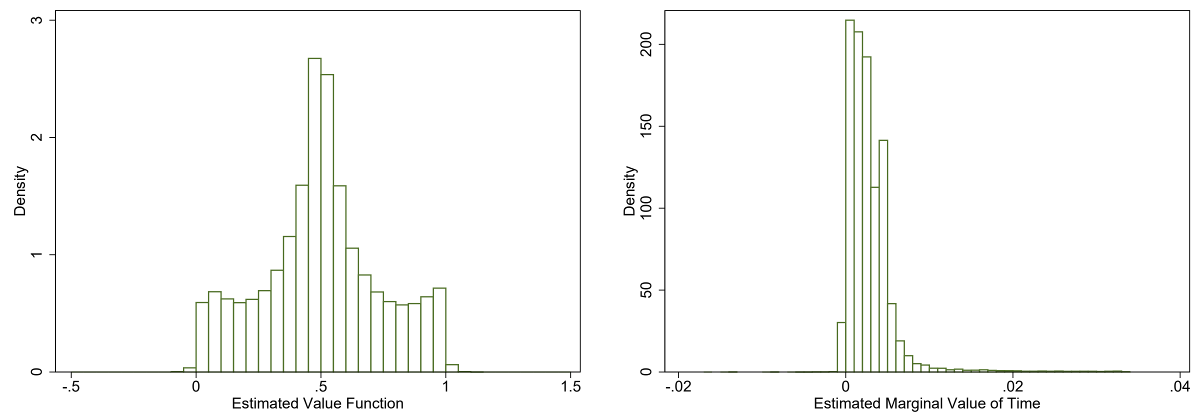


(c) Time



(d) Opponent's Time

Figure 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



(a) Value Function

(b) Marginal Value of Time

Figure 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.

well, suggesting that the local regression strategy is able to adequately capture the curvature 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. In Figure 2, very few of the observations are estimated to have a negative marginal value of time.

In Figure 2, I show the distribution of the estimated value function, and the estimated marginal value of time. My expectation for the value function was that it would range from 0 to 1 and have a peak near $\frac{1}{2}$ since all games start near there. Indeed, that appears to be the case. And although nothing constrains it to be so, there are very few values outside of the [0,1] range. My expectation for the marginal value of time would be that it is almost everywhere positive, that it would peak at a number close to 0 and that it would have a long right-tail. That appears to be the case. 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 later will 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.

In Table 2, 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.[18]

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,[19] 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

[18] 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 very 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.

[19] 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.

Table 2: 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.

same bins of the player's clock, the opponent's clock, the Stockfish evaluation, and the Elo rating of the player.[20] 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) and (6), I am estimating over 20,000 additional coefficients, but the R^2 goes up only about three-tenths of one percent.[21] 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

[20]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.

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

the value function, I am confident that the computer evaluation and the clocks are capturing a large majority of the variation.

3.3 Variation in the Time Spent on a Move

While I can estimate the left-hand side of equation (1) from ordinary least squares, the right-hand side inherently involves the thought experiment of how much better a move might be if the player had more time.

To get at the causal effect of time on the quality of move, I use a matching strategy combined with instrumental variables. The idea behind the strategy is to compare how the move quality in a similar situation but a different time control changes because of the extra time to spend on the move.

For each move in a time control, I try 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 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.[22] The matching is done without replacement. For the seven time controls I consider, I match each with the next higher time control (e.g. 3 minutes is matched to 5 minutes), and repeat the analysis matching it with the next lower time control (e.g. 3 minutes is matched 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 am able to match several hundred thousand moves, although at 15 second-30 second match and the 10 minute-15 minute match, I have only tens of thousands of matches.[23]

To measure how much the quality of the move improves under a different amount of

[22] 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). For each match, the balance in terms of the starting evaluation is quite strong. The difference in starting evaluation across matched observations averages less than two hundredths of a pawn for all six pairs of time controls, and it is not statistically significant in any of them. The largest t statistic in absolute value is 0.52.

[23] Note that to estimate the value function, I do not need to match moves, so I am using millions of data points per time control in that case.

time, I evaluate the value function using the time remaining from the primary move, but the engine evaluation from the matched move. Hence, 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:

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

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.

3.4 Aggregation

Across many different positions, times, and skill-levels, the marginal benefit is not constant. β can be expressed as a weighted average of individual 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 effect of these weights in order to recover the average treatment effect is impossible, so I cannot estimate the unweighted average marginal benefit. However, 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. In practice, the reweighting does not have much effect on the estimates of average marginal cost, but in principle it could matter, and so I do the reweighting throughout the paper.

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.

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. Under the assumption of concave returns, matching moves to a lower time control will provide an upper bound for the marginal benefit. And matching moves to a higher time control will provide a lower bound for the marginal benefit.

This means that 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.

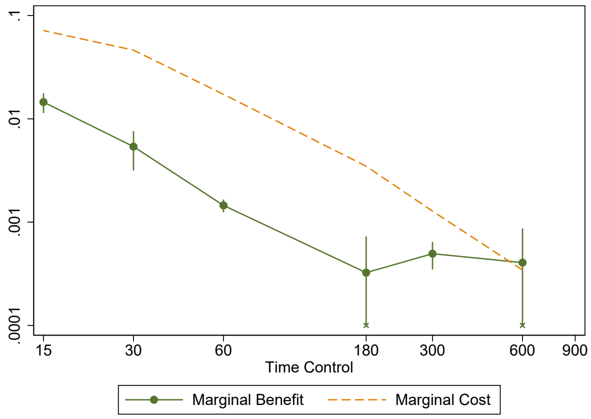
4 Results

4.1 Comparing Marginal Benefit and Marginal Cost

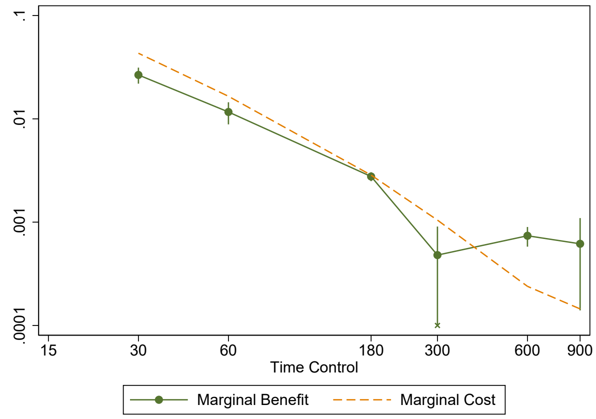
Figure 3 plots the results of $\hat{\beta}_{MB}$ and $\hat{\beta}_{MC}$ for each time control. Note that both panels are presented in a log-log scaling.[24] In panel (a), moves are paired with moves from a higher time control. In panel (b), moves are paired with moves from a lower time control. Hence, theory implies the marginal benefit estimate should be below the marginal cost estimate in panel (a) and the marginal benefit estimate to be above the marginal cost estimate in panel (b). Panel (a) clearly conforms with the null hypothesis predicted by rational inattention. For one point estimate, the order is reversed, but the marginal cost line is well within the

[24] Unfortunately, this requires censoring the 95 percent confidence intervals for a couple of the points. However, none of the affected points are important to the analysis. In all cases, the censoring is not important to evaluate whether marginal cost and marginal benefit are equalized.

range of the marginal benefit's confidence interval.[25]



(a) Lower Bound of Marginal Benefit



(b) Upper Bound of Marginal Benefit

Figure 3: **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 B.2 and B.4. 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.

In Panel (b), there are three observations for which the marginal cost is higher than the marginal benefit, in contrast to theory. This shows that the marginal benefit for the quality of move is below the marginal cost of time for future moves. Hence, players are spending too much time to decide on a move. The other three observations do not exhibit such behavior.

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 and the next, I split the data into categories to examine whether there are deviations from rational inattention.

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

In this section, I divide players into five groups per time control based on their Elo rating (which reflects the results of previous games). Players with a higher Elo rating are better at chess. Players can have different Elo ratings for different time controls, so I base it on the Elo rating in the baseline control, not the matched one. I repeat the analysis from Figure 3 for each of the five groups within each of the time controls.

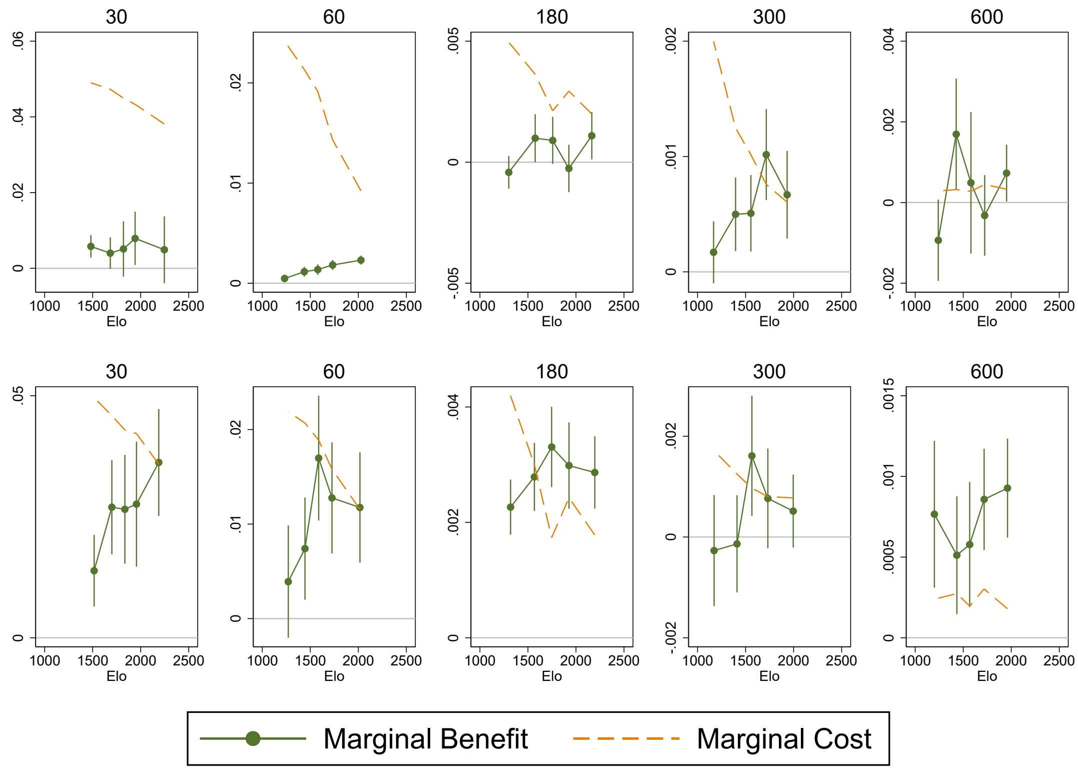


Figure 4: **Heterogeneity in Marginal Benefit and Marginal Cost by Elo rating.** The rational inattention prediction is that the marginal benefit should fall below the marginal cost in the top panels and vice versa in the bottom panels. The number above the panel indicates the seconds in the baseline time control.
Source: Lichess, author's calculations.

The results are presented in Figure 4.[26] Here, because each time control is presented in its own panel with its own y-axis, there is no log-scale. As the time controls increase, the y-axis has a smaller range. In the panels on the upper row, I match moves to the time control below the time control of interest, providing an upper bound on the marginal benefit

[26]To get a sense of what the magnitudes on the x-axis mean, the middle 98 percent of Elo ratings range from about 1000 to 2500. The world's best players have Elo ratings around 3000 on Lichess.

of time. In the panels on the lower row, I match moves to the time control above the time control of interest, providing a lower bound on the marginal value of time. That means that theory would predict that the orange marginal cost line should be above the green marginal benefit line in the top row, and it should be below the green marginal benefit line in the bottom row.

In general, it appears that the top row conforms with theory. However, there are several notable exceptions to the predictions of 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. This pattern would suggest these players are spending too much time on their moves, because marginal benefit is decreasing in attention.

The fact that marginal benefit is far below marginal cost for low-Elo players means that rationally allocating attention is a skill. Bad players are not able to equalize marginal benefit and marginal cost. Good players are at least close enough so as not to be detectable using the empirical strategy proposed in this paper. This is not to say that bad 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 4). It is just not increasing as fast as the marginal cost is.

The difference between low-Elo and high-Elo players comes from both the marginal cost and the marginal benefit of time. Low-Elo players have a lower marginal benefit and a higher average marginal cost. Consistent with both sides, bad players spend longer per move, especially at lower time controls. In 15 second chess, the lowest quintile spends 0.55 seconds per move compared to 0.38 for the highest quintile, a 44 percent difference. At five minute chess, the corresponding values are 11.2 seconds and 10.0 seconds, a difference of only 12 percent.

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.[27] However, they also play more. Looking at only the 3-minute time control, a player in the top quintile plays an average of 135 more moves over the course of a month compared to a player in the bottom quintile, who plays, on average, 128 moves. Similarly, a player in the top quintile is 11 percent more likely to have played in the previous month compared to a player in the bottom quintile, where the baseline is 73 percent. Unfortunately, the data does not allow me to disentangle which features of good players make them better at allocating attention.[28]

4.3 Heterogeneity by Time Remaining

Figures 3 and 4 suggest that any deviations from equation (1) are most prominent in short time controls. In this section, I investigate the role of available time by looking at the heterogeneity of results at different stages in the game. To do that, within each time control, I consider 5 equally sized bins of how much time is remaining in the game. As before, I then run the same analysis within each of these bins.

The results are presented in Figure 5. Like in Figure 4, in the top row, the orange line should be above the green line, and in the bottom row, the orange line should be below the green line.

In general, the top row again seems to conform to the theory. The bottom row rejects the hypothesis of rational inattention near the end of games. For every time control, the marginal cost is below the marginal benefit at the start of the game, but when time is running low, the marginal benefit is consistently below the marginal cost. Because in the

[27] The fact that bad players are optimizing attention suboptimally does not appear to be the reason that 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.

[28] Looking at cuts of the data based on the number of moves played, or whether the player was active in the previous month does not give strong enough results to compare to the effects of Elo ranking, even if one was tempted to interpret such a result causally.

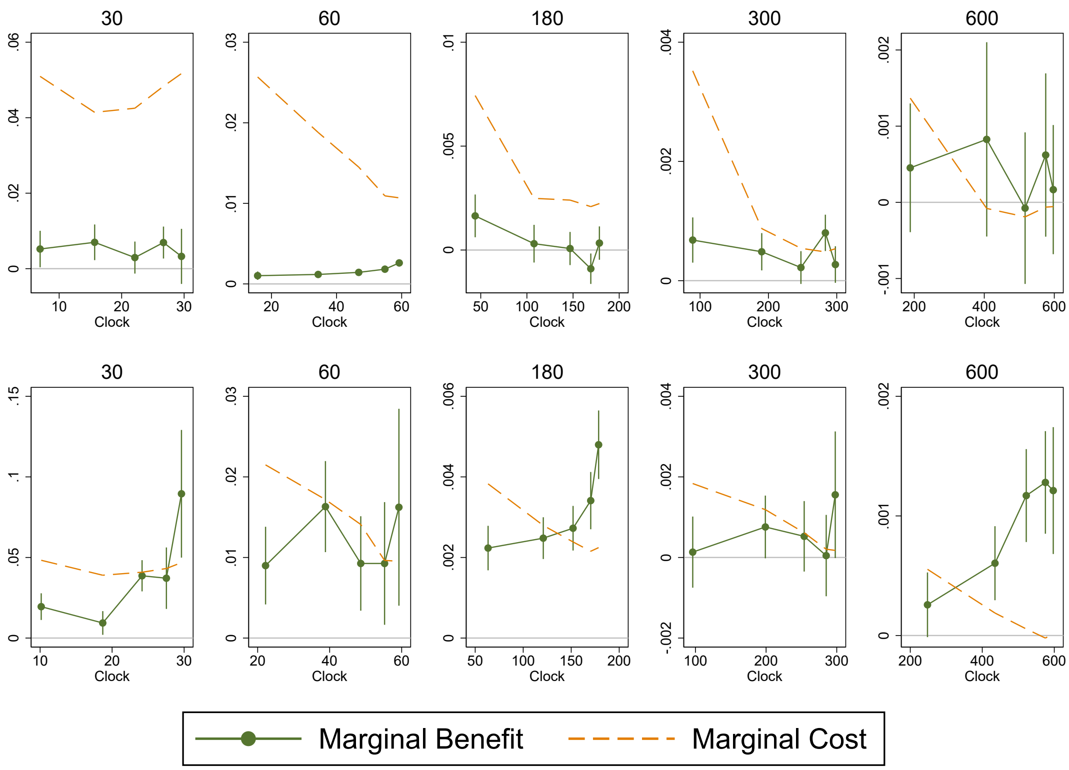


Figure 5: **Heterogeneity in Marginal Benefit and Marginal Cost by Time Remaining at the Beginning of the Turn.** The rational inattention prediction is that the marginal benefit should fall below the marginal cost in the top panels and vice versa in the bottom panels. The number above the panel indicates the seconds in the baseline time control.
Source: Lichess, author's calculations.

bottom row, the marginal benefit is an overestimate, this rejects the hypothesis of rational inattention when the clock is low.

4.4 Discussion of Potential Bias

In this section, I consider whether players playing optimally would necessarily satisfy the tests I propose. In general, I view these considerations as unlikely to pose a threat to the conclusions I drew in the last section.

Dimensions of the Position Not Captured by the Stockfish Evaluation

The biggest threat to the validity of my analysis is that the Stockfish evaluation is not a sufficient summary of the position on the board to effectively calculate win probabilities. For example, there may be positions on the board that a player would pursue because they will help them win, but which does not get a higher score from Stockfish. For example, in very low time situations, it may be optimal to check the opponent's king because if the opponent were planning to play a move that is illegal when their king is in check, the time it takes the opponent to reevaluate gives the mover an advantage. A computer evaluating the position would not consider this, and so it does not factor into my measure of the marginal benefit.

However, the analysis carried out in Table 2 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 check 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, about three-tenth of one percent. 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.

Of course, there is a possibility that there is some other dimension of the position—not considered in Table 2—that does explain a significant amount of win probability. However, putting the opponent in check and capturing a piece seem particularly likely to matter, so other dimensions should matter even less.

Corner Solutions

One possible objection to this analysis 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,[29] then it could be that the marginal benefit is lower than the marginal cost. In that case, it is optimal to take the minimum amount of time, and marginal benefit and marginal cost would not be equalized.

However, in estimating the marginal benefit by matching to a faster time control—which is where I find that marginal benefit is lower than marginal cost—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 up against the minimum time constraint, that move would get, on average, zero weight in the regression. Hence, the average is based only on moves in which the player is not constrained, and this should not be a concern.[30]

Selection from Draw Offers and Resignations

Often when players are confident they are going to lose, they will resign rather than keep playing, and this is sometimes considered good chess etiquette. Alternatively, games sometimes end because both players agree to a draw rather than continuing the game. One might worry that this could potentially bias the results because the sample is selected.

While this selection does affect the measure of the average marginal benefit and the average marginal cost, rational inattention predicts that the marginal benefit is equal to the marginal cost move-by-move, so taking an average over any subset, even a selected one, should still serve as a test of rational inattention.[31]

[29]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.

[30]A similar concern is that sometimes players take all of their remaining time and lose because of it. Because no move is played, it would not be in the dataset. Importantly, the existence of players losing on time does not tell us anything about whether players are rationally inattentive. Such an outcome could be the result of rationally-inattentive play. For example, a player might take a long time on a move because they mistakenly think the game is almost over—and hence the marginal value of time is low—but then end up playing a move that extends the game and causes them to later lose on time.

[31]A bigger concern regards whether Figure 3 supports the idea of concave returns to attention. If players are less likely to resign in 15 second chess, that could be part of the explanation of why I observe differences in the marginal benefit and marginal cost of 15 versus 30 second chess. However, the differences are so large that I do not think this can explain the downward relationship.

Having Already Thought Through the Move

Another objection is that the player could have already spent time thinking through the move, either during a previous turn, or during the opposing player's turn. One possible bias is that in a longer time control, there has been more time to think about the move, so even if only a little bit more time is used, more unmeasured time is spent considering the move. In most cases, this is likely to be small, as the player must think about many possible moves ahead of time, and so any extra time would be split amongst many possibilities. Moreover, this would mean that I am overestimating the marginal benefit of thinking, driving us to think that more players and more time controls are exhibiting sub-optimal behavior.

Thinking Ahead

A related story is that the players are thinking several moves into the future, and since this store of knowledge is not included in the state-space, that 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 to think this 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, as you can always think about future moves when they are in front of you.[32] If there is even a tiny chance that the future position being considered will not occur, then it does not make sense to spend time on it now unless it is helpful to determine the current move.

Nonetheless, even if players only receive information about the current move, that information could be valuable for future moves. For example, a signal that a player should play a certain move might indicate that a subsequent move is more likely to be good. However, at these time controls, players are unlikely to plan moves for far ahead that are that different than their initial instincts in that position. For example, playing a move that forks the queen

[32]By linearity, I simply mean that the cost of a signal about a future move is the same now or later. That seems reasonable as the cost here is measured in time.

and the king is not an issue because planning to take the queen is not all that beneficial since it would be obvious anyway.

More importantly, the heterogeneity that I found suggests that this is not the operative story. If it were the case that planning ahead was an important state-variable to the extent that it could explain my results, I would expect it to occur more for good players who can think more moves ahead and when players have more time so that they can plan elaborate strategies.[33] But I find that it is the bad players and the players with little time that exhibit the deviations from rational inattention.

Learning

A final concern might be that agents are sacrificing time in the present game to become better chess players and win future games. Perhaps by over-analyzing a move, they can make the move faster the next time they encounter a similar situation. This would bias the estimated marginal benefit down from the true marginal benefit.

However, one reason to think this is not the case is that players have unlimited time after the game to analyze moves and become better chess players. Not only that, they can use the help of a computer engine when not actively playing in order to learn better. Lichess makes this particularly easy. In fact, because someone has to use a computer engine to analyze the game for it to show up in the dataset, the players in my data are more likely than average to use this option.

This argument extends to other costs and benefits of attention that I do not include in the model. For example, economists often assume that attention is costly because 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: players are voluntarily playing, likely for fun. However, it is hard to square with why players are choosing to devote too much attention during the game rather

[33] I would be especially concerned regarding this bias for classical chess played by grandmasters.

than afterward.

5 Improving the Allocation of Attention

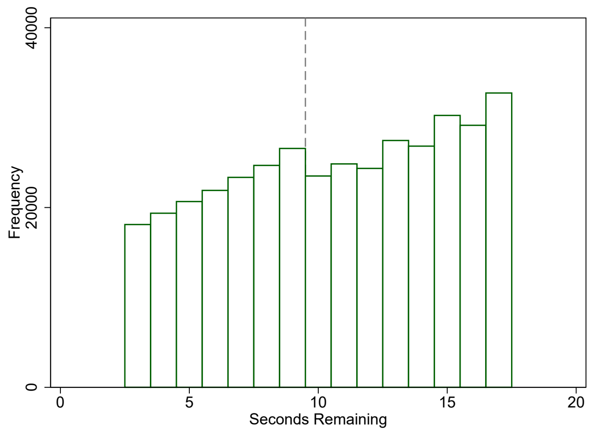
Lichess reminds players that their time is running low by playing a beeping noise and changing the color of the clock that the player sees on their screen from black to red. This reminder occurs at a set time: in 1 minute chess, it is when the player's clock is at 10 seconds.

Here in this section, I show that players take shorter amounts of time and make larger mistakes after this intervention. Because players were previously taking too long to make moves when they had little time, this analysis suggests that the intervention is effective at improving 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.[34]

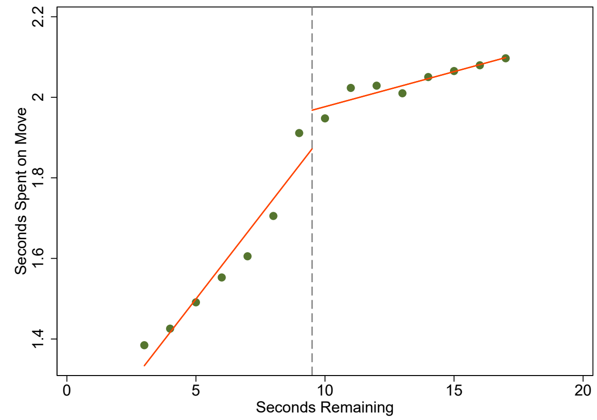
Unfortunately, the nature of this intervention does not allow for a regression discontinuity design. That is because the beep occurs when the clock passes a certain time threshold, and the time at which players make their move is endogenous. As expected, I observe significant bunching in the number of moves made at times just past the threshold rather than just before it (Figure 6a). In this figure, I look at the time remaining in 1 minute games. I only consider moves where the computer evaluation before the player's turn is less than a 5 pawns advantage for either player. In a 1 minute game, the beep occurs at 10 seconds, so I compare outcomes below and above that threshold.[35] Consistent with previous results about marginal cost exceeding marginal benefit for lower-Elo-rated players, I show in Appendix Figure B.2 that the bunching is much stronger for lower rated players. The difference in the number of moves made with 9 seconds compared to 10 seconds is 18 percent for the worst quintile of

[34] Antioch (2020) performs a similar analysis to this section, showing players blunder more when the beep occurs. Antioch argues that this makes the beep a bad thing, but my previous results argue that getting players to play faster and make more mistakes 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 at the time spent on the move, as I do in this section.

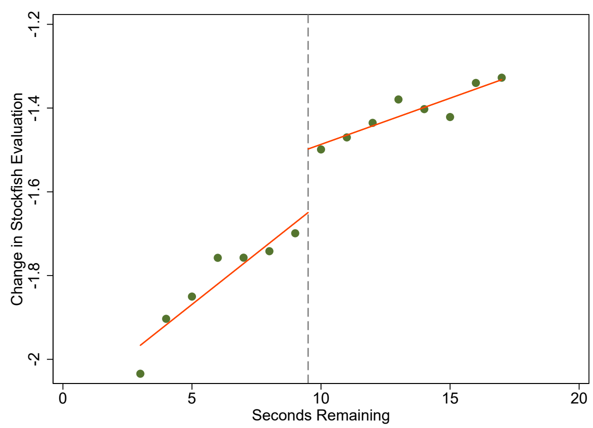
[35] I also show results for 30 second time control in Appendix Figure B.1. At longer time controls, I still observe bunching at times just below the time controls, but the discontinuity in evaluation is not statistically significant.



(a) The Distribution of Time Remaining After a Move



(b) The Amount of Time Spent



(c) The Change in Computer Evaluation

Figure 6: **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.
Source: Lichess, author's calculations.

players, but only 6 percent for the best quintile.

Even though I cannot do a regression discontinuity analysis, the data can still tell us about the players' behavior in response to the beep. The bunching itself reveals that players are choosing to not take much time after hearing the beep, presumably because the beep reminds them that the marginal cost of their time is high. Figure 6b 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 very large drop-off between 9 and 8 seconds, suggestive that players do move much faster after having heard the beep.

Figure 6c shows the quality of the move, as judged by the change in the computer evaluation, by the time remaining when the move is made. The move quality also falls discontinuously after the beep.[36] This means that players make larger mistakes on average after they hear the beep. Presumably, this is due to the fact that players are spending less time on their moves than they would absent the beep. Unfortunately, there is not an regression discontinuity here, or it would be nice to compare the size of the marginal benefit to my previous estimates.

The intervention is presumably beneficial to players 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.[37]

One additional piece of evidence that players benefit from this intervention is that it is very easy to turn off the sound on Lichess. Nonetheless, the beep still clearly has effects on the way people play. So even though it is causing more mistakes, players recognize the value it has of getting them to play faster.

[36] 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.

I cannot use the measured win probability because it is a function of the time remaining, so I simply use the Stockfish evaluation change.

[37] Of course, it could be that they speed up so much that it begins to hurt their win probability. I cannot directly check this.

6 Conclusion

The results in this paper have implications for both models of rational inattention and for further empirical work on rational inattention.

First, I have shown that the predictions of rational inattention seem to hold for good chess players and for chess players with sufficient time. The first-order condition associated with equalizing the marginal benefit and marginal cost of attention cannot be rejected in the data, despite the large dataset. This is generally good news for the way economists model rational inattention in the literature. However, the paper also shows that bad players do not optimally allocate attention, i.e. rational inattention is a skill. So models of rational inattention might make more sense for experts and experienced agents in whatever decision they are making. They might need to be modified for amateurs.

Second, this work also has implications for how to test rational inattention. Being able to reject rational inattention for bad players or players with little time has lessons for running lab experiments, as many tasks done in labs are not things the subjects are experts at. It also shows the importance of not expecting to see rational inattention in experiments where subjects are given very little time to make decisions.

Finally, it shows that simple interventions can help people allocate attention in a better way. The beep that Lichess gives when there are ten seconds remaining in a 1-minute game improves the allocation of attention, especially for the bad players who need it most.

References

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

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

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.

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.

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

Campbell, Murray, A Joseph Hoane Jr, and Feng hsiung Hsu, "Deep blue," *Artificial intelligence*, 2002, *134* (1-2), 57–83.

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.

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.

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.

A The Benefit and Allocation of Attention

To those experienced with chess, it is unsurprising that additional attention would improve the quality of play or that players do strategically allocate their allotted time. Nonetheless, in this appendix, I present econometric evidence to demonstrate these two facts.

Specifically, I want to consider the relationship between the time spent on a move and the quality of that move. As a naive measure of the quality of the move, I will 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 will change only slightly. Because of this, the vast majority of the 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 include player fixed effects interacted with turn fixed effects. 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.[38] These results would suggest that the causal effect of spending an extra second on your move is worth about 0.03 pawns in terms of the engine's evaluation of the position.

Nonetheless, there is a negative correlation between time spent and the engine evaluation:

[38] 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

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 such 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. In the rest of the paper, I dig deeper into whether that allocation of attention is optimal.

B Supplemental Figures and Tables

Tables B.1 to B.4 show the first-stage estimates and the precise numbers behind Figure 3. The first two tables correspond to Panel (a) and the second two tables to Panel (b).

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

	(1)	(2)	(3)	(4)	(5)	(6)
	15 second	30 second	1 min	3 min	5 min	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 B.2: IV Regression, Figure 3a

	(1)	(2)	(3)	(4)	(5)	(6)
	15 second	30 second	1 min	3 min	5 min	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 B.3: First-stage, Figure 3b

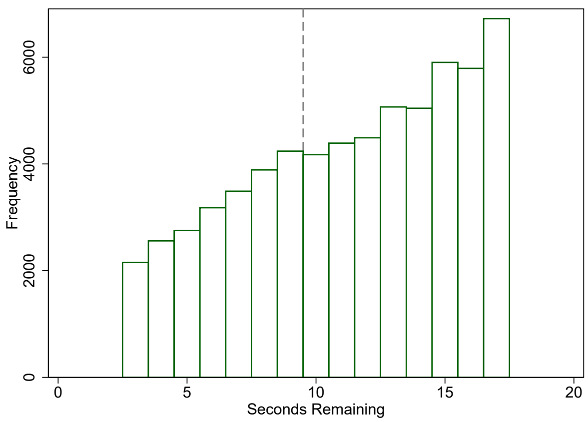
	(1) 30 second	(2) 1 min	(3) 3 min	(4) 5 min	(5) 10 min	(6) 15 min
Time Control Indicator	-0.302***	-0.440***	-2.103***	-1.721***	-3.533***	-2.363***
	(0.00788)	(0.0103)	(0.0132)	(0.0319)	(0.0403)	(0.134)
Observations	33506	43516	225082	122336	229570	49644

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.

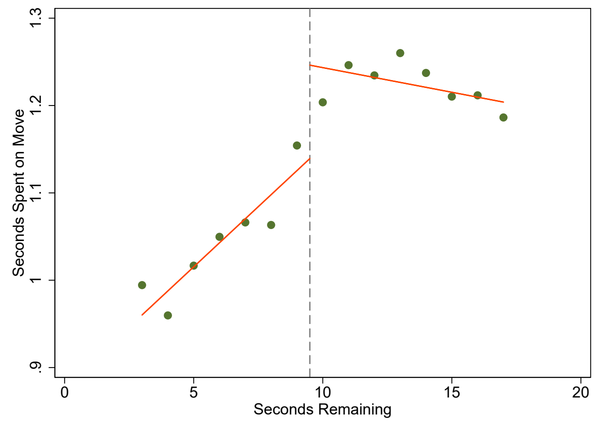
Table B.4: IV Regression, Figure 3b

	(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

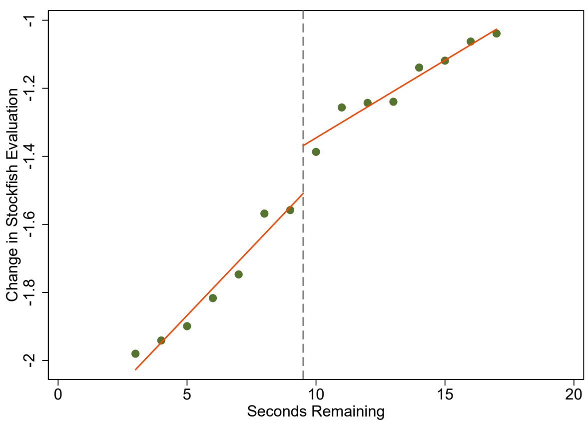
Notes: Instrument variables regression estimating the marginal benefit of attention. Each column represents the regression of the probability of winning based on the position and the clocks in the designated time control on the time spent on the move, using matched moves from the next lowest time control 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.



(a) The Distribution of Time Remaining After a Move



(b) The Amount of Time Spent



(c) The Change in Computer Evaluation

Figure B.1: **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.
Source: Lichess, author's calculations.

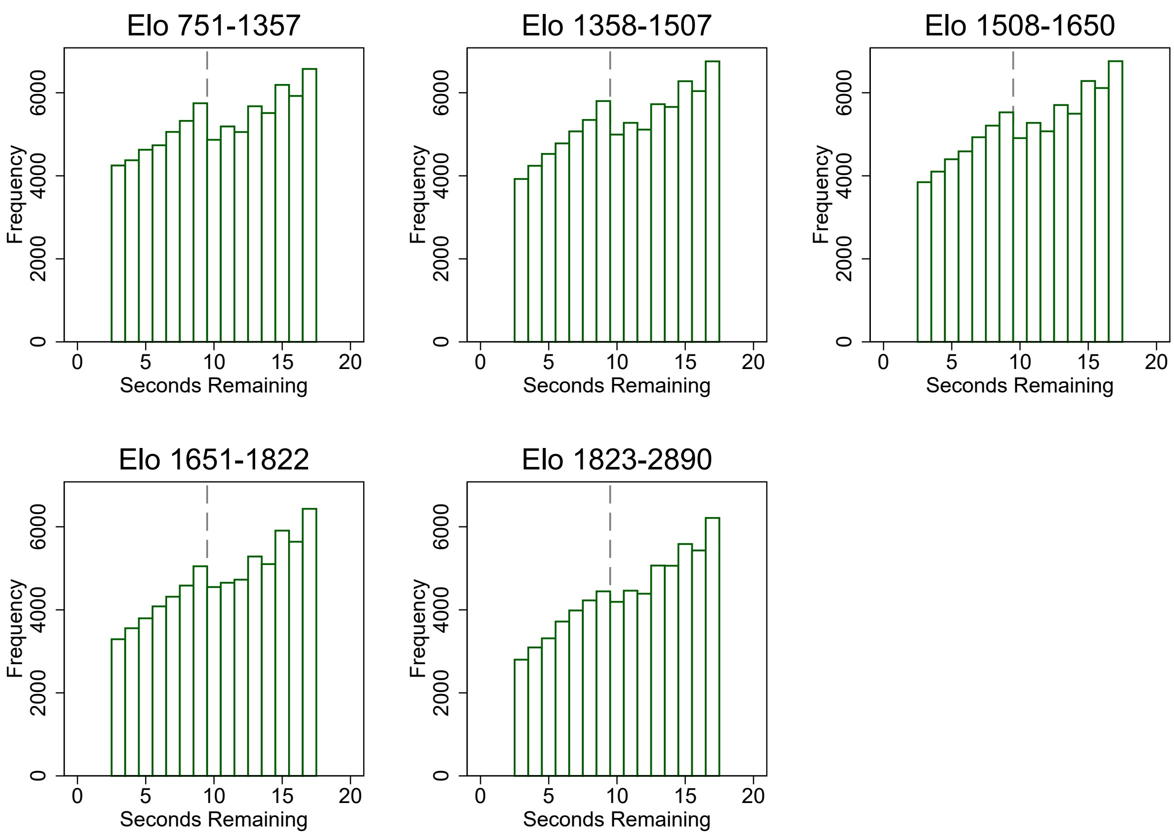


Figure B.2: **Distribution of the Time Remaining, Split over Elo-rating bins.** The figures shows a histogram of the number of moves played by the number of seconds remaining. The number of seconds remaining is measured discretely. The beep occurs when there are 10 seconds remaining.
Source: Lichess, author's calculations.