Complexity and Satisficing: Theory with Evidence from Chess*

Yuval Salant, *Northwestern University*
Jörg L. Spenkuch, *Northwestern University*

August 2023

Abstract

We develop a satisficing model of choice in which the available alternatives differ in their inherent complexity. We assume—and experimentally validate—that complexity leads to errors in the perception of alternatives' values. The model yields sharp predictions about the effect of complexity on choice probabilities, some of which qualitatively contrast with those of maximization-based choice models. We confirm the predictions of the satisficing model—and thus reject maximization—in a novel data set with information on hundreds of millions of real-world chess moves by highly experienced players. These findings point to the importance of complexity and satisficing for decision making outside of the laboratory.

*For helpful conversations and suggestions we thank Sandeep Baliga, Tim Feddersen, Peter Klibanoff, Daniel Martin, Pablo Montagnes, David Smerdon, Uri Zach, as well as audiences at Boston College, Hebrew University, Monash University, Northwestern, Tel-Aviv University, University of Alicante, UC Santa Barbara, the 2022 Barcelona Summer Forum, 2022 North American Summer Meetings of the Econometric Society, and 2021 SITE Conference. We are especially grateful to Sanket Patil, who provided excellent feedback in various stages of this project. The research in this paper was supported in part through the computational resources provided by the Quest high-performance computing facility at Northwestern University. The experiment in Section 4 has been reviewed by the Northwestern IRB, and was preregistered in the OSF Registries at `https://osf.io/6zk9m`. Correspondence can be addressed to the authors at: MEDS Department, Kellogg School of Management, 2211 Campus Dr, Evanston, IL 60208; or by email to `y-salant@kellogg.northwestern.edu` and `j-spenkuch@kellogg.northwestern.edu`.

1. Introduction

The goal of this paper is to better understand decision making when the relevant objects are inherently complex. Insurance contracts, for example, might consist of tens or even hundreds of clauses that jointly determine value. Durable goods can have dozens of relevant attributes, and strategies in dynamic games sometimes include so many contingencies that even enumerating them exceeds the limits of human cognition. A common thread in these and many other examples is that the objects are so large and evaluating them requires so many mental calculations that individuals may struggle to accurately assess their value.

Our analysis begins by modeling the idea that complexity makes it harder to assess value. Each alternative in our model is characterized by its value to the decision maker (DM) and its inherent complexity. When assessing an alternative's value, the DM only obtains a noisy estimate, whose dispersion increases in the complexity of the object. As a consequence, the DM's perception of value is less accurate for objects that are more complex.

We incorporate this notion of complexity into an empirically testable theory of choice. Here, we build on Simon's (1955; 1972) seminal work on bounded rationality and satisficing. According to Simon, individuals may not consider all possible alternatives and pick the best one, but examine a rather small number, making a choice as soon as they find an alternative that they regard as satisfactory. In our model, the decision maker has in mind an aspiration level that she wishes to exceed. She lists all available alternatives in random order, and sequentially evaluates them until she encounters one whose estimated value exceeds her aspiration level. This is the alternative she chooses.

After developing the key predictions of our satisficing-with-evaluation-errors model, we compare them to those of maximization-based choice models. Following Manzini and Mariotti (2014), we postulate a two-stage maximization procedure that includes standard maximization as a special case. In the first stage, the DM reduces the set of available alternatives by drawing a consideration set. In the second stage, the DM evaluates all alternatives in the consideration set, and chooses the one with the highest estimated value. We depart from Manzini and Mariotti (2014) in assuming that object evaluations are noisy and depend on complexity.

Our main theoretical result establishes that satisficing and maximization yield qualitatively different predictions about the effect of object complexity on choice probabilities. Under satisficing, increasing the complexity of a high-value alternative increases the choice probabilities of any other available alternative. Maximization, however, predicts that such an increase in complexity reduces the choice probabilities of objects with weakly higher value. The theory thus points to a new empirical test that leverages object complexity to distinguish between satisficing and maximization.

In our empirical analysis, we implement this test in the context of chess endgames. As a finite, two-player, zero-sum game with perfect information, chess is theoretically trivial; yet it remains practically intractable. Evaluating individual moves often strains the bounds of human cognition, which makes chess an almost ideal setting to study the role of complexity and satisficing in decision making.

In chess, every board configuration corresponds to a choice set in which the alternatives are all available legal moves. By Zermelo's Theorem (1913), any chess move is of one of three types. A *winning* move allows the current player to force a win under subsequent optimal play. A *losing* move enables her opponent to guarantee himself a win, whereas a *drawing* move lets both players force a draw. While computing these types is generally infeasible in the opening and middlegame phases, endgames with up to six pieces have been definitively solved by modern computers. Unlike human players, we can therefore assign an unambiguous, ordinal measure of value to virtually any endgame move.

Chess also admits natural proxies for object complexity. As in any dynamic game, assessing the value of a chess move requires the DM to examine the ensuing subgame. Because larger game trees are likely harder to evaluate than smaller ones, we posit that the complexity of a particular move is closely linked to the size of the subgame. Although computational constraints prevent us from calculating the total number of nodes in every relevant game tree, we can proxy for the size of each subgame by determining its "depth" and "width." Our measure of subgame depth corresponds to what chess players call depth to mate (DTM). It is a theoretical metric of how fast the dominant player can force a checkmate when the losing player resists as long as possible. By width we mean the number of moves that are available to the opponent directly after the current player makes a particular choice. By construction, both depth and width are strongly correlated with the number of nodes in the subgame.

The empirical analysis has three parts. First, to test the idea that complexity leads to evaluation errors, we conduct an online experiment with nearly four thousand chess players. In the experiment, each participant is asked to assess the type of a particular move (i.e., winning, drawing, or losing) in twenty-five randomly chosen endgame positions.[1] Consistent with our notion of complexity, we find that the accuracy of participants' responses declines significantly in moves' inherent complexity. That is, more complex moves are more difficult to evaluate.

In the second part of the empirical analysis, we test the theory's predictions about the effect of object complexity on choice probabilities. Data on choice behavior come from `lichess.org`, one of the three most popular internet chess servers. We have information on the universe

[1]The instructions carefully defined each type of move, although this may not have been necessary given the subject population. About 78% of participants indicated that they had already known about winning, drawing, and losing moves before encountering our definitions.

of moves in all rated games on the platform from January 2013 through August 2020.[2] Our analysis focuses on choices in endgame positions by nearly a quarter million highly experienced users. In total, we examine about 227 million choices from sets with approximately 4.6 billion alternatives.

As predicted by the satisficing-with-evaluation-errors model, we find that, for winning moves, higher complexity is associated with a lower probability of being chosen. For losing moves the opposite holds.

Next, we directly pit satisficing against maximization. To this end, we ask how increasing the complexity of one winning move affects the choice probabilities of *other* winning moves in the same set. Under satisficing, these choice probabilities should increase, whereas they should decrease if players are maximizing. Regardless of whether we rely on depth or width to measure complexity, whether we consider only small choice sets, or restrict attention to games with long time controls, the data are inconsistent with maximization.

This finding raises the question of how widespread departures from maximization are. Are we rejecting the null hypothesis of maximization because some or because most of the DMs in our data appear to be satisficing instead? To speak to this question, we go on to test the null on the individual level. Focusing on players for whom we observe at least one thousand choices, we statistically reject (at the 5%-significance level) maximization for more than 80% of individuals.

We conclude the empirical analysis by studying response times. Based on the idea that complex alternatives are more difficult to evaluate, we hypothesize that object complexity and response times should be correlated in our experiment. This is, indeed, the case. Turning to the Lichess data, we ask whether players' deliberation time prior to choosing a move depends on the composition of values in the choice set (i.e., the number of winning, drawing, and losing moves). Satisficing predicts that replacing a losing move with a winning move of similar complexity would lead to a decrease in response times. This comparative static holds as well.

Related Literature. The work in this paper speaks directly to the theoretical literature on how complexity considerations affect outcomes in single- and multi-person environments. This research usually conceives of complexity as affecting behavior through constraints on agents' computational abilities and memory (e.g., Neyman 1985; Rubinstein 1986; Abreu and Rubinstein 1988; Kalai and Stanford 1988; Salant 2011; Wilson 2014; Jakobsen 2020). A high-level takeaway is that computational constraints can greatly affect both individual and strategic outcomes. Our contribution relative to extant theoretical work is twofold. First, we

[2] Rated games are consequential in the sense that their outcomes directly affect users' strength ratings and rankings on the site. Anecdotal evidence suggests players care intensely about their ratings.

develop a tractable model of decision making that incorporates complexity at the level of individual alternatives. Second, we provide empirical evidence on the relevance of complexity for real-world decision making.

In addition, our work complements a growing experimental literature on satisficing and complexity in decision making (see, e.g., Huck and Weizsäcker 1999; Gabaix et al. 2006; Bossaerts and Murawski 2017; Oprea 2022; Enke et al. 2023). Rubinstein (2007, 2016), for instance, shows that decisions that require explicit cognitive reasoning take far longer to complete than those that do not. Caplin, Dean, and Martin (2011) provide evidence that individuals rely on satisficing in choice environments in which evaluating each option takes time and effort. Oprea (2020) develops a revealed-preference methodology to measure the cost of complexity. He finds subjects are willing to pay significant amounts in order to avoid tasks that are inherently complex. Overall, laboratory experiments support the idea that complexity affects decision making.

Outside of the laboratory, however, tests of fundamental decision-theoretic concepts remain rare.[3] As Chiappori, Levitt, and Groseclose (2002) note, nonexperimental settings are often intractable, with choice sets that need not be known in their entirety, or even be specified ex ante. Moreover, theoretical predictions may hinge on subtle properties of utility functions, intricacies of payoff structures, and individuals' beliefs—all of which are typically unobserved by the econometrician. As a result, we know little about how complexity and satisficing shape decision making in real-world environments.

Chess endgames provide an almost ideal empirical setting to study this question. In addition to yielding observable variation in complexity and admitting an objective measure of alternatives' value, chess possesses at least three additional attractive features. First, the rules of the game are known to players and there is virtually no uncertainty about primitives such as choice sets. Second, data on chess games are abundant, affording us enough statistical power to test even subtle theoretical predictions. Third, we study experienced players in a familiar environment, thus minimizing the risk that our findings are due to an unfamiliar setting or driven by learning.[4]

Our chief contribution relative to extant experimental work is threefold. First, we document

[3]A related literature asks whether some of the basic tenets of game theory are consistent with observed behavior in different real-world environments. Walker and Wooders (2001), Chiappori, Levitt, and Groseclose (2002), Palacios-Huerta (2003), and Hsu, Huang, and Tang (2007) all study minimax play in professional sports, while Spenkuch, Montagnes, and Magleby (2018) examine backward induction in sequential voting. On the whole, the evidence from these settings corroborates theory more closely than one might have guessed based on an abundance of negative findings from the laboratory (see, e.g., Camerer 2003 for a review).

[4]For conflicting evidence as to whether experience and skill in one strategic environment transfer to another one, see Palacios-Huerta and Volij (2008, 2009), Wooders (2010), Levitt, List, and Reiley (2010), and Levitt, List, and Sadoff (2011).

the importance of complexity and satisficing for decision making outside of the laboratory. Second, we provide evidence to suggest that complexity is a key driver of evaluation errors. Third, we develop a new empirical test that has the potential to distinguish satisficing from maximization-based choice behavior in both observational and experimental data.

2. Theory

Our analysis begins by developing a model of decision making in which the available alternatives differ in their inherent complexity. The model has two key components. The first one formalizes the idea that complexity leads to errors in the perception of value. The second component incorporates our notion of complexity into satisficing behavior. After deriving comparative statics on how complexity affects choice probabilities when decision makers (DMs) are satisficing, we establish that a maximization-based choice procedure would yield predictions that are qualitatively different.

2.1. *Complexity and Evaluation Errors*

Let X be a finite grand set of alternatives. An object in X is characterized by a pair (v, σ), where v denotes the value of the object and σ is its inherent complexity. To fix ideas, it is useful to think of complexity as the size of the respective object, or the number of mental operations that are required to calculate its value. For example, the complexity of a strategy in a dynamic game may be a function of the number of nodes in the ensuing subgame, whereas the complexity of a contract might be approximated by the number of non-redundant clauses.

The DM does not know the value of any of the alternatives. When assessing an alternative's value, she obtains a scalar score u, which is drawn from a non-degenerate cumulative distribution function (CDF) F, which depends on v and σ. The score corresponds to the alternative's perceived utility at the end of the DM's evaluation process. The score distribution summarizes all possible perceived utilities after deliberation, or, alternatively, the distribution of scores in a population of DMs.

Here are a few examples of possible evaluation processes and the resulting score distributions.

Example 1 (Maximum Likelihood): *The DM has no prior knowledge about v. She obtains a signal and conducts maximum likelihood estimation to determine the most likely value of the object. The score u then corresponds to the maximum likelihood estimate given the signal realization. For example, if the signal is drawn from a normal distribution with mean v and standard deviation σ, then the score is also distributed $N(v, \sigma)$. If the DM takes "several looks" at the object, i.e., obtains k i.i.d. draws from $N(v, \sigma)$, then the score distribution becomes $N\left(v, \frac{\sigma}{\sqrt{k}}\right)$.*

Example 2 (Bayesian Updating): *The DM has a prior belief about the value of the object, which she updates based on the signal(s) she receives. The score corresponds to the mean of her posterior.*

Example 3 (Partial Confidence): *The DM is partially confident that the value of the object equals $\hat{v}$. She consults a supplementary source of information to obtain a potentially alternative value y, and forms the score $\alpha\hat{v}+(1-\alpha)y$, where α is her initial degree of confidence in $\hat{v}$. The score is then distributed according to the respective linear transformation of the distribution of the supplemental information.*

Let f denote probability density function (PDF) associated with the score distribution F and let $\mu = \mu(v)$ be the mean score according to f. We assume that the score distribution has the following three properties:

(i) *Responsiveness*: The mean score $\mu(v)$ increases in v. We allow $\mu(v)$ to differ from v because we want to accommodate evaluation processes as in Examples 2 and 3.

(ii) *Symmetry*: The density f satisfies $f(\mu - \epsilon) = f(\mu + \epsilon)$ for any $\epsilon \in \mathcal{R}$. Symmetry says that the DM does not systematically over- or underestimate value beyond the distortion allowed by responsiveness.

(iii) *Unimodality*: The density f weakly increases to the left of μ. The essence of unimodality is that tail scores are less likely than "about average" realizations.

In our theory, complexity increases the amount of noise that the DM needs to contend with in assessing value. Complexity is therefore a property of the family of the score distributions that are associated with different objects in X. We require that this family satisfies:

Condition 1: *For every two alternatives a and b in X with values v_a and v_b and $\sigma_a < \sigma_b$, the corresponding CDFs, F_a and F_b, satisfy*

$$F_b(\mu(v_b) - \epsilon) - F_a(\mu(v_a) - \epsilon) \geq 0$$

for any $\epsilon > 0$, with strict inequality whenever $F_b(\mu(v_b) - \epsilon) > 0$.

An increase in complexity thus corresponds to a shift of probability mass from the center of the distribution to its tails.

Several well-known families of distributions satisfy responsiveness, symmetry, unimodality, and Condition 1. A leading example is the family of normal distributions when, for any alternative in the choice set, the mean and standard deviation of the associated score distribution are increasing functions of v and σ, respectively. Another example is the family of uniform distributions, where for any alternative, the associated score is distributed on

$[\mu(v)-\sigma, \mu(v)+\sigma]$. Yet other examples include the Logistic and Laplace families with location parameters corresponding to objects' values and scale parameters corresponding to their complexities.

2.2. *Satisficing Behavior*

Following Simon (1955), we assume that the DM has in mind an aspiration level T that she wishes to exceed. This aspiration level corresponds to the minimal score that the DM regards as satisfactory. When choosing from a set of alternatives $A \subseteq X$, the DM first lists all objects in A in some random order. She then proceeds by sequentially evaluating the available alternatives. Starting with the first one, the DM examines the current object in order to obtain its score. The alternative is chosen if the score exceeds T. Otherwise, the DM proceeds to the next object. She continues in this fashion until she makes a choice or until she reaches the end of the list. In the latter case, the DM chooses the last alternative she evaluated.[5]

We allow for any distribution of evaluation orders that assigns positive probability to all orderings and satisfies *value invariance*. Value invariance means that if any two orderings of alternatives, O_1 and O_2, give rise to the same sequence of values, then the probabilities assigned to O_1 and O_2 are the same.[6]

The outcome of this satisficing-with-evaluation-errors procedure can be summarized by a random choice function C that maps every choice set A and every $a \in A$ to the probability $C(a, A)$ of selecting a from A. Choice behavior is stochastic because object evaluations are noisy and because the evaluation order is random.

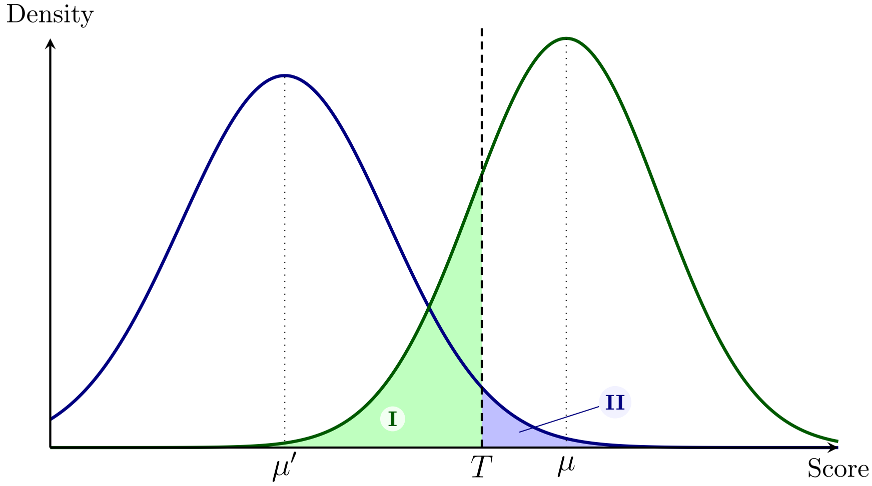
2.3. *Comparative Statics*

To develop intuition for how complexity affects choice probabilities, consider Figure 1. The figure depicts two normal score distributions—one for an alternative with expected score above T, and another one for an alternative with expected score below T. In this setup, higher object complexity directly corresponds to higher variance. Hence, all else equal, an increase in the complexity of an alternative with expected score above T leads to more probability

[5] An alternative formulation would be to assume that the DM recalls all scores and that she chooses the object with the highest score whenever she exhausts all alternatives without finding a satisfactory one. This form of recall makes satisficing closer to standard utility maximization and hence the distinction between the two models less sharp. This is because choice probabilities, when stopping prior to the last alternative, follow the predictions of satisficing, whereas choice probabilities when all alternatives are exhausted follow the predictions of utility maximization. The likelihood of the latter event declines exponentially fast as the number of alternatives in the set grows.

[6] Let a_k^n denote the nth alternative in ordering O_k. Value invariance requires that if two orderings, O_k and $O_{k'}$, satisfy $v_{a_k^n} = v_{a_{k'}^n}$ for all $n = 1, \ldots, |A|$, then O_k and $O_{k'}$ are drawn with the same probability.

Figure 1: An Example of the Noisy Evaluation Process



Notes: Figure illustrates the evaluation process in our model when scores are normally distributed. The solid green line shows the PDF of the score for an alternative with $\mu > T$, whereas the solid blue line corresponds to the PDF of the score for an alternative with $\mu' < T$. The dotted lines mark the expected scores of both alternatives, i.e., μ and μ'. The dashed line marks the DM's aspiration level, T.

mass in area I, which in turn implies that, conditional on being examined by the DM, the corresponding object is chosen with lower probability. As for the remaining alternatives, their choice probabilities do not change if they were examined prior to the alternative that is now more complex. The choice probabilities of all subsequent alternatives, however, increase because they must offset the decline in the choice probability of the object that is now more complex. By contrast, for an alternative with expected score below T, an increase in object complexity leads to more probability mass in area II, which means that the comparative statics reverse.

The following proposition establishes that these predictions carry over to any family of distributions satisfying responsiveness, symmetry, unimodality, and Condition 1.

PROPOSITION 1: *Consider two alternatives a and b with the same value v and with $\sigma_a < \sigma_b$ such that $F_b(T) \notin \{0, 1\}$. Let A and B be two choice sets such that $\{a\} = A - B$ and $\{b\} = B - A$.*

If $\mu(v) > T$, then:

(a) The choice probability of a in A is larger than the choice probability of b in B.

(b) The choice probability of any other alternative in A is smaller than that of the same alternative in B.

If, however, $\mu(v) < T$, then the reverse rankings of choice probabilities hold.

2.4. *Satisficing vs. Maximization*

Our choice model combines two related but conceptually distinct ideas. First, we stipulate a noisy evaluation process that resembles the familiar random-utility framework in discrete-choice models (Luce 1959; Marschak 1960; McFadden 1974). An important difference between our theory and the standard discrete-choice setup is that the errors in our model are due to difficulty in coping with complexity. This feature is important for generating predictions on how complexity affects choice probabilities. Second, we incorporate noisy evaluations into satisficing choice behavior. In doing so, we depart from the standard paradigm of utility maximization; and it is a priori unclear that such a departure is warranted.

To address this question, we first need to specify what we mean by maximization. In the tradition of random-utility models, maximization postulates that the DM considers all alternatives in the choice set, assesses their values, and chooses the one with the highest score. This *full*-maximization assumption is plausible in many settings. It may be demanding, however, when choice sets are very large, when the available alternatives are complex, or when choices need to be made under time pressure. In such environments, it is possible that the DM conducts *partial*-maximization. That is, the DM may first identify a consideration set (i.e., a subset of the available alternatives), assess only the objects in this set, and then choose the alternative with the highest score from within the consideration set.

When comparing the model predictions under satisficing with those under maximization, we would like to allow for both full and partial maximization. We, therefore, follow Manzini and Mariotti's (2014) model of maximization from consideration sets. Our key departure from Manzini and Mariotti (2014) lies in the assumption that object evaluations depend on complexity.

Formally, we say that the DM applies a *maximization-from-consideration-sets procedure* if for every choice set A, the DM follows a two-stage process. In the first stage, the DM draws a consideration set $S \subseteq A$, with $|S| \geq 2$, according to some probability distribution P_A. In the second stage, the DM relies on the noisy evaluation process above to assess the values of all objects in S, after which she chooses the alternative with the highest score.

We require that the family of distributions $\{P_A\}$ satisfies a value-invariance property that is analogous to the one for satisficing. Specifically, for any two choice sets A and B and any two corresponding consideration sets S_A and S_B, we require that S_A and S_B are drawn with the same probability, i.e., $P_A(S_A) = P_B(S_B)$, if the composition of values in A and S_A is the same as that in B and S_B, respectively.[7] Since full maximization trivially satisfies value invariance, it is a special case of the maximization-from-consideration-sets procedure.

[7]Let $|v_a|^A$ denote the number of alternatives with value v_a in A. Two sets A and B have the same composition of values if, for every $a \in A \cup B$, $|v_a|^A = |v_a|^B$.

With this definition in hand, we are able to establish that satisficing and maximization from consideration sets yield qualitatively different predictions about the effect of object complexity on choice probabilities.

PROPOSITION 2: *Suppose the DM uses a maximization-from-consideration-sets procedure. Fix two alternatives a and b with the same value v and with $\sigma_a < \sigma_b$. Let A and B be two choice sets such that $\{a\} = A - B$ and $\{b\} = B - A$. Then, the choice probability of every alternative $c \in A \cap B$ with value $v_c \geq v$ is weakly larger in A than in B.*

It is larger if (i) the supports of the score densities of a and c are not finite, and (ii) there exists a consideration set that is drawn with positive probability and contains a, c, and at least one more alternative.

In words, Proposition 2 implies that, under maximization from consideration sets, an increase in the complexity of one object reduces the choice probabilities of alternatives with weakly higher values.

To illustrate the driving force behind this result in the context of a simple example, consider two alternatives, a and c, with $v_c > v_a$. Assume also that the DM can assess the value of c almost perfectly, so that almost all the mass of the score distribution of c is concentrated in a (narrow) interval above $\mu(v_a)$. For an increase in the complexity of a to affect the choice probability of c, both alternatives must be part of the DM's consideration set. For such a consideration set, the DM would choose a over c only if the score of the former exceeds that of the latter. Since higher complexity corresponds to more probability mass in the tails of the score distribution, an increase in the complexity of a makes it more likely that the associated score exceeds the score of c.

Proposition 2 conflicts with Proposition 1(b). If DMs are satisficing, then an in increase in complexity should result in *higher* choice probabilities for any other object in the choice set. Thus, taken together, Propositions 1 and 2 provide a theoretical foundation for a new empirical test that leverages object complexity to distinguish between satisficing and maximization from consideration sets. In Section 6, we implement this test to detect satisficing in the context of chess.

3. Application to Chess

3.1. *Model Primitives*

In chess, the grand set of alternatives X includes all legal moves in all board positions, and a choice set corresponds to the collection of all legal moves in a given position. By Zermelo's Theorem, starting from any given position, either White can force a win, Black can force a win, or both sides can guarantee themselves a draw. It is therefore possible to associate every move in any board configuration with the ultimate outcome of the game under subsequent optimal play. A move that allows the DM to force a win yields the largest payoff W, whereas a move that enables her opponent to do so produces the lowest payoff L. Moves that lead to draws generate a payoff of $L < D < W$. Hence, for any move a, we have that $v_a \in \{W, D, L\}$.

Since assessing a move's value requires the DM to examine contingencies in the ensuing subgame, we equate the complexity of a given move with the number of subsequent contingencies—or the size of the game tree following this move. Given that we empirically analyze hundreds of millions of choices from sets with several billion alternatives, calculating the exact size of every subgame in our data is computationally infeasible. We, therefore, settle on two proxies: a subgame's "depth" and "width." By width, we mean the number of moves that are available to the opponent directly after the current player makes a particular choice.[8] By depth, we refer to the number of moves until mate if the dominant player attempts to win as quickly as possible, while her opponent resists as long as possible. The latter metric assumes best-response play, and is commonly known as depth to mate (DTM). In the data, both depth and width are predictive of mistakes, suggesting that they are, indeed, related to complexity (cf. Sections 4.2 and 5.2).

3.2. *Measuring Value and Complexity*

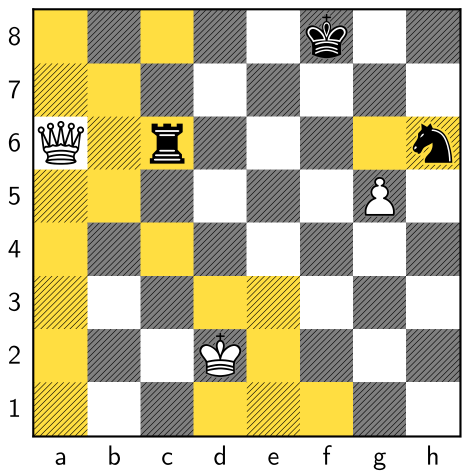
While it is theoretically possible to compute the value of any legal move in any stage of a chess game, doing so in the opening and middlegame phases is computationally infeasible. Our empirical analysis therefore focuses on endgame positions with up to six pieces on the board, which have been definitively solved by computer algorithms.

These algorithms begin by constructing an exhaustive list of all possible (up to symmetry) legal board configurations with three chess pieces.[9] Every configuration is examined, and the ones in which the player to move is in checkmate are stored as "mated in 0." Next, all configurations with the other side to move are evaluated. If one of them can reach a configuration that has previously been determined to be "mated in 0" by executing a legal move, then it is stored as "mate in 1." To find the set of configurations that are "mated in 2," the algorithm looks for configurations from which *all* possible legal moves lead to "mate in 1" configurations; and to determine configurations that are "mate in 3," it subsequently checks for configurations from which it is possible to directly reach a configuration that is known to be "mated in 2." Proceeding recursively, a configuration is classified as "mated in l" if every legal move results in a configuration that is "mate in $w \leq l - 1$, with equality for at least

[8] Almost mechanically, width is positively correlated with the number of possible moves that are available to the current player after the opponent moves, the number of moves available to the opponent after the current player moves again, and so on.

[9] Configurations with two lone kings are automatically drawn.

Figure 2: Example of an Endgame Table



Move	Evaluation
♕c6	W in 19
♕a8	W in 21
♕a3	W in 47
♕a7	D
♕a5	D
♕a4	D
♕a2	D
♕a1	D
♕b5	D
♕d3	D
♕e2	D
♕f1	D

Move	Evaluation
♕b7	D
♙h6	L in 34
♕c4	L in 34
♕c8	L in 34
♔e2	L in 32
♔e3	L in 32
♔d3	L in 32
♕b6	L in 32
♔e1	L in 30
♔d1	L in 28
♙g6	L in 28

Notes: Figure provides an example of the information in endgame tablebases. The left panel shows the board configuration that is to be evaluated, assuming it is White's turn to move. Yellow-colored squares help visualize the set of available moves. The right panel shows the computer evaluation of each legal move, drawing on the Nalimov endgame tables. The letters W, D, and L denote winning, drawing, and losing moves from the perspective of the current player.

one move. By contrast, a configuration is marked as "mate in w" if it is possible to move to another one that is "mated in $w-1$." This procedure continues until no further progress at classifying configurations is made, at which point all remaining configurations with three chess pieces are designated as "drawn." Essentially the same algorithm is next applied to board configurations with four pieces, then five, and then six.

The end result is a so-called tablebase in which *board configurations* are classified as either "drawn," "mated in l," or "mate in w." A particular *move* is said to be of type W with DTM d if it results in a new board configuration that, with the other player to move, is known to be "mated in $d-1$." Thus, the minimal DTM among all available W-moves from any configuration that is "mate in w" is, by construction, equal to w. Similarly, a move is said to be an L-move with DTM d if it leads to a configuration that is "mate in $d-1$." The maximal DTM among all L-moves from any configuration that is "mated in l" equals l. Moves that result in "drawn" configurations are classified as type D.[10]

Figure 2 provides a concrete example of the content of a tablebase. The left panel depicts the board configuration that is being examined, with the data for each available legal move shown on the right. The assessment of a move consists of two components: its type (i.e., W, D, or L), and, for a W- or L-move, its DTM. In this particular example, ♕c6 corresponds to "W in 19," which means that, if White moves the queen to c6, then White can force checkmate in nineteen moves regardless of Black's response.

Note, tablebases do not contain a measure of subgame depth for D-moves. To the best of

[10]For additional information on algorithmic analysis of chess endgames, see, e.g., Thompson (1986).

our knowledge, there is no general approach to even identify D-moves by means other than elimination, which does not lend itself to measuring subgame depth. We therefore refrain from quantifying depth for D-alternatives, and restrict our empirical tests to W- and L-moves.

For the analyses below, we look up the value of every endgame move in extant tablebases. In the case of W- and L-moves, we also retrieve their DTM. Since tablebases do not record subgame width, we compute this alternative measure of complexity by counting the number of legal moves in any board position that can be reached by some move in our data.

3.3. *Predictions*

In order to translate the model's comparative statics into concrete predictions for the case of chess, we need to specify players' threshold scores. We assume that the average score associated with a winning move, $\mu(W)$, exceeds the DM's aspiration level, whereas the average score of a losing move, $\mu(L)$, is below the threshold.

ASSUMPTION 1: *The threshold T is between $\mu(L)$ and $\mu(W)$.*

In other words, if evaluations were not noisy, players would find W-moves acceptable but reject L-alternatives.

Under this assumption, we have the following testable predictions of the satisficing-with-evaluation-errors model (cf. Proposition 1):

PREDICTION 1: *Holding the values and complexities of all other moves in the choice set fixed, an increase in the complexity of a W-move decreases the frequency with which this move is chosen. For an L-move, however, an increase in complexity leads to a higher choice frequency.*

To distinguish satisficing from maximization, we examine how the complexity of W-moves affects the choice frequency of other W-moves in the same set (cf. Propositions 1(b) and 2).

PREDICTION 2: *Under satisficing, an increase in the complexity of one W-move increases the choice frequency of all other W-moves in the choice set. Under maximization, however, an increase in the complexity of a W-move decreases the choice frequency of all other W-moves.*

4. Does Complexity Lead to Evaluation Errors?

The fundamental idea behind our notion of complexity is that higher complexity leads to noisier perceptions of value (cf. Condition 1). Noisier perceptions should in turn increase the frequency with which DMs misclassify moves, e.g., identify a W-move as a D- or an L-move. To test whether complexity is indeed associated with errors of this kind, we conducted an experiment with nearly four thousand online-chess players.

4.1. *Experimental Design*

The experiment took place on a custom-built website over the four-week period starting April 7, 2023. We recruited participants via targeted ads on social media and through forum posts on two of the largest online chess platforms, `lichess.org` and `chess.com`. To ensure that we were recruiting online-chess players, we required all participants to provide their Lichess and Chess.com usernames, which our website verified in real time. Out of the 3,966 participants, 584 provided a Lichess username, 2,471 submitted a Chess.com username, and 911 subjects provided both.

The experiment consisted of twenty-five rounds. In each round, participants were shown a chess board with a randomly sampled endgame position in which one legal move was highlighted. They were then asked to indicate whether the highlighted move is a winning, drawing, or losing move—as in the example in Figure 3. Subjects had between five and forty-five seconds to submit their answer, and they knew that moves of each type were a priori equally likely to be shown.[11] We settled on this experimental task because it allows us to relate moves' complexity to the accuracy of participants' subjective evaluations. At the same time, it resembles the kind of puzzles that are popular among chess players.[12]

In order to present subjects with moves that they might realistically evaluate in a real-world chess game, we extracted a random subset of 30,000 legal moves from a representative set of board configurations in our observational data from Lichess, which is analyzed in Sections 5–7.[13] We then constructed sampling probabilities so that participants could expect to see an equal number of W-, D-, and L-moves, subject to depth being approximately uniformly distributed between zero and fifty. Importantly, subjects were never asked whether they would choose any given move. Our experimental design thus tests the idea that higher object complexity is associated with more classification errors, independent of whichever choice procedure players may use.

We incentivized subjects by awarding one virtual lottery ticket for every move they correctly evaluated. After the experiment, all lottery tickets were entered into a raffle for twenty $100 Amazon gift certificates. The median participant earned 15 tickets and spent about 9 minutes on the experiment.[14] For additional details on the experimental setup, see Appendix D.

---

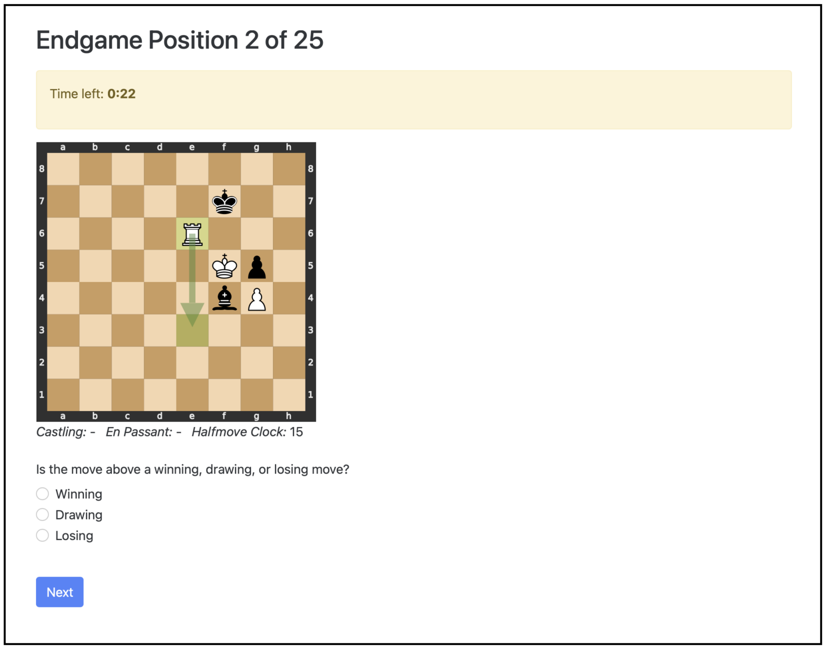
[11] The time limit was uniform i.i.d. across rounds.

[12] We designed the experiment to test different hypotheses. In this paper, we focus one of them (i.e., H1 in the preregistration), leaving tests of other hypotheses for future work.

[13] The only constraint we imposed is that the depth and width of extracted moves do not exceed 50 and 18, respectively. Both numbers correspond roughly to the 95th percentiles of the respective marginal distributions.

[14] About 20% of participants did not finish the experiment. The analysis below uses data from all participants—regardless of the total number of evaluations they submitted—subject to passing basic attention checks. Results are qualitatively and quantitatively similar if we exclude anyone who did not complete the experiment.

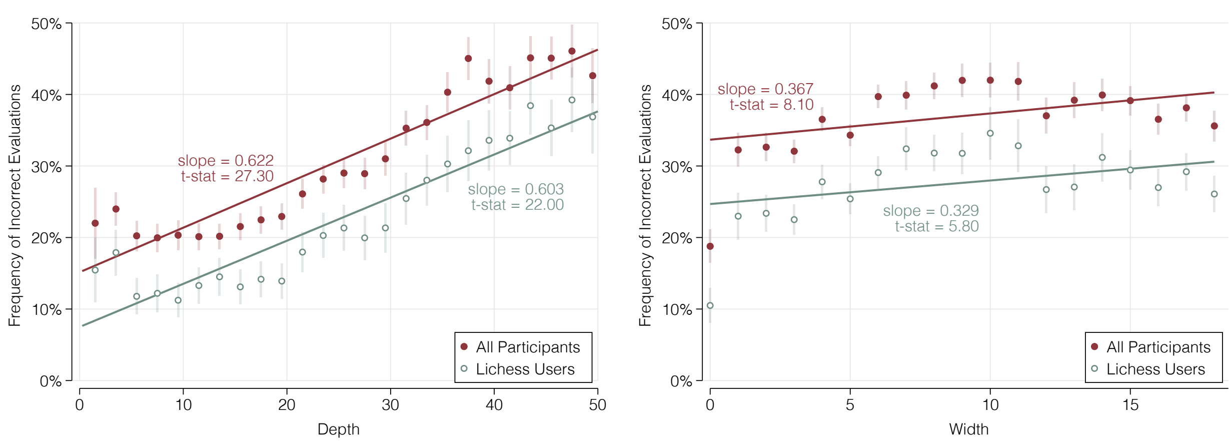
Figure 3: Screenshot of Experimental Task



Notes: Figure shows a screenshot from our experiment, in which subjects are asked to identify the type of a particular move.

Figure 4: Object Complexity Predicts Incorrect Evaluations

(a) Based on Depth (b) Based on Width



Notes: Figure shows binscatter plots of the raw relationship between the frequency of incorrect move evaluations (y-axis) and the respective moves' complexity (x-axis). Panel (a) uses depth to measure complexity, whereas panel (b) relies on width. The underlying data come from the experiment described in the text. When focusing on Lichess users only, observations are reweighted to approximate the distribution of strength ratings in the observational choice data from Lichess that we use in Sections 5–7. Error bars correspond to 95%-confidence intervals, accounting for two-way clustering by participant and move.

4.2. *Experimental Results*

Figure 4 plots the raw frequency of incorrect evaluations against moves' complexity. In the left panel we use subgame depth to measure complexity, while the right panel uses width instead. Since our analyses in Sections 5–7 rely on observational data from Lichess, we present results pooling across all subjects and restricting attention to Lichess users only. For the latter, we reweight observations so that the distribution of strength ratings among Lichess users in our experiment approximates that in the real-world data below.[15]

Consistent with the idea that object complexity injects noise into value assessments, Figure 4 shows that relatively simple moves are more likely to be correctly evaluated than more complex ones. Reassuringly, we observe a similar, approximately linear and statistically significant relationship for all participants and Lichess users only—though the latter do, on average, better. We also observe that evaluation errors are more sensitive to moves' depth than to their width.[16] Among all players, a one standard deviation increase in depth is associated with a 9.0 percentage points (p.p.) increase in the rate of errors, while a standard deviation increase in width is only associated with a 2.0 p.p. increase.[17] Viewed through the lens of our model, this suggests that depth might be a better proxy for object complexity than width.

Table 1 presents results from estimating variants of the following linear probability model:

$$(1) \qquad Incorrect_a = \kappa \, Complexity_a + \psi_A + \xi_a,$$

where $Incorrect_a$ is an indicator for whether the subject made a mistake in identifying the type of move a in endgame position A, $Complexity_a$ denotes the move's complexity (i.e., depth or width), and ψ_A is a fixed effect for the board position, i.e., the exact configuration of all pieces. By including ψ_A we account for general features of the board that might affect subjects' evaluations, such as the number, type and positioning of chess pieces. In our most inclusive specification, we interact ψ_A with a fixed effect for the specific piece executing move a. In these regressions, all identifying variation comes from comparing different moves with the same piece in the same board configuration. For example, in the context of Figure 3 we might be comparing ♖e3 with ♖e7. The former is an L-move with depth equal to 44, while the latter is an L-move whose depth is 40.

[15] The Lichess users participating in the experiment have a strength rating that is nearly 150 points lower than that of the (highly experienced) players in our real-world sample. The results below would be slightly stronger if we did not reweight observations to approximate the ratings distribution in the real-world data.

[16] In the right panel, excluding moves with a width of zero (i.e., moves that result in checkmate or stalemate) would yield a slope estimate of 0.233 with a standard error of 0.046 for all users, and an estimate of 0.204 with a standard error of 0.060 for Lichess users.

[17] In the experimental data, the standard deviation of depth equals 14.4, while that of width is about 5.4.

Table 1: Experimental Results

Panel A: All Online-Chess Players

	Probability of Incorrectly Identifying Type of Move							
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)	(7A)	(8A)
Depth (÷ 100)	0.622		0.987	1.067			0.969	1.053
	(0.023)		(0.042)	(0.046)			(0.043)	(0.047)
Width (÷ 100)		0.367			0.573	0.471	0.548	0.380
		(0.045)			(0.087)	(0.095)	(0.099)	(0.111)
Fixed Effects:								
Board Position	No	No	Yes	No	Yes	No	Yes	No
Board Position × Piece	No	No	No	Yes	No	Yes	No	Yes
Mean of LHS Variable (%)	31.033	36.976	31.033	31.033	36.976	36.976	31.033	31.033
R^2	0.038	0.002	0.139	0.176	0.125	0.165	0.139	0.176
N	58,470	87,060	58,470	58,470	87,060	87,060	58,470	58,470

Panel B: Lichess Users

	Probability of Incorrectly Identifying Type of Move							
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)	(7B)	(8B)
Depth (÷ 100)	0.603		0.829	0.851			0.825	0.850
	(0.027)		(0.059)	(0.070)			(0.060)	(0.070)
Width (÷ 100)		0.329			0.358	0.334	0.144	0.026
		(0.057)			(0.112)	(0.125)	(0.132)	(0.152)
Fixed Effects:								
Board Position	No	No	Yes	No	Yes	No	Yes	No
Board Position × Piece	No	No	No	Yes	No	Yes	No	Yes
Mean of LHS Variable (%)	26.624	32.227	26.624	26.624	32.227	32.227	26.624	26.624
R^2	0.043	0.002	0.212	0.263	0.193	0.254	0.212	0.263
N	22,382	33,422	22,382	22,382	33,422	33,422	22,382	22,382

Notes: Entries are coefficients and standard errors from estimating κ in variants of eq. (1) by ordinary least squares. The set of included fixed effects varies across columns. The unit of observation is always a participant's evaluation of a particular move. There are differences in the number of observations across columns because depth is not defined for D-moves. The regressions in the upper panel use data from all participants in our experiment, whereas those in the lower panel restrict attention to registered users of Lichess. In the latter case, observations are reweighted to approximate the distribution of strength ratings in the real-world Lichess data that we use in Sections 5–7. All estimates are scaled to correspond to the percentage-point change in the probability of incorrectly identifying the type of a move associated with a one-unit increase in the respective regressor. Standard errors are two-way clustered by participant and move, and are shown in parentheses.

The first two columns of Table 1 reproduce the evidence in Figure 3. The results in the next four columns establish that the estimates are robust to controlling for the exact board position and chess piece. If anything, including controls strengthens the relationship between moves' complexity and errors in evaluation. The specifications in the last two columns of Table 1 show that either complexity measure is related to evaluation errors even after controlling for the other one, although the point estimates are not statistically significant for width in the subsample of Lichess users.

In sum, the experimental results support the fundamental idea behind our notion of complexity. More complex moves are more difficult to evaluate.

5. Data on Choice Behavior

Having established a connection between complexity and evaluation errors, we proceed to introduce a new, large observational data set that contains information on choice behavior in chess endgames. These data allow us to test the model predictions on how complexity affects choice probabilities.

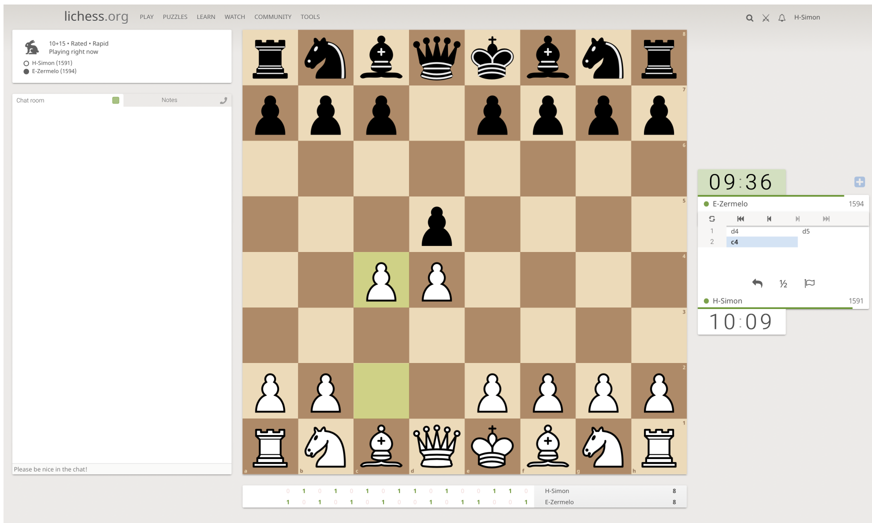
5.1. *Data Sources*

The core of the data comes from `lichess.org`, one of the most popular online chess platforms. Funded by donations, Lichess is ad free and allows anyone to play live chess games at no cost through a high-quality graphical user interface (see Figure 5 for a screenshot of a typical game). Although Lichess offers a choice between many different time limits, the majority of games that are actually hosted on the platform can be broadly classified as "speed chess."[18] Lichess further distinguishes between casual and rated games. The latter determine player ratings and are therefore only available to registered users. In a nutshell, a player's strength rating increases (decreases) whenever she wins (loses) a rated game, and it increases (decreases) by more the stronger (weaker) her opponent was. Anecdotal evidence from online messaging boards suggests that users care intensely about their rating. Since high ratings tend to be a source of pride among chess players, Lichess has a strict policy against computer-assisted play. Enforcement of this policy relies on a variety of methods, including community reporting of suspected offenders and automatic detection algorithms.

We have data on the universe of rated games between human players from January 2013 through August 2020. The available information includes players' usernames, ratings and

[18] The three most popular time control formats on Lichess are Bullet, Blitz, and Rapid. In a 40-move Blitz game, each player has about eight minutes to deliberate. The corresponding numbers for Bullet and Rapid games are three and twenty-five, respectively. Some of our analysis restricts attention to games with Classical and Correspondence time controls, which last longer and in which time pressure tends to be less of an issue. We have also conducted robustness checks in which we directly control for time pressure. The results are qualitatively equivalent to those below (cf. Appendix C).

Figure 5: Screenshot of a Rated Game on Lichess



Notes: Figure shows a screenshot from a rated game between registered users on `lichess.org`. The green squares highlight the most recent move, i.e., ♙c4.

real-world titles (if any), the date and start time of the game, its outcome, as well as the sequence and timing of moves. We can therefore reconstruct all choice sets that a player faced as well as the moves she chose.

We complement these data with information on moves' values and complexities. As explained above, extant computer analyses have determined the values and, for W- and L-moves, the depth to mate of essentially all legal moves in endgame positions with six of fewer pieces on the board (cf. Section 3.2).[19] We retrieve this information by running several billion queries against the Syzygy and Nalimov tablebases (Nalimov et al. 2000; Man 2013).[20] To compute width, we adopt a brute-force approach. For every legal endgame move in the data, we construct the resulting board position and count the number of moves that would be available to the opponent if the current player executed the respective move.

Our final sample contains nearly 227 million decision problems with a total of over 4.6 billion alternatives. There are five distinct sources of selection into this sample. First, because we need information on alternatives' values and complexities, we restrict attention to board

[19] The only exceptions are positions with castling rights and configurations in which a lone king faces five other pieces. The former are extremely rare in endgames ($<.01\%$ of available legal moves in our data), while the latter are uninteresting (because 98.8% of available moves are of type W). Our empirical work excludes all board configurations for which information on DTM is not available.

[20] As a technical side note, the Syzygy tablebases do not contain information on DTM. In contrast to the Nalimov tables, they do, however, take into account the fifty-move stalemate rule. In rare instances, the fifty-move rule matters for correctly determining whether one player can unilaterally invoke a draw. We therefore rely on win-draw-loss information from the Syzygy database, while information on DTM comes from Nalimov's database. The commercially available Lomonosov tablebases contain information on values and DTM for board configurations with up the seven pieces, but require about 140TB of storage. They are thus too large to be usable in most computing environments.

positions with six or fewer pieces.

Second, we focus on choice sets that contain at least one W- and at least one D- or L-move. We adopt this restriction because it enables us to test Predictions 1 and 2 without changing samples. A disadvantage of this restriction is that the sets in our sample contain an above average share of W-moves.

The third and related source of selection pertains to how often different individuals reach a winning position, i.e., an endgame position with at least one W-move. The strongest players, for instance, may often mate their opponents before reaching the endgame stage. Similarly, very weak players may rarely be in a position to win endgames and might thus also be underrepresented in our sample. We address this issue in two ways. First, whenever appropriate, we control for player fixed effects. Second, we reweight observations so that all players receive equal weight in the analysis. The results below should hence be interpreted as referring to a typical decision by the average player in our sample.

Fourth, to minimize the risk that our findings are due to an unfamiliar setting or a lack of experience with similar decision problems, we exclude, for every player, the first one thousand endgame moves from winning positions. This leaves us with approximately 237,000 highly experienced DMs who are very familiar with the task at hand.

Finally, users on Lichess are not a random subset of all experienced chess players. In the appendix, we address this potential source of concern by replicating our main results in an independent data set covering a large number of chess games in international tournaments. These data come from the online publication *The Week in Chess* (TWIC), which covers "all the latest news and games from international chess." The most important disadvantage of this alternative data set is that there is significantly less variation in the skill of players, and that it is several orders of magnitude smaller than the Lichess data. These limitations notwithstanding, the TWIC data yield qualitatively similar conclusions (cf. Appendix C.4).[21]

5.2. *A First Look at the Data*

Table 2 displays summary statistics for select variables in the Lichess data. On average, 15.5 out of 20.6 available moves are W-moves; yet only about 6% of observed choices are mistakes in the sense that a player chooses a D- or an L-move instead of a W-move. Mistakes thus occur at about one quarter the rate one would expect if DMs were choosing at random. At the same time, the raw data also imply that mistakes do occur with a certain regularity. They are not rare events. Moreover, mistakes are consequential. A player whose current move is a

[21] Out of the twenty point estimates in Appendix C.4 sixteen are statistically significant and have the same sign as their counterparts in Tables 4 and 5 below. For the remaining four coefficients the 95%-confidence intervals include both negative and positive values.

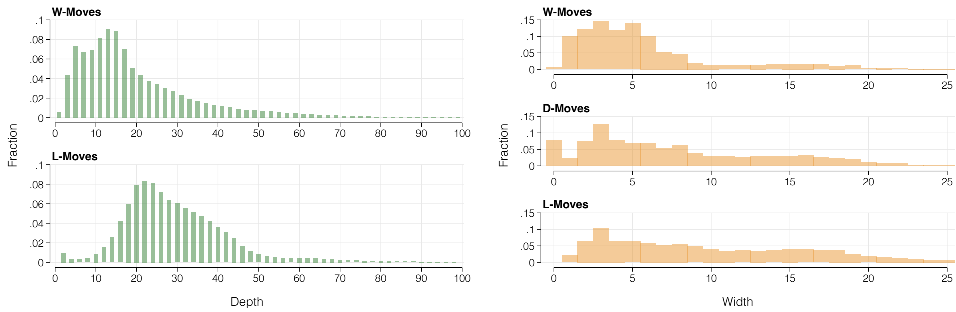
Table 2: Summary Statistics

Variable	Mean	SD	Percentile 25%	50%	75%	95%	N
A. Move Characteristics							
Type:							
W-Move	0.69	0.46					4,617,441,573
D-Move	0.23	0.42					4,617,441,573
L-Move	0.08	0.27					4,617,441,573
Depth:							
W-Moves	25.89	17.88	13	23	33	59	3,457,878,398
L-Moves	30.35	13.46	22	28	36	50	296,522,573
Width:							
W-Moves	6.66	5.01	3	5	8	18	3,457,878,398
D-Moves	6.33	5.55	3	4	8	18	863,040,602
L-Moves	8.95	6.16	4	7	13	20	296,522,573
B. Choice-Set Composition							
Total Number of Legal Moves	20.58	10.35	13	20	28	38	226,955,095
Number of W-Moves	15.48	11.09	6	15	24	34	226,955,095
Number of D-Moves	3.81	4.30	1	2	5	13	226,955,095
Number of L-Moves	1.29	2.73	0	0	2	7	226,955,095
C. Outcomes							
Mistakes:							
Any Type of Error	0.06	0.24					226,955,095
Choose D-Move	0.05	0.23					226,955,095
Choose L-Move	0.01	0.08					226,955,095
Result of Game:							
If Current Move is Mistake:							
Win Game	0.31	0.46					13,052,773
Draw	0.49	0.50					13,052,773
Lose Game	0.19	0.40					13,052,773
If Choose W-Move:							
Win Game	0.74	0.44					213,902,322
Draw	0.20	0.40					213,902,322
Lose Game	0.05	0.22					213,902,322
D. Timing							
Time Left on Clock (in sec.)	72.35	192.87	8	22	69	296	212,295,223
Deliberation Time (in sec.)	1.66	3.07	0	1	2	5	212,249,738
E. Player Characteristics							
Total Number of Endgame Moves	2,584	2,469	1,297	1,793	2,877	6,696	237,232
Average Rating	1,733	281	1,533	1,716	1,917	2,222	237,232
Real-World Title	0.01	0.10					237,232

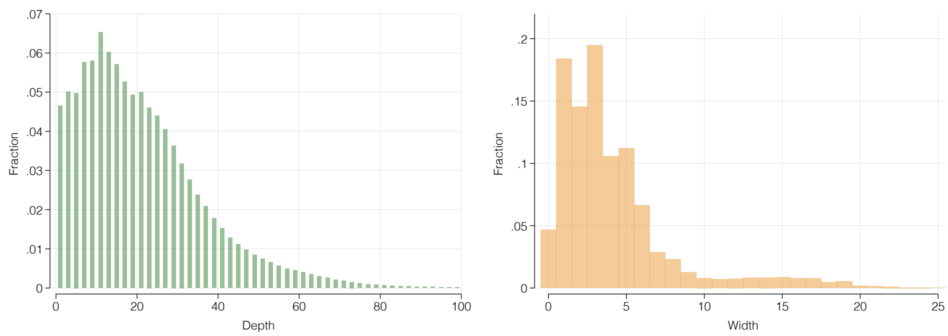
Notes: Table displays summary statistics for selected variables in the Lichess data. Each observation in panel A corresponds to a legal move, and observations in panels B–D correspond to decision problems. Panel E contains player-level information. Observations are reweighted so that all players and all decision problems for a given player receive equal total weight. The number of observations related to the timing of moves is smaller because the raw data do not include this information for games that were played prior to April 2017.

Figure 6: Distribution of Complexity Measures

(a) Depth, by Type of Move | (b) Width, by Type of Move



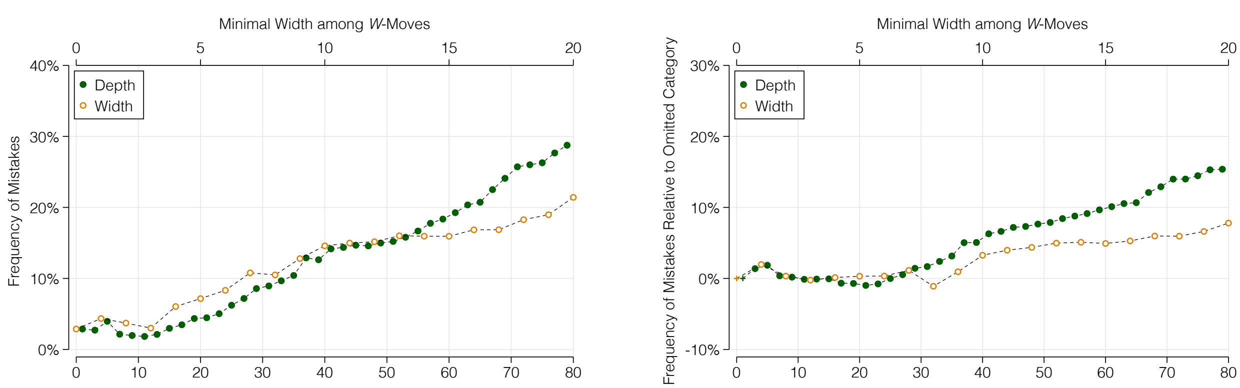
(c) Minimal Depth Among W-Moves | (d) Minimal Width Among W-Moves



Notes: Panel (a) presents a histogram of moves' depth, separately by type of move. Panel (b) does so for moves' width. Panels (c) and (d) respectively depict the distribution of the minimal depth and width among all W-moves in the choice set.

Figure 7: Greater Object Complexity is Associated with More Mistakes

(a) Raw Data | (b) Controlling for Set Composition



Notes: Figure shows the relationship between the frequency of mistakes (y-axis) and the minimal depth and width among the available W-moves (x-axis). Panel (a) does so based on the raw data, whereas a panel (b) presents estimates of the same relationship after controlling for the composition of the choice set, i.e., a fixed effect for the combination of the number of available W-, D-, and L-moves. As explained in the text, the DM is said to make a mistake when she chooses a D- or L-move in the presence of a W-alternative. The graphs do not show confidence intervals because they are too small to be visually apparent.

mistake is about 43 p.p.—or roughly 58%—less likely to ultimately win the game than one who chooses a W-move, while the probability of a loss more than doubles.[22]

We next turn to the distribution of our complexity measures. The upper two panels in Figure 6 display histograms for individual moves' depth (left) and width (right). On average, W-moves have a depth of about 25.9 and a width of 6.7. The corresponding numbers for L-alternatives are 30.4 and 8.9, respectively. Important for our purposes, there is a great amount of variation in depth and, to a somewhat lesser extent, width.

The lower two panels of Figure 6 plot the distribution of the minimal depth (left) and minimal width (right) among W-moves at the choice-set level. From a theoretical perspective minimal depth and width correspond to the lowest amount of complexity with which the DM needs to contend. Empirically, both measures are highly correlated with other summary statistics for the complexity of available alternatives, such as the mean or median depth and width. Taking either measure at face value, the data include choice sets in which evaluating at least some of the moves is relatively easy, others where accurately classifying any move likely exceeds the bounds of human cognition, and a great range of intermediate cases.

Figure 7 plots the minimal depth and width among W-moves against the observed frequency of mistakes. Regardless of whether we rely on minimal depth or width—or other low-dimensional summary measures—we find that complexity predicts mistakes. The left panel of Figure 7 shows this based on the raw data, while the right panel reveals a similar relationship after controlling nonparametrically for the composition of the choice set (i.e., the number of available winning, drawing, and losing moves). In line with our experimental results, the evidence in Figure 7 suggests that depth is a better predictor of mistakes than width. For this reason, we treat depth as our preferred measure of complexity.

One potential issue with relying on depth as a proxy for complexity is that players may care about more than just winning. For instance, preferences may be lexicographic over the outcome of the game and its duration. That is, players may prefer winning to drawing and losing, but winning quickly might be better than winning slowly. If there is such a preference for winning quickly and if players are able to identify the depth to mate of individual moves, then it is possible that high- and low-depth moves may differ not only in their complexity but also in their instrumental value.

To rule out that this possibility drives our results we follow a two-pronged approach. First, we conduct robustness checks in which we test Predictions 1 and 2 for choice sets in which the minimal depth among W-moves exceeds fifty (cf. Appendix C). While it is conceivable

[22] Mistakes do not always result in forgone wins because the player's opponent may subsequently also make a mistake. Similarly, due to potential future mistakes, choosing a W-move now does not guarantee that the player will win the game with certainty.

that players rank winning moves according to depth to mate when some of them are simple, it seems unlikely that they are able to do so when the available alternatives are all very complex. In fact, our experimental data suggests that players often struggle to identify even the value of such highly complex moves.

Second, and perhaps more importantly, we present robustness checks that exploit variation in width *conditional on depth to mate*. To understand why this is helpful, consider Table 3. The table presents regression results that relate both of our complexity measures to the actual length of subsequent play. Given the definition of depth, the positive coefficient in column (1) verifies that subgames that should, in theory, take longer do, on average, take longer.[23] The coefficient in column (2) reveals that width is also positively correlated with length of play.[24] The third column of Table 3, however, demonstrates that the relationship between width and length of play reverses if we condition on the depth of the move. Columns (4)–(9) show that this reversal is driven by W-moves. Upon controlling for depth, there is almost no relationship between the width of L-moves and the length of subsequent play. For W-moves, however, the partial correlation is negative.

The reason for the negative conditional correlation goes back to the definition of depth. Depth to mate is a metric of how quickly the dominant player can force checkmate when the opponent resists as long as possible, i.e., if the opponent always picks the L-move with the highest depth to mate. Choices, however, are noisy. Sometimes the losing player chooses a move that allows for a quicker mate, and the probability of (inadvertently) executing such a move is larger when his choice set contains more alternatives. Thus, conditional on depth, higher-width moves are associated with quicker wins.

We build on this observation to address the possibility that players have a preference for winning quickly. While our preferred specifications rely on depth to measure complexity, we present robustness checks that exploit variation in complexity due to moves' width *conditional on their depth*. For the latter set of specifications, any bias arising from a desire to win quickly would go in the opposite direction.

[23]Depth to mate may differ from the realized length of play for several reasons. For example, if the dominant player succeeds in mating her opponent but does not take the shortest path to victory, then the total number of subsequent moves may exceed the initial move's depth. If, however, the losing player resigns or does not hold out as long as possible, then there will be fewer subsequent moves than implied by depth to mate.

[24]The correlation between depth and width is approximately 0.6, and a simple mediation analysis implies that the coefficients in columns (1) and (2) are consistent with the idea that the "effect" of width on subsequent length of play works entirely through depth.

Table 3: Object Complexity and Length of Subsequent Play

	Number of Subsequent Moves								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Depth	0.340			0.343			0.121		
	(0.000)			(0.000)			(0.002)		
Width		0.731	–0.090		0.758	–0.093		–0.081	0.002
		(0.001)	(0.001)		(0.001)	(0.001)		(0.003)	(0.003)
Fixed Effects:									
Player	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Depth	No	No	Yes	No	No	Yes	No	No	Yes
Type of Chosen Move	W- or L-Move	W- or L-Move	W- or L-Move	W-Moves	W-Moves	W-Moves	L-Move	L-Move	L-Move
Mean of LHS Variable	15.631	15.631	15.631	15.647	15.647	15.647	13.288	13.288	13.288
R^2	0.195	0.126	0.234	0.197	0.129	0.235	0.329	0.314	0.340
N	215,258,880	215,258,880	215,258,880	213,902,320	213,902,320	213,902,320	1,356,566	1,356,566	1,356,566

Notes: Entries are coefficients and standard errors from regressing the number of moves after the current one in the same game on that move's inherent complexity, as proxied by its depth and width. All regressions control for player fixed effects. Other fixed effects vary across columns. Since depth is only defined for W- and L-moves, the sample includes only decision problems in which the DM chose a move of either type. Observations are reweighted so that all decision problems for a particular player and all players receive equal weight. Standard errors are two-way clustered by subject and move, and are shown in parentheses.

6. How Does Complexity Affect Choice Frequencies?

We now proceed to test Predictions 1 and 2. Prediction 1 relates moves' complexity to the probability that the respective moves are chosen. The effect of complexity on own choice probabilities should be negative for W-moves and positive for L-alternatives. Prediction 2 concerns the impact of one W-move's complexity on the choice frequencies of other W-moves in the same set. The sign of this effect allows us to distinguish between satisficing and maximization.

6.1. *Tests of Prediction 1*

To investigate the connection between object complexity and own choice frequencies, we estimate the following econometric model separately for W- and L-moves:

$$Choose_a = \beta\, Complexity_a + \chi_p + \phi_{A \setminus a} + \varepsilon_a. \tag{2}$$

Here, $Choose_a$ is an indicator for whether player p facing choice set A picked move a, $Complexity_a$ denotes the move's depth, χ_p is a player fixed effect, and $\phi_{A \setminus a}$ corresponds to a fixed effect for the other moves in the same choice set. In constructing this fixed effect, we assume that, in line with the theory, moves can be reduced to their types and complexity. Since we do not measure depth to mate for D-moves, $\phi_{A \setminus a}$ conditions (only) on the vector of depth values for W- and L-moves and the type composition of the choice set, i.e., the number of W-, D-, and L-alternatives. By including $\phi_{A \setminus a}$, we aim to approximate the thought experiment in which we vary the object complexity of one particular alternative, holding the values and complexity of all other moves fixed.

As explained above, we complement the results based on this specification with robustness checks that condition on moves' depth and rely on their width as an alternative source of variation in object complexity. In these regressions, $Complexity_a$ corresponds to the width of alternative a and $\phi_{A \setminus a}$ is additionally interacted with the depth of a.

The upper panel of Table 4 shows results from estimating the regression model in eq. (2) using depth to measure complexity, while the lower panel implements our robustness checks based on width. In the first two columns within each panel, we study W- and L-moves from all board configurations. In the last two columns, we restrict attention to choice sets that do not contain any D-moves. The assumption that the included fixed effects appropriately control for the complexity of all other moves is most plausible in the latter set of specifications. Regardless of which sample we consider and irrespective of whether we exploit variation in depth or width conditional on depth, we find that individual W-moves are significantly less likely to be chosen as object complexity increases. By contrast, the choice probabilities

Table 4: Choice Frequencies as a Function of Object Complexity

Panel A: Based on Depth

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Depth (÷ 100)	–0.775	0.014	–0.227	0.034
	(0.002)	(0.001)	(0.003)	(0.002)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves				
× Depth of Other Moves	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	16.713	0.472	20.480	0.657
R^2	0.494	0.232	0.663	0.232
N	3,457,878,398	296,522,573	398,856,135	111,905,262

Panel B: Based on Width

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Width (÷ 100)	–0.335	0.008	–0.744	0.031
	(0.002)	(0.001)	(0.003)	(0.002)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves				
× Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	16.713	0.472	20.480	0.657
R^2	0.555	0.284	0.693	0.284
N	3,457,878,398	296,522,573	398,856,135	111,905,262

Notes: Entries are coefficients and standard errors from estimating β in variants of eq. (2) by ordinary least squares. The regressions in the upper panel use moves' depth as a proxy for their inherent complexity, while those in the lower panel rely on width. All estimates control for player fixed effects. The regressions in the upper panel additionally include fixed effect for the combination of the number of W-, D-, and L-moves and the vector of depths of all other W- and L-moves in the same choice set. The regressions in the lower panel interact that fixed effect with the respective move's depth. The unit of observation in each regression is an available W- or L-move. Observations are reweighted so that all moves of the same type in a particular decision problem and all players receive equal weight. The sample in the first two columns in both panels includes all board configurations in our data, whereas the last columns restrict attention to configurations for which the associated choice sets do not contain D-moves. All estimates are scaled to correspond to the percentage-point change in choice probability associated with a one-unit increase in the respective regressor. Standard errors are two-way clustered by player and game, and are shown in parentheses.

of L-moves increase in their complexity. The results in Table 4 are thus consistent with Prediction 1.

Moreover, the point estimates are economically large. According to the coefficients in the first column of each panel, a one standard deviation increase in the depth of a W-move is associated with a decline in the same move's choice probability of about 13.9 p.p.; and a standard deviation increase in width is associated with a 3.9 p.p. decrease in the probability that the respective move is chosen. Our findings therefore suggest that object complexity is an empirically important determinant of choice.

6.2. *Tests of Prediction 2*

To pit satisficing against maximization we restrict attention to W-moves and modify the regression specification in eq. (2) by replacing the left-hand-side variable with an indicator for whether the player picked a W-move other than a. In symbols:

$$\text{(3)} \qquad \textit{Choose Other W-Move}_a = \gamma\, \textit{Complexity}_a + \chi_p + \phi_{A\setminus a} + \eta_a.$$

Table 5 presents results from estimating this model on different subsets of our data. The results in cols. (1A) and (1B) show that, in the full sample, the inherent complexity of one W-move is positively correlated with the choice frequency of other W-moves in the same set. The positive point estimates in these columns are inconsistent with maximization from consideration sets.

The next two columns demonstrate that the coefficients' sign remains unchanged when we only consider choice sets that do not contain any D-moves, or when we exclude the simplest W-move from each set. The remaining three columns restrict attention to settings that meet one of the following criteria: (i) long time controls (so that each player has, in expectation, at least twenty-five minutes for deliberation per game); (ii) small choice sets (with ten or fewer moves); (iii) none of the available W-moves are easily recognizable as good (because minimal depth exceeds fifty). Although these are settings in which maximization might be a priori especially appealing, the evidence continues to point to satisficing.

This finding raises the question of how widespread departures from maximization are. Are we rejecting the null of maximization because a few or because most of the DMs in our data are satisficing? To speak to this question, we test Prediction 2 at the individual level. Since the theory requires us to hold fixed the type and complexity of all other available moves, our individual-level tests focus on players for whom our sample contains at least one thousand decisions. There are 61,337 such individuals.

Table 5: Complexity and Choice Frequencies of Other W-Moves

Panel A: Based on Depth

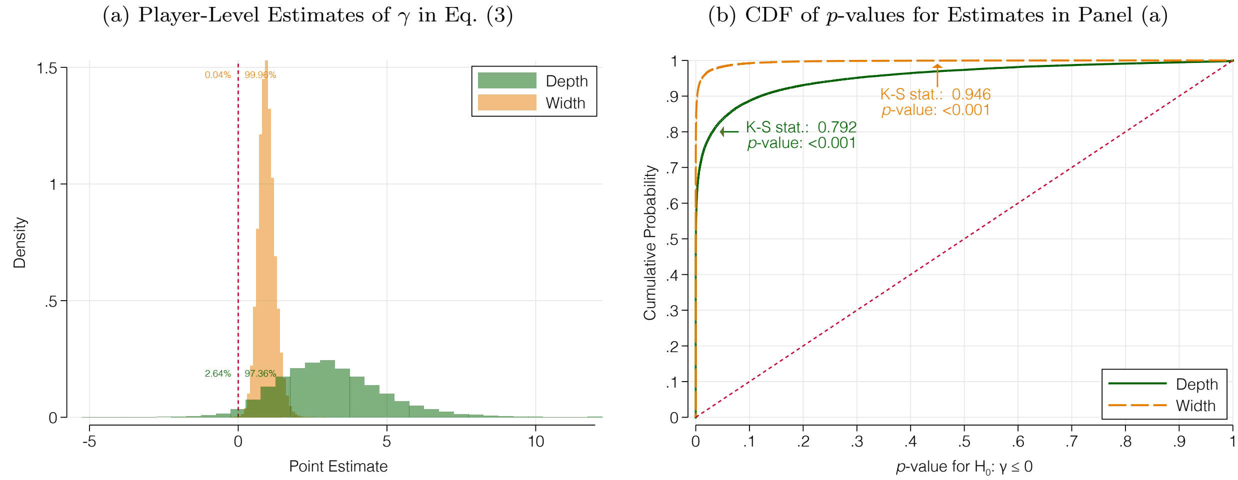
	Probability of Choosing Other W-Move					
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)
Depth ($\div$ 100)	1.497 (0.003)	1.524 (0.010)	0.324 (0.002)	0.911 (0.007)	1.635 (0.004)	1.782 (0.021)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- $\times$ D- $\times$ L-Moves $\times$ Depth of Other Moves	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	85.651	88.959	92.346	70.629	67.843	86.036
R^2	0.407	0.448	0.319	0.509	0.331	0.454
N	3,435,257,516	395,100,888	2,986,665,574	89,336,853	148,599,747	92,878,586

Panel B: Based on Width

	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width ($\div$ 100)	0.543 (0.001)	0.805 (0.003)	0.407 (0.002)	0.098 (0.020)	0.163 (0.003)	0.674 (0.011)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- $\times$ D- $\times$ L-Moves $\times$ Depth of Other Moves $\times$ Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	85.651	88.959	90.700	59.955	67.843	86.036
R^2	0.468	0.485	0.442	0.521	0.460	0.507
N	3,435,257,516	395,100,888	2,822,067,449	13,401,877	148,599,747	92,878,586

Notes: Entries are coefficients and standard errors from estimating γ in variants of eq. (3) by ordinary least squares. The regressions in the upper panel use moves' depth as a proxy for their inherent complexity, while those in the lower panel rely on width. All estimates control for player fixed effects. The regressions in the upper panel additionally include fixed effect for the combination of the number of W-, D-, and L-moves and the vector of depths of all other W- and L-moves in the same choice set. The regressions in the lower panel interact that fixed effect with the respective move's depth. The unit of observation in each regression is an available W-move. Observations are reweighted so that all moves in a particular decision problem and all players receive equal weight. The sample varies across columns but always exlcudes board configurations that only admit one W-move. The first column in each panel imposes no further restrictions. The second column excludes boards that admit D-alternatives. The third column excludes the simplest W-move in each board configuration, as measured by moves' depth (upper panel) or width (lower panel). The fourth column restricts attention to board configurations in which all W-moves have a depth of at least 50 (upper panel) and a width of at least 15 (lower panel). The fifth column focuses on moves from choice sets with no more than 10 alternatives. The last column excludes all observations from games that were played subject to faster-than-classical time controls. All estimates are scaled to correspond to the percentage-point change in choice probability associated with a one-unit increase in the respective regressor. Standard errors are two-way clustered by player and game, and are shown in parentheses.

Figure 8: Testing Prediction 2 at the Player Level



Notes: Figure presents player-level tests of Prediction 2. Panel (a) plots histograms of player-level estimates of γ in eq. (3), restricting attention to the 61,337 players for whom we observe at least 1,000 decisions in our sample. Estimates are scaled to be directly comparable to their counterparts in the first column of Table 5. Panel (b) shows the empirical CDF of the one-sided p-values associated with the point estimates in panel (a) (i.e., $H_0: \gamma \leq 0$). It also shows results from a Kolmogorov-Smirnov test against the null hypothesis of a uniform distribution of p-values. All p-values account for clustering across moves in the same game.

For every one of these players, we estimate the regression model in eq. (3). We then plot the distribution of the resulting coefficients in the left panel of Figure 8. Irrespective of whether we rely on depth or width conditional on depth to measure complexity, we obtain positive point estimates for the vast majority of DMs.

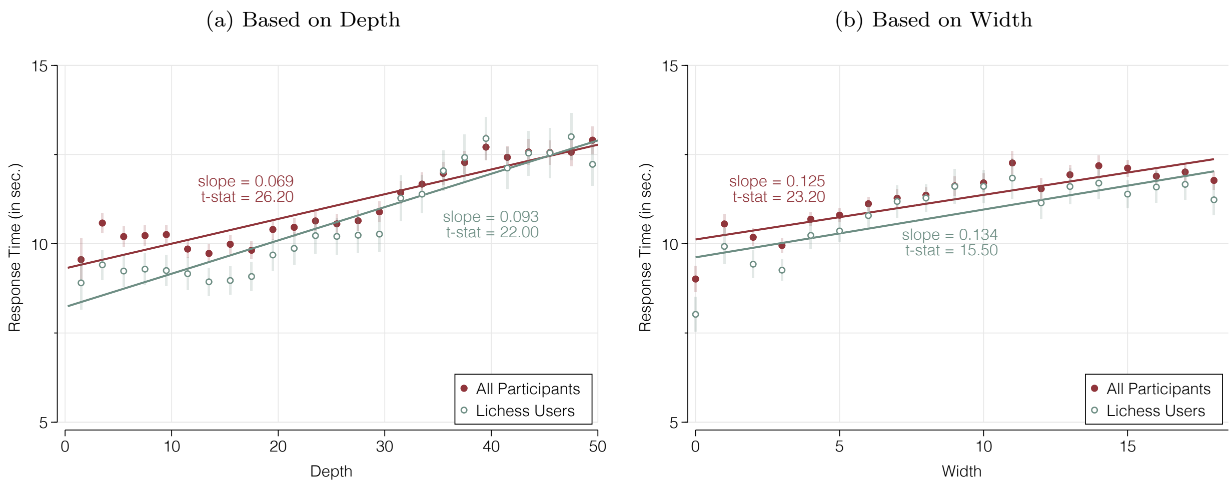
The right panel of Figure 8 shows the empirical CDFs of the one-sided p-values for our individual-level estimates. The relevant p-values are one-sided because the null hypothesis of maximization from consideration sets implies that $\gamma \leq 0$ (cf. Prediction 2). Under this null, the distribution of p-values should first-order stochastically dominate the uniform distribution.[25] This, however, is not what we observe. For either complexity measure the actual distribution of p-values is itself first-order stochastically dominated by the uniform distribution, and a formal Kolmogorov-Smirnov test rejects the limit case of uniformity at the 99%-confidence level. Even if we relied on two-sided p-values, we would reject, at the 5%-significance level, the null hypothesis of maximization from consideration sets for more than 80% of players.

7. Response Times

Going beyond the theory, we next explore response times. We first analyze how complexity affects the speed with which DMs evaluate individual alternatives. We then examine how

[25] That is, we should have $\Pr(p \leq \alpha \mid H_0) \leq \alpha$ for all $\alpha \in [0, 1]$. This claim follows from Definition 8.3.26 and Theorem 8.3.27 in Casella and Berger (2001). The intuition behind it is as follows. If $\gamma = 0$, then the observed p-values should be uniformly distributed over the unit interval. If $\gamma < 0$, however, then we would expect to see fewer small (one-sided) p-values and more large ones, implying first-order stochastic dominance.

Figure 9: Object Complexity Predicts Response Times



Notes: Figure shows binscatter plots of the raw relationship between response times in the experimental task (y-axis) and the respective moves' complexity (x-axis). Panel (a) uses a move's depth to measure complexity, whereas panel (b) relies on width. The underlying data come from the experiment described in Section 4. When focusing on Lichess users only, observations are reweighted to approximate the distribution of strength ratings in the observational Lichess data that is used in the previous section. Error bars correspond to 95%-confidence intervals, accounting for two-way clustering by participant and move.

long DMs deliberate before choosing from a set of alternatives.

Our model assumes that complexity makes it more difficult to assess the value of any given alternative, which in turn results in noisier evaluations. Another way in which this difficulty might manifest itself are longer response times. When an object is more complex, it likely takes more time evaluate it. Since we record response times in the experiment, we can test whether this intuition holds in the data.[26]

Figure 9 demonstrates that response times do, indeed, increase in the complexity of moves. We observe a clear positive relationship for both of our complexity measures, and when we limit the sample to Lichess users. Appendix Table AT.10 shows that response times increase in complexity even when we control for the exact board configuration as well as the piece that executes the move. Based on this evidence, we conclude that more complex alternatives take more time to evaluate.

The sequential nature of satisficing suggests an additional empirical regularity. Consider replacing one of the alternatives in the DM's choice set with another one of equal complexity but higher value. Unless the evaluation order favors lower- over higher-value alternatives, such a switch should reduce the time it takes to choose a move. The reason is that the object with the higher value is more likely to have a satisfactory score, which means that, on average, the DM stops and makes a choice earlier.

[26] We did not preregister any analyses with response times as outcome.

To test this prediction, we return to the Lichess data and compare response times for choice sets in which the simplest move is of type L with response times for sets in which the simplest move is (nearly) equally complex but of type W. Restricting attention to decision problems in which the simplest available move is of either type and finely partitioning the data according to the complexity of the simplest alternative, we implement these comparisons by estimating the following regression model separately for each partition:

$$(4) \qquad \mathit{Response\,Time}_{A,p} = \lambda\,\mathbf{1}\big[\mathit{Type}(\lfloor A \rfloor) = W\big] + \tau\,\mathit{Time\,Pressure}_{A,p} + \chi_p + \phi_{A\setminus\lfloor A\rfloor} + \upsilon_{A,p}.$$

Here, $\mathit{Response\,Time}_{A,p}$ denotes how long player p deliberated before picking a move from choice set A, $\mathit{Type}(\lfloor A \rfloor)$ stands for the type of the simplest move in A, $\mathbf{1}[\cdot]$ represents the indicator function, and χ_p is a player fixed effect.[27] $\mathit{Time\,Pressure}_{A,p}$ controls for how fast p must, on average, execute every move if she is to achieve checkmate before her clock expires (assuming she takes the shortest path to mate and her opponent holds out as long as possible). To approximate the thought experiment of replacing the simplest move holding everything else fixed, we control for $\phi_{A\setminus\lfloor A\rfloor}$, a fixed effect for the type composition and complexity of all other moves in A.

The parameter of interest in eq. (4) is λ. Since we partition the data according the complexity of the simplest available move, a given λ measures the expected change in response time from replacing the simplest L-move with a W-move of approximately the same complexity. Note, W- and L-moves never have exactly the same depth. By definition, the depth to mate of a W-move is an odd number while that of an L-moves is always even. For this reason, we construct partitions so that we only compare choice sets in which the simplest move is an L-move with depth d with sets in which the simplest move is a W-move with depth $d + 1$. Since more complex moves take, on average, longer to evaluate, the resulting estimates of λ will be upward biased relative to the counterfactual in which only the type of the simplest move changes. For width such an issue does not arise. Thus, when using width to measure complexity, each partition of the data includes only choice sets in which the simplest available moves have exactly the same width.

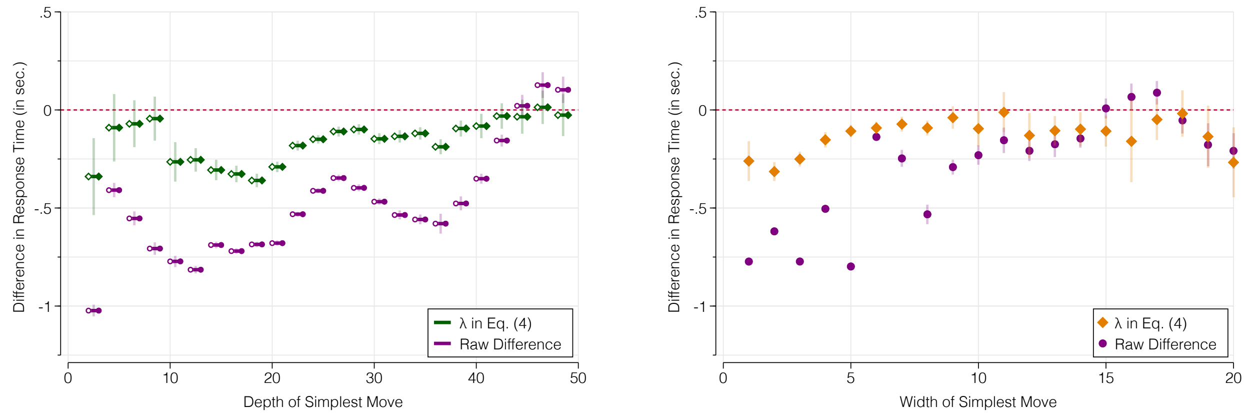
Figure 10 presents estimates of λ according to the complexity of the simplest move. For comparison, the figure also shows raw differences in response times between choice sets in which the simplest move is an L-move and those in which it is of type W. Although these differences narrow upon carefully controlling for the characteristics of the other moves in the set, and despite the upward bias in our estimates based on depth, all but one of the regression estimates in the left panel of Figure 10 are negative, and most are statistically

[27] We exclude choice sets in which there is more than one simplest move whenever these moves are of different types.

Figure 10: Response Times Depend on the Type of the Simplest Available Move

(a) Estimates of λ in Eq. (4), Based on Depth

(b) Estimates of λ in Eq. (4), Based on Width



Notes: Panel (a) presents point estimates and 95%-confidence intervals for λ in eq. (4) based on depth as a measure of complexity. As explained in the text, the relevant comparisons are across choice sets in which the simplest move is an L-move with depth d and sets in which the simplest move is a W-move with depth $d+1$. The panel also shows raw differences in response times between these sets. Panel (b) presents point estimates and 95%-confidence intervals for λ in eq. (4) based on width, as well as raw differences in response times between choice sets in which the simplest moves are L- and W-moves with exactly the same width. In either panel, negative values indicate lower response times for sets in which the simplest move is of type W. Error bars account for two-way clustering by player and game.

significant at the 5%-level. In the panel on the right, all twenty regression estimates are negative and fourteen are statistically significant. The evidence, therefore, suggests that replacing an L-move with an equally complex W-move would shorten response times.

We further observe that switches involving simpler W- and L-moves are, on average, associated with larger reductions in response times than switches of more complex alternatives. This comparative static is predicted by satisficing. To see this, note that DMs are more likely to stop searching for an acceptable alternative when they encounter a simple rather than a complex W-move, but are less likely to stop when they encounter an L-move of similar complexity. Taken together, the results in Figure 10 imply that response times shorten as the number of "good" alternatives in the set increases, especially when they are simple to evaluate. Both observations are consistent with satisficing behavior.

8. Concluding Remarks

The starting point of this paper is the observation that many objects that are of interest to economists are inherently complex and therefore difficult to evaluate. Our first contribution is to propose a tractable model of object complexity, and to experimentally validate its key assumption, i.e., that higher inherent complexity leads to evaluation errors.

Our second contribution is to examine how DMs cope when choosing from a set of complex alternatives. We consider two leading mechanisms. According to the first one, DMs evaluate objects sequentially, stopping as soon as a satisfactory alternative is found. According to the second mechanism, DMs identify a (possibly small) subset of alternatives, evaluate all objects in this consideration set, and choose the one with the highest perceived value. We develop a new empirical test that relies on variation in object complexity to distinguish between these mechanisms. Our data on endgame moves in chess are consistent with the former mechanism but not with the latter one.

While chess provides an almost ideal setting to study decision making outside of the laboratory, object complexity is likely important in many other economic environments. Contracts, for example, may span tens of pages detailing different contingencies. Financial products can have dozens of attributes that are relevant to consumers. Doctors may need to diagnose patients with long medical histories and multiple symptoms. Firms often choose between suppliers or job candidates that differ along more than just one dimension. Going beyond specific examples, we suspect that object complexity affects decision making whenever the available alternatives are composed of many payoff-relevant "ingredients" that need to be "integrated" in order to assess value.

Do economic agents commonly resort to satisficing when choosing among complex objects? Given that our theory and empirical test are not specific to chess, it is straightforward to analyze this question in any setting that satisfies the following conditions. (i) Choices and choice sets are observable. (ii) The available alternatives can be ordered according to their value to the economic agent. (iii) It is possible to measure, or at least approximate, objects' inherent complexity. Additional evidence from a variety of settings would not only help to better understand the prevalence of satisficing behavior but it might also shed light on its boundary conditions.

Another direction for future research is to examine the implications of complexity and satisficing for firm behavior and competition. For example, when consumers' choice sets include products by different firms, then firms with superior offerings might wish to reduce the inherent complexity of their products in order to make it easier for consumers to evaluate them. Firms with inferior offerings, however, have an incentive to complicate, say, product descriptions.[28] In addition, if consumers are satisficing, then firms might seek to influence the order in which their products are being considered—even if doing so is costly. Similar considerations are likely relevant in the design and presentation of policy proposals, and in various other settings.

References

Abreu, D. and A. Rubinstein (1988). The Structure of Nash Equilibrium in Repeated Games with Finite Automata. *Econometrica 56*(6), 1259–1281.

[28]Consistent with this idea, Célérier and Vallée (2017) document that the complexity of structured financial products predicts markups relative to their actuarially fair value.

Bossaerts, P. and C. Murawski (2017). Computational Complexity and Human Decision-Making. *Trends in Cognitive Sciences 21*(12), 917–929.

Camerer, C. (2003). *Behavioral Game Theory.* Princeton, NJ: Princeton University Press.

Caplin, A., M. Dean, and D. Martin (2011). Search and Satisficing. *American Economic Review 101*(7), 2899–2922.

Casella, G. and R. Berger (2001). *Statistical Inference* (2nd ed.). Boston: Cenage.

Célérier, C. and B. Vallée (2017). Catering to Investors through Security Design: Headline Rate and Complexity. *Quarterly Journal of Economics 132*(3), 1469–1508.

Chen, D., M. Schonger, and C. Wickens (2016). oTree - An Open-Source Platform for Laboratory, Online and Field Experiments. *Journal of Behavioral and Experimental Finance 9*, 88–97.

Chiappori, P.-A., S. Levitt, and T. Groseclose (2002). Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer. *American Economic Review 92*(4), 1138–1151.

Enke, B., T. Graeber, and R. Oprea (2023). Complexity and Time. NBER Working Paper No. 31047.

Gabaix, X., D. Laibson, G. Moloche, and S. Weinberg (2006). Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model. *American Economic Review 96*(4), 1043–1068.

Hsu, S.-H., C.-Y. Huang, and C.-T. Tang (2007). Minimax Play at Wimbledon: Comment. *American Economic Review 97*(1), 517–523.

Huck, S. and G. Weizsäcker (1999). Risk, Complexity, and Deviations from Expected-Value Maximization: Results of a Lottery Choice Experiment. *Journal of Economic Psychology 20*(6), 699–715.

Jakobsen, A. (2020). A Model of Complex Contracts. *American Economic Review 110*(5), 1243–1273.

Kalai, E. and W. Stanford (1988). Finite Rationality and Interpersonal Complexity in Repeated Games. *Econometrica 56*(2), 397–410.

Levitt, S., J. List, and D. Reiley (2010). What Happens in the Field Stays in the Field: Exploring Whether Professionals Play Minimax in Laboratory Experiments. *Econometrica 78*(4), 1413–1434.

Levitt, S., J. List, and S. Sadoff (2011). Checkmate: Exploring Backward Induction among Chess Players. *American Economic Review 101*(2), 975–990.

Luce, D. (1959). *Individual Choice Behavior.* New York: Wiley.

Man, R. d. (2013). Syzygy tablebases. available at https://github.com/syzygy1/tb.

Manzini, P. and M. Mariotti (2014). Stochastic Choice and Consideration Sets. *Econometrica 82*(3), 1153–1176.

Marschak, J. (1960). Binary-Choice Constraints and Random Utility Indicators. In K. Arrow, S. Karlin, and P. Suppes (Eds.), *Stanford Symposium on Mathematical Methods in the Social Sciences*, pp. 312–329. Stanford, CA: Stanford University Press.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zambreka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

Nalimov, E., G. Haworth, and E. Heinz (2000). Space-Efficient Indexing of Chess Endgame Tables. *ICGA Journal 23*(3), 148–162.

Neyman, A. (1985). Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoners' Dilemma. *Economics Letters 19*(3), 227–229.

Oprea, R. (2020). What Makes a Rule Complex? *American Economic Review 110*(12), 3913–3951.

Oprea, R. (2022). Simplicity Equivalents. mimeographed, UC Santa Barbara.

Palacios-Huerta, I. (2003). Professionals Play Minimax. *Review of Economic Studies 70*(2), 395–415.

Palacios-Huerta, I. and O. Volij (2008). Experientia Docet: Professionals Play Minimax in Laboratory Experiments. *Econometrica 76*(1), 71–115.

Palacios-Huerta, I. and O. Volij (2009). Field Centipedes. *American Economic Review 99*(4), 1619–1635.

Rubinstein, A. (1986). Finite Automata Play the Repeated Prisoner's Dilemma. *Journal of Economic Theory 39*(1), 83–96.

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times. *Economic Journal 117*(523), 1243–1259.

Rubinstein, A. (2016). A Typology of Players: Between Instinctive and Contemplative. *Quarterly Journal of Economics 131*(2), 859–890.

Salant, Y. (2011). Procedural Analysis of Choice Rules with Applications to Bounded Rationality. *American Economic Review 101*(2), 724–748.

Simon, H. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics 69*(1), 99–118.

Simon, H. (1972). Theories of Bounded Rationality. In C. McGuire and R. Radner (Eds.), *Decision and Organization*, pp. 161–176. Amsterdam: North-Holland.

Spenkuch, J., P. Montagnes, and D. Magleby (2018). Backward Induction in the Wild? Evidence from Sequential Voting in the US Senate. *American Economic Review 108*(7), 1971–2013.

Thompson, K. (1986). Retrograde Analysis of Certain Endgames. *ICGA Journal 9*(3), 131–139.

Walker, M. and J. Wooders (2001). Minimax Play at Wimbledon. *American Economic Review 91*(5), 1521–1538.

Wilson, A. (2014). Bounded Memory and Biases in Information Processing. *Econometrica 82*(6), 2257–2294.

Wooders, J. (2010). Does Experience Teach? Professionals and Minimax Play in the Lab. *Econometrica 78*(3), 1143–1154.

Zermelo, E. (1913). Über eine Anwendung der Mengenlehre auf die Theorie des Schachspiels. In E. Hobson and A. Love (Eds.), *Proceedings of the Fifth International Congress of Mathematicians (Cambridge 1912)*, pp. 501–504. Cambridge, UK: Cambridge University Press.

Online Appendix

Contents

List of Tables

Appendix A: Proofs

Proof of Proposition 1. Fix two alternatives a and b and two choice sets A and B as in the statement of Proposition 1.

Suppose $\mu(v) > T$. Fix two orders of evaluation, O_A for A and O_B for B, that are identical except for b substituting a in O_B. By value invariance, the probabilities of these two orders are identical. If a and b appear last in O_A and O_B respectively, then the probability of reaching them is identical because O_A and O_B are identical prior to reaching the last alternative. In this case, the two alternatives' choice probabilities conditional on reaching them are also identical because they are equal to 1.

Otherwise, the choice probability of a in O_A is

$$Pr(\text{DM did not stop prior to } a) \times Pr\,(y \geq T)$$

where y is some score realization. The first component in this expression is identical to the probability of not stopping prior to reaching b in O_B because O_B is identical to O_A prior to reaching b. The second component in this expression is equal to $1 - F_a(T)$. This component is larger than $1 - F_b(T)$. This is because we have that (i) $\mu(v) > T$ and (ii) $F_b(T) > 0$ implying that $F_b(T) > F_a(T)$ by Condition 1. Thus, the choice probability of a in O_A is larger than the choice probability of b in O_B. Consequently, the choice probability of any alternative appearing after a in O_A is smaller than the choice probability of the same alternative in O_B. Since this holds for any order other than orders in which a and b appear last and all orders are drawn with positive probability, the result follows.

The proof for $\mu(v) < T$ is analogous. *Q.E.D.*

Proof of Proposition 2. Fix three alternatives a, b, and c and two choice sets A and B as in the statement of Proposition 2.

The choice probability of c in A is the sum of expressions of the form $P_A(S)P(c, S)$ where $P_A(S)$ denotes the probability of drawing consideration set S and $P(c, S)$ is the probability that c is the highest-order statistic in S.

Consider the mapping $M(S) = S - a + b$. This mapping is from the power set of A to the power set of B. It is one-to-one and onto. By order invariance, $P_A(S) = P_B(M(S))$ for every $S \subseteq A$. It therefore suffices to show that

$$P(c, S) - P(c, M(S)) \geq 0 \qquad (5)$$

for every S in order to establish the required inequality of choice probabilities. If $a \notin S$, then $M(S) = S$ and inequality (5) holds.

Suppose $a \in S$. Because the PDFs and CDFs corresponding to alternatives in S and $M(S)$ are identical other than for the alternatives $a \in A$ and $b \in B$, we can write the difference above as:

$$\begin{aligned} P(c,S) - P(c,M(S)) &= \int_{-\infty}^{\infty} \left(\prod_{l \notin \{a,b,c\}} F_l(y) \right) f_c(y) \left(F_a(y) - F_b(y) \right) dy \\ &=_{(i)} \int_{-\infty}^{\infty} P(y)G(y)dy =_{(ii)} \int_{-\infty}^{\infty} P(\mu(v)+\epsilon)G(\mu(v)+\epsilon)d\epsilon \\ &=_{(iii)} \int_{0}^{\infty} \left(P(\mu(v)+\epsilon) - P(\mu(v)-\epsilon) \right) G(\mu(v)+\epsilon)d\epsilon. \end{aligned}$$

Here, equality (i) follows from denoting $P(y) = \left(\prod_{l \notin \{a,b,c\}} F_l(y) \right) f_c(y)$ and $G(y) = F_a(y) - F_b(y)$, equality (ii) follows from substituting y with $\mu(v) + \epsilon$, and equality (iii) follows from the fact that by symmetry:

$$G(\mu(v)+\epsilon) = 1 - F_a(\mu(v)-\epsilon) - (1 - F_b(\mu(v)-\epsilon)) = F_b(\mu(v)-\epsilon) - F_a(\mu(v)-\epsilon) = -G(\mu(v)-\epsilon).$$

Thus, to establish inequality (5), it suffices to show that the integrand in (iii) is non-negative.

Because b is more complex than a, we have that $G(\mu(v)+\epsilon) \geq 0$ by Condition 1. The expression $P(\mu(v)+\epsilon) - P(\mu(v)-\epsilon)$ is also non-negative because (i) $f_c(\mu(v)+\epsilon) \geq f_c(\mu(v)-\epsilon)$ since f_c is symmetric around its mean and increases up to its mean which is weakly larger than the $\mu(v)$, and (ii) CDFs are weakly increasing functions. The first part of Proposition 2 follows.

For the second part of the proposition, fix a consideration set $S \subseteq A$ of size ≥ 3 that includes a and c and that is drawn with positive probability. Suppose that the supports of f_a and f_c are not finite. By symmetry and unimodality, each support is the real line. By Condition 1, the support of f_b is also the real line, and hence $G(\mu(v) + \epsilon) > 0$. To complete the proof, it thus suffices to show that $P(\mu(v)+\epsilon) - P(\mu(v)-\epsilon) > 0$ on a non-empty interval I of ϵ's.

Let $d \notin \{a, c\}$ be some alternative in S. By unimodality and symmetry, the support of f_d is either an interval or the real line. In either case, F_d increases in some interval $(d_{\min}, d_{\max})$. Let

$$I = \begin{cases} (d_{\min} - \mu(v), d_{\max} - \mu(v)) & \text{if } d_{\min} > \mu(v) \\ (0, \min\{d_{\max} - \mu(v), \mu(v) - d_{\min}\}) & \text{if } d_{\min} \leq \mu(v) < d_{\max} \\ (\mu(v) - d_{\max}, \mu(v) - d_{\min}) & \text{otherwise.} \end{cases}$$

Then, for every $\epsilon \in I$, $F_d(\mu(v)+\epsilon) > F_d(\mu(v)-\epsilon)$ implying that the product term in $P(\mu(v)+\epsilon)$ is larger than in $P(\mu(v)-\epsilon)$ and hence that $P(\mu(v)+\epsilon) - P(\mu(v)-\epsilon) > 0$.

Q.E.D.

Appendix B: Data Appendix

Our observational data on endgame moves come from `lichess.org`. Every month, Lichess releases database extracts covering all rated chess games between two human players that were hosted on its platform during the previous month. These extracts are made available in the human-readable PGN format at `https://database.lichess.org`, and include basic facts about each game (including players' usernames and ratings, date and time of the game, time controls, ultimate outcome, etc.), the exact sequence of moves, as well as, starting April 2017, the clock reading at the end of each move.

We downloaded and processed all extracts through August 2020, filtering on endgame positions with six or fewer pieces. We then spent about 600,000 CPU-hours querying the Nalimov and Syzygy endgame tablebases for information on depth to mate (DTM) and the type of each available legal move (i.e., W, D, or L) in these positions. The 6-men Syzygy and Nalimov endgame databases are available at `http://tablebase.sesse.net` (Syzygy: 150GB; Nalimov: 1.2TB). Because Syzygy tablebases take into account the 50-move rule, we rely on them to determine the type of each move, whereas information on DTM comes from Nalimov's database. The only board configurations with six or fewer pieces that are not covered in the latter are *(i)* ones in which a lone king faces five other pieces, and *(ii)* positions with castling rights. The former are generally uninteresting because 98.8% of available legal moves are of type W, and the latter are extremely rare in the Lichess data ($<$ .01% of moves in nontrivial endgame positions).

The sample for our main analysis restricts attention to decision problems in *(i)* board positions with six or fewer pieces with *(ii)* available information on the types of all available legal moves and the DTM of all available W- and L-moves, in which *(iii)* there are one or more legal W-moves and at least one D- or L-alternative, *(iv)* excluding the first 1,000 such decision problems for every user.

Appendix C: Robustness Checks

C.1. *Restricting Attention to Board Positions with High Minimal DTM*

In Appendix Tables AT.1–AT.2, we replicate our main results, restricting attention to board positions in which the minimal DTM among W-moves exceeds fifty. These are positions in which it is *a priori* unlikely that players can accurately discriminate between moves according to their depth. Reassuringly, the results from this smaller sample are qualitatively equivalent to those in the main text.

C.2. *Controlling for Time Pressure*

Since the timing of decisions is endogenous, we do not control for it in our main analysis. We do, however, obtain qualitatively equivalent findings when we account for it. To show this, we replicate Tables 4 and 5 in the main text, controlling for time pressure. Specifically, we control for the number of seconds per move the player has left if she takes the shortest path to mate, while her opponent holds out as long as possible. As the results in Appendix Tables AT.3–AT.4 illustrate, our findings remain qualitatively unchanged.

C.3. *Restricting Attention to First Move in Series*

One potential concern with the results in the main text is serial-dependence in the decisions of players. A player who sees a winning strategy and follows it in each subsequent move enters our data set multiple times. To rule out that this issue is driving our main results, we replicate Tables 4 and 5 in the text restricting attention only to the first move in a series of moves from winning positions in a given game. If a player sees and executes a winning strategy, then she would thus only enter our data set once per game.[1] As the results in Appendix Tables AT.5–AT.6 illustrate, our findings remain qualitatively unchanged.

C.4. *Replication with Data from The Week in Chess*

Appendix Tables AT.7–AT.8 replicate Tables 4 and 5 in the main text, using an independent dataset that we obtained from *The Week in Chess* (TWIC). TWIC is a free, weekly publication that "rounds up the most important chess" games from the previous week (see `https://theweekinchess.com`). Most of these games are played between elite players in national and international tournaments, or chess leagues.

Our data include all games covered in TWIC between September 1994 and May 2020. In total, we observe 536,674 decision problems in endgame positions with six or fewer pieces, one or more legal W- and at least one D- or L-move. The choice sets in these decision problems contain 9,067,040 legal moves.

Besides being several orders of magnitude smaller, the most important difference between the TWIC and Lichess data is that the former admit much less variation in players' skill. Chess players in high-profile tournaments tend to be better than the average experienced player on Lichess. This fact is reflected in a *much* lower frequency of mistakes in the TWIC data. Since

[1] A player can enter more than once per game if she and her opponent both make mistakes, in which case strategies would need to be recomputed.

tournament-level players almost never choose L-moves in winning positions, our estimates of the effect of object complexity on the choice frequency of L-moves are economically and statistically indistinguishable from zero. Nonetheless, out of the 20 estimates in Appendix Tables AT.7–AT.8, 16 are statistically highly significant and have the same sign as their counterparts in the main text. The remaining 4 estimates can only be imprecisely estimated, so that their 95%-confidence intervals include both positive and negative values.

Appendix D: Experimental Instructions and Further Details

As explained in the main text, the experiment took place over the four-week period starting April 7, 2023. We recruited participants via targeted ads on Facebook, Twitter, and Reddit, as well as through forum posts on `lichess.org` and `chess.com`. All ads and forum posts contained a link that directed participants to a website that we had custom-built for the experiment using oTree (Chen et al. 2016).

After consenting to participate in the experiment, we required all subjects to provide their Lichess and Chess.com usernames, which the website verified in real time by querying the APIs of the the respective platforms.[2] Since we wanted to recruit only online-chess players, providing a valid username to at least of one these platforms was a precondition for participation. Out of the 3,966 subjects that met this condition, 584 provided a Lichess username, 2,471 submitted a Chess.com username, and 911 subjects provided both.

The experiment consisted of five stages: 1. Consent; 2. Username Verification; 3. Instructions; 4. Experimental Task (25 rounds); 5. Background Questionnaire (8 questions).

The actual experimental task consisted of twenty-five rounds. In each round, participants were shown a chess board with a randomly sampled endgame position in which one legal move was highlighted (cf. Figure 3 in the main text). They were then asked to indicate whether the highlighted move is a winning, drawing, or losing move. These types had been carefully defined in the instructions; though this may not have been strictly necessary, given that about 78% of subjects indicated that they had already known about winning, drawing, and losing moves before participating the experiment. The instructions had also explicitly stated that moves of each type were a priori equally likely to be shown.

The population of moves that could in principle be shown to subjects in order to be evaluated had been extracted from a random subset of all legal moves in 4,196 representative endgame positions from our observational Lichess data. In total, we extracted 30,000 randomly chosen moves subject to their depth and width not exceeding 50 and 18, which corresponds to about the 95th percentiles of the respective marginal distributions. We then constructed

[2]These APIs queries verified the existence of the usernames and retrieved basic information about users' activity on the platform, including their strength ratings.

sampling weight to achieve that W-, D-, and L-moves would be shown to participants with approximately the same probability, subject to the depth of the W- and L-moves that were shown being approximately uniformly distributed between zero and fifty.

Subjects had between five and forty-five seconds to submit their evaluation. The time limit was randomized and distributed uniformly and i.i.d. across rounds.

Subjects earned one virtual lottery ticket for every move they correctly evaluated. After the experiment, all lottery tickets were entered into a raffle for twenty $100 Amazon gift certificates.

The median participant earned 15 tickets and spent slightly less than 9 minutes on the experiment. About 22% of participants did not finish the experiment. That is, they did not submit evaluations for all 25 moves or the evaluations that were submitted did not pass basic attention checks.[3,4]

Appendix Table AT.9 presents descriptive statistics for our experimental data. On the next page, we reproduce the text that was shown to participants during the experiment, with horizontal lines demarcating screens.

[3] An evaluation fails our attention checks if (i) the subject submits her answer less than two seconds after being shown the board, or (ii) if she lets the time run out for this and all subsequent evaluation tasks.

[4] The numbers above do not include individuals that clicked on our ads but did not proceed past the consent screen.

Experimental Screens

Research Survey

Research Study: Understanding Strategic Reasoning under Time Pressure (STU00219176)
Principal Investigators: Dr. Yuval Salant; Dr. Jörg Spenkuch
Supported By: This research is funded by Northwestern University.

Welcome to our survey of chess players! The purpose of this study is to better understand how chess players reason under time pressure. We are very grateful for your help!

To take this survey you must be a registered user of either Lichess.org or Chess.com. Below we provide additional information on this study in order to help you decide whether you'd like to participate.

To begin our survey, you need to provide your consent by pressing the PROCEED button at the bottom of this page.

What should I expect?
Your participation is voluntary. If you choose to participate, you will first be asked to provide your username on Lichess.org and/or Chess.com. We will then ask you to rate several chess moves in endgame positions, followed by a handful of questions about your age, gender, and experience playing chess. We estimate that it will take about 10-15 minutes to complete the survey.

Will I be paid?
We will reward your participation in this survey with a chance to win one of twenty $100 gift certificates to Amazon.com. Everyone who successfully completes the survey becomes eligible to participate in the raffle for these gift certificates. Your chances of winning will depend on how many other users complete the survey and on how well you evaluate the endgame moves that we will show you. The winners of the gift certificates will be contacted via the messaging function on Lichess and Chess.com by May 31, 2023.

Are there any risks?
We foresee little risk from participating in this survey, and we do not guarantee that you will receive any benefits beyond a chance to win an Amazon gift certificate.

How will my information be used?
The information collected through this survey will be exclusively used for research purposes. All data will be handled and stored in accordance with Northwestern University policy. There is minimal risk that participants might be identified from the information provided. The research team will take extensive precautions to keep all data secure in order to protect confidentiality. As part of this effort, your actual identity will remain unknown to the researchers conducting this study. The results of this research may be published, but only in anonymized form.

Who can I talk to?

If you have questions, concerns, or complaints, you can contact the Principal Investigators at chessresearch@u.northwestern.edu. This research has been reviewed and approved by an Institutional Review Board (IRB) — an IRB is a committee that protects the rights of people who participate in research studies. You may contact the IRB by phone at +1 (312) 503-9338 or by email at irb@northwestern.edu if:

- Your questions, concerns, or complaints are not being answered by the research team.
- You cannot reach the research team.
- You want to talk to someone besides the research team.
- You have questions about your rights as a research participant.
- You want to get information or provide input about this research.

By proceeding to the next screen, you are consenting to participate in the survey.

Lichess / Chess.com Username

Are you a registered user of Lichess.org? If so, please enter your username. If not, leave the textbox below empty.

Are you a registered user of Chess.com? If so, please enter your username. If not, leave the textbox below empty.

Please do not enter your real name, but your username on Lichess.org and/or Chess.com (e.g., chessmaven19). Please enter both usernames if you play on both platforms. If you do not enter at least one valid userhandle, then we won't be able to contact you if you win one of the $100 gift certificates.

Evaluating Chess Moves: Instructions

We are interested in better understanding how chess players evaluate moves under time pressure. To this end, we will show you 25 legal chess moves in endgame positions. You are being asked to evaluate them.

When evaluating a move, you can choose between the following three possibilities:

- **Winning move** = If the current player makes this move, then the current player will win under subsequent perfect play.

- **Losing move** = If the current player makes this move, then the opponent will win under subsequent perfect play.
- **Drawing move** = If the current player makes this move, then perfect play by both players will result in a draw.

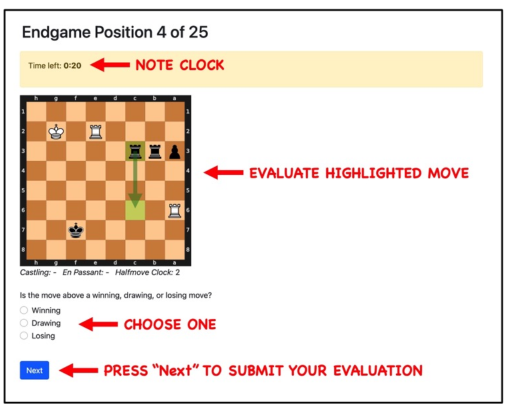
We will compare your evaluations to the respective moves' actual theoretical values (i.e., Winning, Losing, or Drawing), and you will earn one virtual lottery ticket for every move that you correctly evaluate. Pooling all lottery tickets earned by the participants in this survey, we will randomly draw 20 tickets and award $100 Amazon gift certificates to the respective owners. Thus, your chances of winning a gift certificate depend directly on how many evaluations you get right.

For every move you see, this website randomly determines how much time you have to submit your evaluation. For some evaluations, you might have as little as 5 seconds, whereas for others you may take up to 45 seconds.

To be clear, we are *not* asking you to evaluate whether a particular move is the best move in the given board position. We are asking you to determine whether the move is a Winning, Drawing, or Losing move, as defined above.

Please proceed to the next screen to see an example of what exactly you're being asked to do.

Evaluating Chess Moves: Example



The move that is highlighted in the screenshot above is theoretically a Losing move. If you were asked to evaluate this move, you would earn one lottery ticket if you chose the "Losing" option and pressed the "Next" button before the clock at the top of the screen ticks down to zero. You would not earn a lottery ticket if you chose either the "Winning" or "Drawing" options, or if you didn't submit your answer in time. Once the clock expires, you will automatically be moved to the next screen.

Note

The chess boards that you're about to see have been chosen from a large number of positions that were actually played in various online games.

We have picked these boards such that you can expect to see Winning, Drawing, and Losing moves in roughly equal proportions.

To see the first move, please press "Next".

Endgame Position 1 of 25

Time Left: ____

Castling: _ En Passant: _ Halfmove Clock: _

Is the move above a winning, drawing, or losing move?

- Winning
- Drawing
- Losing

⋮ ⋮ ⋮

⋮ ⋮ ⋮

Results

Round	Your Guess	Correct Answer
1	Winning	Winning
2	Drawing	Losing
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
25	Losing	Losing

Based on these results, you have earned __ **lottery tickets**.

Please tell us a little bit about yourself.

How old are you?

(drop-down list)

What is your gender?

- Male
- Female
- Other
- Prefer No to Say

Where do you live?

(drop-down list)

We will now ask you some questions about your experience playing chess.

Approximately how long have you been playing chess?

(drop-down list)

How often do you typically play over-the-board chess, i.e., in real life?

(drop-down list)

How often do you typically play online chess?

(drop-down list)

Thank you for your participation!

Thank you for contributing to our research. Before you go, we have just two more questions that will help us improve this survey.

Before taking our survey, did you already know about Winning, Drawing, and Losing moves in endgame positions?

- Yes
- No

Did you feel that the instructions you received about evaluating endgame moves were clear?

- Yes
- No

If you have any other comments, please enter them below. We would be very interested in hearing your feedback.

Goodbye

This completes our survey. Your answers have been recorded. Thank you for your help!

We'll contact you via Lichess.com and/or Chess.com if you end up winning an Amazon gift certificate.

Appendix Tables

Appendix Table AT.1: Replication of Table 4, Board Positions with High Minimal Depth to Mate

Panel A: Based on Depth

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Depth (÷ 100)	–0.282	0.017	–0.207	0.023
	(0.003)	(0.002)	(0.006)	(0.006)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	23.970	0.569	21.319	0.902
R^2	0.433	0.334	0.563	0.398
N	92,066,019	39,282,812	22,051,140	9,811,399

Panel B: Based on Width

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Width (÷ 100)	–0.419	0.003	–1.076	0.007
	(0.006)	(0.001)	(0.009)	(0.002)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	23.970	0.569	21.319	0.902
R^2	0.484	0.388	0.569	0.440
N	92,066,019	39,282,812	22,051,140	9,811,399

Notes: See Table 4 in the main text. The only difference between this table and that in the text is that results above restrict attention to board configurations in which the minimal depth among W-moves is at least 50.

Appendix Table AT.2: Replication of Table 5, Board Positions with High Minimal Depth to Mate

Panel A: Based on Depth

	Probability of Choosing Other W-Move					
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)
Depth (÷ 100)	0.911 (0.007)	1.328 (0.023)	0.259 (0.007)	0.911 (0.007)	0.946 (0.009)	1.075 (0.044)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	70.629	85.755	79.501	70.629	56.506	72.482
R^2	0.509	0.417	0.516	0.509	0.308	0.573
N	89,336,853	21,796,372	72,891,541	89,336,853	5,832,993	3,048,972

Panel B: Based on Width

	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width (÷ 100)	0.717 (0.005)	1.106 (0.009)	0.494 (0.008)	0.098 (0.020)	0.309 (0.012)	0.821 (0.076)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	70.629	85.755	78.899	59.955	56.506	72.482
R^2	0.549	0.447	0.584	0.521	0.389	0.603
N	89,336,853	21,796,372	69,705,830	13,401,877	5,832,993	3,048,972

Notes: See Table 5 in the main text. The only difference between this table and that in the text is that results above restrict attention to board configurations in which the minimal depth among W-moves is at least 50.

Appendix Table AT.3: Replication of Table 4, Controlling for Time Pressure

Panel A: Based on Depth

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Depth ($\div$ 100)	–0.775	0.014	–0.226	0.035
	(0.002)	(0.001)	(0.003)	(0.002)
Seconds Left per Move	0.000	–0.000	0.002	–0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- $\times$ D- $\times$ L-Moves				
$\times$ Depth of Other Moves	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	16.693	0.475	20.481	0.660
R^2	0.494	0.233	0.663	0.234
N	3,238,254,715	276,991,030	372,397,046	104,556,541

Panel B: Based on Width

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Width ($\div$ 100)	–0.335	0.008	–0.746	0.031
	(0.002)	(0.001)	(0.003)	(0.002)
Seconds Left per Move	0.000	–0.000	0.001	–0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- $\times$ D- $\times$ L-Moves				
$\times$ Depth of Other Moves $\times$ Own Depth	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	16.693	0.475	20.481	0.660
R^2	0.555	0.285	0.693	0.286
N	3,238,254,715	276,991,030	372,397,046	104,556,541

Notes: See Table 4 in the main text. The only difference between this table and that in the text is that results above control for time pressure, i.e., the number of seconds per move the player has left she follows the shortest W-path.

Appendix Table AT.4: Replication of Table 5, Controlling for Time Pressure

Panel A: Based on Depth

	Probability of Choosing Other W-Move					
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)
Depth (÷ 100)	1.499	1.526	0.324	0.917	1.638	1.802
	(0.003)	(0.010)	(0.002)	(0.007)	(0.004)	(0.028)
Seconds Left per Move	0.001	0.000	0.002	0.286	–0.000	–0.000
	(0.000)	(0.000)	(0.000)	(0.012)	(0.000)	(0.000)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	85.649	88.956	92.335	70.601	67.824	86.057
R^2	0.408	0.448	0.320	0.511	0.331	0.459
N	3,217,115,234	368,886,629	2,797,083,668	83,387,752	138,639,890	56,772,033

Panel B: Based on Width

	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width (÷ 100)	0.546	0.808	0.410	0.112	0.165	0.714
	(0.001)	(0.003)	(0.002)	(0.021)	(0.003)	(0.012)
Seconds Left per Move	0.002	0.001	0.002	0.281	0.013	0.000
	(0.000)	(0.000)	(0.000)	(0.022)	(0.001)	(0.000)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	85.649	88.956	90.695	59.929	67.824	86.057
R^2	0.469	0.485	0.443	0.523	0.460	0.510
N	3,217,115,234	368,886,629	2,643,066,500	12,504,657	138,639,890	56,772,033

Notes: See Table 5 in the main text. The only difference between this table and that in the text is that results above control for time pressure, i.e., the number of seconds per move the player has left she follows the shortest W-path.

Appendix Table AT.5: Replication of Table 4, First Move in Series Only

Panel A: Based on Depth

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Depth (÷ 100)	–0.452 (0.002)	0.023 (0.001)	–0.034 (0.004)	0.070 (0.003)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	33.899	0.938	36.255	1.509
R^2	0.525	0.248	0.722	0.270
N	223,801,960	75,288,017	59,250,718	29,795,003

Panel B: Based on Width

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Width (÷ 100)	–0.158 (0.005)	0.013 (0.002)	–0.712 (0.008)	0.043 (0.003)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	33.899	0.938	36.255	1.509
R^2	0.593	0.314	0.759	0.333
N	223,801,960	75,288,017	59,250,718	29,795,003

Notes: See Table 4 in the main text. The only difference between this table and that in the text is that results above restrict attention to only the very first move in a series of moves from winning positions.

Appendix Table AT.6: Replication of Table 5, First Move in Series Only

Panel A: Based on Depth

	Probability of Choosing Other W-Move					
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)
Depth (÷ 100)	1.059 (0.004)	1.453 (0.016)	0.243 (0.003)	0.790 (0.012)	1.101 (0.005)	1.244 (0.026)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	76.431	85.180	87.770	62.998	63.249	76.812
R^2	0.512	0.504	0.515	0.603	0.360	0.556
N	214,042,424	57,226,280	180,924,795	15,344,997	20,229,107	5,328,759

Panel B: Based on Width

	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width (÷ 100)	0.439 (0.004)	0.865 (0.008)	0.321 (0.005)	0.130 (0.086)	0.154 (0.007)	0.661 (0.067)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	76.431	85.180	84.758	52.085	63.249	76.812
R^2	0.579	0.534	0.611	0.606	0.478	0.594
N	214,042,424	57,226,280	172,871,972	2,464,839	20,229,107	5,328,759

Notes: See Table 5 in the main text. The only difference between this table and that in the text is that results above restrict attention to only the very first move in a series of moves from winning positions.

Appendix Table AT.7: Replication of Table 4, TWIC Data

Panel A: Based on Depth

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Depth (÷ 100)	–0.911	0.002	–0.987	–0.000
	(0.020)	(0.006)	(0.109)	(0.001)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves				
× Depth of Other Moves	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	26.542	0.077	23.835	0.150
R^2	0.645	0.422	0.582	0.367
N	5,132,343	1,177,248	853,980	296,615

Panel B: Based on Width

	Probability of Choosing Move			
	W-Moves	L-Moves	W-Moves	L-Moves
Width (÷ 100)	–0.511	–0.002	–0.566	–0.009
	(0.018)	(0.002)	(0.032)	(0.005)
Fixed Effects:				
Player	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves				
× Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes
Board Configurations	All	All	No D-Moves	No D-Moves
Mean of LHS Variable (%)	26.542	0.077	23.835	0.150
R^2	0.679	0.357	0.504	0.337
N	5,132,343	1,177,248	853,980	296,615

Notes: See Table 4 in the main text. The only difference between this table and that in the text is that results above are based on data from The Week in Chess.

Appendix Table AT.8: Replication of Table 5, TWIC Data

Panel A: Based on Depth

	Probability of Choosing Other W-Move					
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)
Depth (÷ 100)	2.485 (0.061)	3.149 (0.309)	0.410 (0.032)	1.513 (0.099)	2.663 (0.081)	2.506 (0.063)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	81.576	87.459	93.359	74.058	68.824	81.525
R^2	0.539	0.473	0.511	0.599	0.479	0.539
N	5,064,773	847,303	4,311,578	958,023	339,083	4,583,088

Panel B: Based on Width

	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width (÷ 100)	0.554 (0.019)	0.566 (0.032)	0.418 (0.019)	1.850 (0.522)	0.851 (0.121)	0.554 (0.019)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	81.576	87.459	88.375	65.657	68.824	81.525
R^2	0.577	0.455	0.585	0.606	0.583	0.576
N	5,064,773	847,303	4,070,864	147,925	339,083	4,583,088

Notes: See Table 5 in the main text. The only difference between this table and that in the text is that results above are based on data from The Week in Chess.

Appendix Table AT.9: Summary Statistics, Experimental Data

Variable	Mean	SD	Percentile				N
			25%	50%	75%	95%	
A. Subject Characteristics							
Registration:							
Lichess User	0.38	0.48					3,966
Chess.com User	0.85	0.35					3,966
Age (in Years):							
$\leq$ 20	0.44	0.50					3,112
21 – 30	0.41	0.49					3,112
31 – 40	0.11	0.31					3,112
41 – 50	0.02	0.14					3,112
51 – 60	0.01	0.09					3,112
$>$ 60	0.01	0.10					3,112
Gender:							
Male	0.92	0.27					3,112
Female	0.04	0.19					3,112
Other and Prefer Not to Say	0.05	0.21					3,112
Region:							
North America	0.51	0.50					3,112
Central and South America	0.04	0.19					3,112
Western Europe	0.21	0.41					3,112
Eastern Europe	0.07	0.26					3,112
East Asia	0.02	0.15					3,112
South Asia	0.05	0.22					3,112
Australia and Oceania	0.03	0.17					3,112
Africa	0.02	0.14					3,112
Other	0.05	0.21					3,112
Experience Playing Chess (in Years):							
$<$ 1	0.23	0.42					3,101
1 – 2	0.29	0.45					3,101
3 – 5	0.20	0.40					3,101
6 – 10	0.10	0.30					3,101
$>$ 10	0.18	0.38					3,101
Frequency Playing Online Chess:							
Daily	0.49	0.50					3,101
Weekly	0.34	0.47					3,101
Monthly	0.12	0.32					3,101
Almost Never	0.06	0.23					3,101
Never	0.00	0.06					3,101
Frequency Playing Over-the-Board Chess:							
Daily	0.03	0.16					3,101
Weekly	0.16	0.37					3,101
Monthly	0.22	0.42					3,101
Almost Never	0.46	0.50					3,101
Never	0.13	0.34					3,101
Understanding of Instructions:							
Already Knew about Move Types	0.78	0.42					3,076
Instructions Were Clear	0.85	0.36					3,075
Didn't Know about Move Types and Instructions Weren't Clear	0.05	0.22					3,071
B. Move Evaluations							
Subject Level:							
Evaluations Completed	21.95	6.94	25	25	25	25	3,966
Tickets Earned	13.85	5.91	11	15	18	22	3,966
Move-Level Performance:							
Correctly Evaluated	0.63	0.48					87,060
Response Time (in Seconds)	11.25	6.53	7	10	14	24	87,059
C. Move Characteristics							
True Type:							
Winning	0.34	0.47					87,060
Drawing	0.33	0.47					87,060
Losing	0.34	0.47					87,060
Complexity:							
Depth	25.52	14.45	13	26	38	48	58,470
Width	9.00	5.41	4	8	14	18	87,060

Notes: Table displays summary statistics for selected variables in the data from our experiment.

Appendix Table AT.10: Experimental Response Times Increase in Complexity

Panel A: All Online-Chess Players

	Response Time							
	(1A)	(2A)	(3A)	(4A)	(5A)	(6A)	(7A)	(8A)
Depth	0.069		0.081	0.085			0.077	0.082
	(0.003)		(0.005)	(0.006)			(0.005)	(0.006)
Width		0.125			0.136	0.121	0.106	0.087
		(0.005)			(0.010)	(0.011)	(0.012)	(0.013)
Fixed Effects:								
Board Position	No	No	Yes	No	Yes	No	Yes	No
Board Position × Piece	No	No	No	Yes	No	Yes	No	Yes
Mean of LHS Variable (in sec.)	11.079	11.246	11.079	11.079	11.246	11.246	11.079	11.079
R^2	0.025	0.011	0.104	0.133	0.099	0.131	0.105	0.133
N	58,469	87,059	58,469	58,469	87,059	87,059	58,469	58,469

Panel B: Lichess Users

	Response Time							
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)	(7B)	(8B)
Depth	0.093		0.087	0.092			0.084	0.089
	(0.004)		(0.009)	(0.010)			(0.009)	(0.011)
Width		0.134			0.129	0.103	0.100	0.069
		(0.009)			(0.017)	(0.019)	(0.022)	(0.024)
Fixed Effects:								
Board Position	No	No	Yes	No	Yes	No	Yes	No
Board Position × Piece	No	No	No	Yes	No	Yes	No	Yes
Mean of LHS Variable (in sec.)	10.942	11.136	10.942	10.942	11.136	11.136	10.942	10.942
R^2	0.043	0.012	0.201	0.253	0.189	0.248	0.202	0.253
N	22,381	33,421	22,381	22,381	33,421	33,421	22,381	22,381

Notes: See Table 1 in the main tex. The only difference between this table and that in the text is that regression results above rely in response times as dependent variable.