Complexity and Satisficing: Theory with Evidence from Chess*

Yuval Salant
Kellogg School of Management, Northwestern University

Jörg L. Spenkuch
Kellogg School of Management, Northwestern University and *NBER*

November 2024

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. Looking beyond chess, our work offers a blueprint for incorporating complexity at the level of individual objects into models of choice and for detecting satisficing outside of the laboratory.

*This manuscript supersedes our previous working paper entitled "Complexity and Choice" (Salant and Spenkuch 2022). 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 Bar-Ilan University, Boston College, Chicago Booth, 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, 2021 SITE Conference, and 2024 BRIC 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 6 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 objects' 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.[1]

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 some 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 a class of maximization-based models. Following Manzini and Mariotti (2014), we postulate a two-stage 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-based models yield qualitatively different predictions about the effect of object complexity on choice

[1]The term "complexity" is typically associated with features of the environment that increase the cognitive costs of making decisions. Such costs may arise for multiple reasons. Here, we focus on the idea that individual objects might be inherently complex and, therefore, difficult to evaluate. Object complexity, in turn, gives rise to complexity at the choice-set level. We note, however, that decision problems may be complex without individual alternatives being difficult to evaluate, e.g., when choice sets are large (Iyengar and Lepper 2000). In Salant and Spenkuch (2022), we test for choice overload among chess players. After controlling for the quality of available moves, there is no evidence to suggest that choice-set size adversely affects decision-making in our setting.

probabilities. Under satisficing, increasing the complexity of a high-value alternative decreases the probability that the corresponding object is deemed satisfactory, which, in turn, increases the choice probabilities of all other available alternatives. Maximization, however, predicts that an increase in complexity reduces the choice probabilities of objects with weakly higher values. The intuition behind this comparative static is that noisier evaluations increase the probability that the respective object will be perceived as better than superior ones. The theory thus yields a new, general empirical test that leverages object complexity to distinguish between satisficing and maximization-based choice models.

As a proof of concept, we implement this test in the context of chess endgames. Chess is a finite, two-player, zero-sum game with perfect information. Although it is theoretically trivial, chess remains practically intractable. Evaluating individual moves often strains the bounds of human cognition, which makes chess an ideal setting to leverage complexity in order to pit satisficing against maximization.

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, i.e., the number of decision nodes. Although computational constraints prevent us from calculating the total number of nodes in every relevant game tree, we can proxy for the size of the tree 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.

In order to validate our complexity measures and to provide evidence on the link between object complexity and 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.[2] We find that the accuracy of participants' responses declines significantly with moves' depth and width. That is, more complex moves are more difficult to evaluate.

Our main empirical results relate object complexity to 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 of moves in all rated games on the platform from January 2013 through August 2020.[3] 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.

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 significantly affect both individual and strategic outcomes.

Our contribution relative to extant theoretical work is twofold. First, we propose and

[2]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.

[3]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.

experimentally test a new notion of complexity. According to this notion, complexity manifests at the level of individual alternatives and leads to errors in the perception of value—in line with recent work on cognitive imprecision (see, e.g., Woodford 2020).[4] Second, we derive comparative static results relating object complexity to choice behavior in different models of decision making.

In addition, our work complements a growing experimental literature on complexity and satisficing (see, e.g., Huck and Weizsäcker 1999; Gabaix et al. 2006; Bossaerts and Murawski 2017; Oprea 2022). Rubinstein (2007, 2016) uses response times to distinguish instinctive choices from those that require significant cognitive effort. Caplin et al. (2011) provide evidence that individuals satisfice 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. Enke and Graeber (2023) demonstrate that cognitive uncertainty depends on the complexity of the experimental task, and that it can rationalize seemingly distinct behavioral phenomena. Overall, laboratory experiments confirm the idea that complexity can greatly affect decision making.

Outside of the laboratory, however, tests of fundamental decision-theoretic concepts remain rare.[5] As Chiappori et al. (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 affects decision making outside of the laboratory; and we do not have empirical tests to detect satisficing in real-world environments.

Chess endgames provide an almost ideal setting to study complexity and test for satisficing. 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

[4] Another relevant literature is on the drift-diffusion model. The emphasis in this model is on binary choice, and errors arise because perceptions of the difference in alternatives' values evolve stochastically. As the difference in values decreases or as the amount of noise in the stochastic process increases, the binary choice problem may be thought of as being more complex. See Gonçalves (2024) for a recent discussion.

[5] 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 et al. (2002), Palacios-Huerta (2003), and Hsu et al. (2007) all study minimax play in professional sports, while Spenkuch et al. (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).

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.[6]

Our chief contribution relative to extant experimental work is threefold. First, we document the importance of complexity and satisficing for decision making outside of the laboratory. Second, we provide evidence to suggest that object complexity is a key driver of evaluation errors. That is, we provide evidence on the mechanism through which complexity affects choice behavior. Third, we develop a new empirical test that has the potential to distinguish satisficing from maximization-based choice behavior in both experimental and observational data sets. This test is not specific to chess.

2. Theory

Our analysis begins by developing the notion of object complexity. After incorporating this notion into two leading choice models—satisficing and maximization—we establish that these models yield qualitatively different comparative statics regarding the effect of object complexity on choice probabilities.

2.1. *Object Complexity*

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 complexity. To fix ideas, complexity may be interpreted as the size of the respective object, or the length of its description. The DM does not know v, and she may or may not know σ.

When assessing an alternative's value, the DM needs to contend with noise due to object complexity. Her assessment may be affected by beliefs, memory, information, computational constraints, etc. We abstract from these specifics and focus on the output of the evaluation process. This is the relevant component for our analysis. We call each potential output a *score* and the (non-degenerate) distribution of potential outputs a *score distribution.* One useful way of thinking about the score distribution is as a summary of all possible "perceived values" of the object after deliberation.

To demonstrate the richness and flexibility of our framework, here are a few examples of natural evaluation processes that are accommodated by the analysis below.

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*

[6]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), and Levitt et al. (2010, 2011).

the object. The score 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. For example, suppose that the prior belief is given by $N(v_0, \sigma_0)$ and the signal is distributed $N(v, \sigma)$. Then, given signal realization y, the score equals $v_0 + \frac{\sigma_0^2}{\sigma_0^2+\sigma^2}(y - v_0)$, and the score distribution is $N\left(v_0 + \frac{\sigma_0^2}{\sigma_0^2+\sigma^2}(v - v_0), \frac{\sigma_0\sigma}{\sqrt{\sigma_0^2+\sigma^2}}\right)$.*

EXAMPLE 3 (Partial Confidence): *The DM is partially confident that the value of the object is v_0. She consults a supplementary source of information to obtain a potentially different value y, and forms the score $\alpha v_0 + (1 - \alpha)y$, where α is her initial degree of confidence. The score is then distributed according to the respective linear transformation of the distribution of the supplemental information.*

Let F and f denote the CDF and PDF associated with the score distribution, 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, for any magnitude, positive and negative deviations from the mean of the score distribution are equally likely to occur.

(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, object complexity increases the amount of noise that the DM needs to contend with in the evaluation process. It is, therefore, a property of the family of the score distributions that are associated with different alternatives 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 for any $\epsilon > 0$:*

1. *$F_b(\mu(v_b) - \epsilon) - F_a(\mu(v_a) - \epsilon) \geq 0$, with strict inequality whenever $F_b(\mu(v_b) - \epsilon) > 0$; and*

2. $F_b(\mu(v_b)+\epsilon) - F_a(\mu(v_a)+\epsilon) \leq 0$*, with strict inequality whenever* $F_b(\mu(v_b)+\epsilon) < 1$.[7]

Condition 1 is a single-crossing property. It states that if both score distributions were to be demeaned, then their CDFs would cross exactly once at zero. An increase in object 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 every object, 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 if, for any alternative, the associated score is distributed on $[\mu(v)-\sigma, \mu(v)+\sigma]$. Other examples include the Logistic and Laplace families with location parameters corresponding to objects' values and scale parameters corresponding to their complexities.

2.2. *Choice Behavior: Satisficing vs. Maximization*

We consider two models that map scores into choice behavior. First, building on Simon (1955), we incorporate object complexity into a *satisficing* procedure. In satisficing, 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 deems satisfactory.[8] When choosing from a set of alternatives $A \subseteq X$, the DM first lists all objects in A in some random order. She then evaluates them sequentially. Starting with the first alternative, 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. The DM 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.[9]

We allow for any distribution of evaluation orders that assigns positive probability to all orderings and satisfies *value invariance*. Value invariance means that if 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.[10]

[7] Part 2 of Condition 1 is implied by symmetry. We include it so that the condition is self-contained.

[8] When the aspiration level T is a function of previous scores, the first (second) part of Theorem 1.A below continues to hold for any alternative whose mean score exceeds the supremum (infimum) of aspiration levels.

[9] Our results continue to hold if the DM chooses any alternative with equal probability when reaching the end of the list. Assuming, however, that the DM chooses the highest-score alternative when reaching the end of the list makes satisficing closer to 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 maximization. The likelihood of the latter event declines exponentially fast as the number of alternatives in the set grows.

[10] Let $O_1 = (a_1, \ldots, a_n)$ and $O_2 = (b_1, \ldots, b_n)$ with $n = |A|$. Value invariance requires that if $v_{a_k} = v_{b_k}$ for $k = 1, \ldots, n$, then O_1 and O_2 are drawn with the same probability.

The second model we consider incorporates object complexity into the *maximization-from consideration-sets* procedure of Manzini and Mariotti (2014). In this model, the DM chooses from a choice set A using a two-stage process. In the first stage, she draws a consideration set $S \subseteq A$, with $|S| \geq 2$, according to some probability distribution P_A. In the second stage, the DM evaluates 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., $P_A(S_A) = P_B(S_B)$, whenever the composition of values in A and S_A is the same as that in B and S_B, respectively.[11] Note, value invariance is trivially satisfied if $P_A(A) = 1$ for every $A \subseteq X$. In this case, we obtain the standard random utility model.

Both satisficing and maximization induce random choice functions that assign to every choice set A a probability distribution over the alternatives in A. Choice behavior is stochastic because object evaluations are noisy and because either the evaluation order (in satisficing) or the consideration set (in maximization) is random. Our main result, stated in Theorem 1 below, establishes how changes in object complexity affect choice probabilities in either model. The key takeaway is that object complexity has qualitatively different effects, depending on whether DMs rely on satisficing or maximization.

THEOREM 1: *Fix 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 $\{b\} = B - A$.*

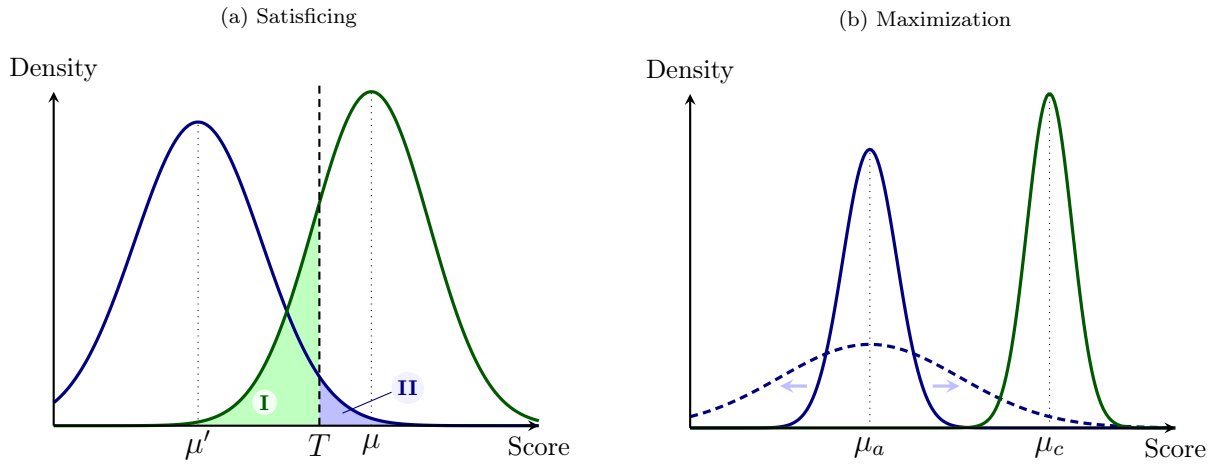
PART A: *Suppose the DM satisfices. If $\mu(v) > T$, then the choice probability of a in A is larger than the choice probability of b in B, and any other non-zero choice probability in A is smaller than in B. If $\mu(v) < T$, then the rankings of the choice probabilities reverse.*

PART B: *Suppose the DM uses a maximization-from-consideration-sets procedure. 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) there is a consideration set $S \subseteq A$ such that $\{a, c\} \subsetneq S$ with $P_A(S) > 0$, and (ii) the support of the score densities of a and c is the real line.*

To develop intuition for how object complexity affects choice probabilities under satisficing, consider the left panel in Figure 1. It 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

[11] Let $|v|^A$ denote the number of alternatives with value v in A. Two sets A and B have the same composition of values if, for every $v \in \mathcal{R}$, $|v|^A = |v|^B$.

Figure 1: Choice with Noisy Evaluations



Notes: Figure provides graphical intuition for Parts A (left panel) and B (right panel) of Theorem 1.

above T leads to more probability mass in area Ⓘ, 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 these probabilities must offset the decline in the choice probability of the object that is now more complex.[12]

By contrast, Theorem 1.B 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, we turn to the right panel in Figure 1. There are two alternatives, a and c, with $v_c > v_a$. Almost all the mass of their score distributions is concentrated in narrow intervals. 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. Given such a consideration set, the DM would choose a only if its score exceeds that of c. Since higher object complexity corresponds to more probability mass in the tails of the score distribution, an increase in the complexity of a makes this event more likely.

Theorem 1 provides a theoretical foundation for a new empirical test that leverages object complexity to distinguish between satisficing and maximization. In what follows, we implement this test to detect satisficing in the context of chess.

[12]For an alternative with an expected score below T, an increase in object complexity leads to more probability mass in area Ⓘ, which means that the comparative statics reverse.

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 position 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 chooses a particular move.[13] 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).

We validate these complexity measures in Section 6, where we report results from 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. The experimental results indicate that the accuracy of participants' responses declines significantly with moves' depth and width, suggesting that both are useful proxies of object complexity.

As for moves' values, 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 rely on backward-induction logic to determine the values of all legal moves in every endgame position, which are then stored in so-called tablebases.[14] For W- and

[13] 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.

[14] For additional details on the algorithmic analysis of endgame positions and tablebases, see Appendix B.

L-moves, these algorithms also calculate DTM. For D-moves, however, there is no general approach to even identify them 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.

3.2. *Predictions*

In order to translate the comparative statics of the satisficing model into concrete predictions for the case of chess, we need to specify players' aspiration levels. 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 two testable implications of Theorem 1.

PREDICTION 1: *If players satisfice, then, 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.*

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.*

The latter prediction directly pits satisficing against maximization.

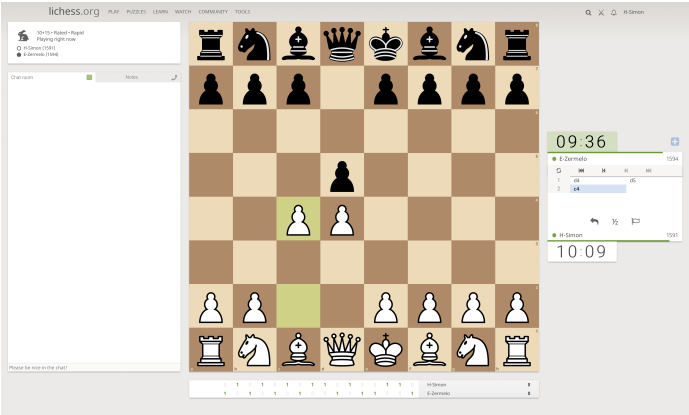
4. Data Sources and Descriptive Statistics

In order to test Predictions 1 and 2, we introduce a new, large observational data set that contains information on choice behavior in chess endgames.

4.1. *Data Sources*

The core of our 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 2 for a screenshot of a typical game). Although Lichess offers a choice between many different time limits, the majority of

Figure 2: 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.

games that are hosted on the platform can be broadly classified as "speed chess."[15] Lichess further distinguishes between casual and rated games. The latter determine player ratings and are therefore only available to registered users. 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 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 DTM of essentially all legal moves in endgame positions with six of fewer pieces.[16] We retrieve this information by running several billion queries against the Syzygy and Nalimov tablebases (Nalimov et al. 2000; Man 2013).[17] To compute width, we construct for every legal endgame

[15]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 analyses restrict 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 D).

[16]The only exceptions are positions with castling rights and positions 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 analysis excludes all board positions for which information on DTM is not available.

[17]As a technical side note, the Syzygy tablebases do not contain information on DTM. In contrast to the

move in the data 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 depth, we restrict attention to board 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 D.4).[18]

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, retrieve information on values 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.

[18]Out of the twenty point estimates in Appendix D.4 sixteen are statistically significant and have the same sign as their counterparts in Tables 3 and 4 below. For the remaining four coefficients the 95%-confidence intervals include both negative and positive values.

Table 1: 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.

4.2. *A First Look at the Data*

Table 1 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 mistake is about 43 percentage points (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.[19]

We next turn to the distribution of our complexity measures. The upper two panels in Figure 3 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 3 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 in order to correctly identify at least one winning move. 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 one move is relatively easy, others where accurately classifying any move likely exceeds the bounds of human cognition, and a great range of intermediate cases.

Figure 4 plots the observed frequency of mistakes as a function of the minimal depth and width among W-moves. Regardless of whether we rely on minimal depth or width—or other low-dimensional summary measures—we find that either predicts mistakes. The left panel of Figure 4 shows this based on the raw data, while the right panel reveals a similar relationship after controlling for the number of W-, D-, and L-moves in the choice set.

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 such a preference for winning quickly exists and if players are able to identify the depth 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.

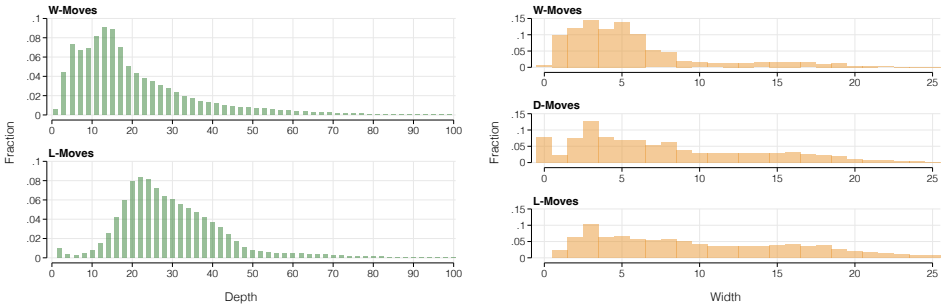
To rule out that this possibility drives our results, we follow a two-pronged approach. First, we conduct robustness checks in which we test Predictions 1 and 2 for choice sets in which the minimal depth among W-moves exceeds fifty (cf. Appendix D). While it is conceivable that players rank winning moves according to depth to mate when some of them are simple,

[19]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 a win.

Figure 3: Distribution of Complexity Measures

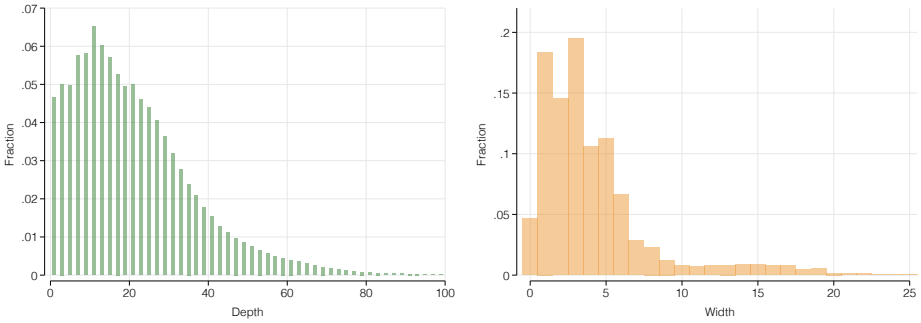
(a) Depth, by Type of Move

(b) Width, by Type of Move



(c) Minimal Depth Among W-Moves

(d) Minimal Width Among W-Moves

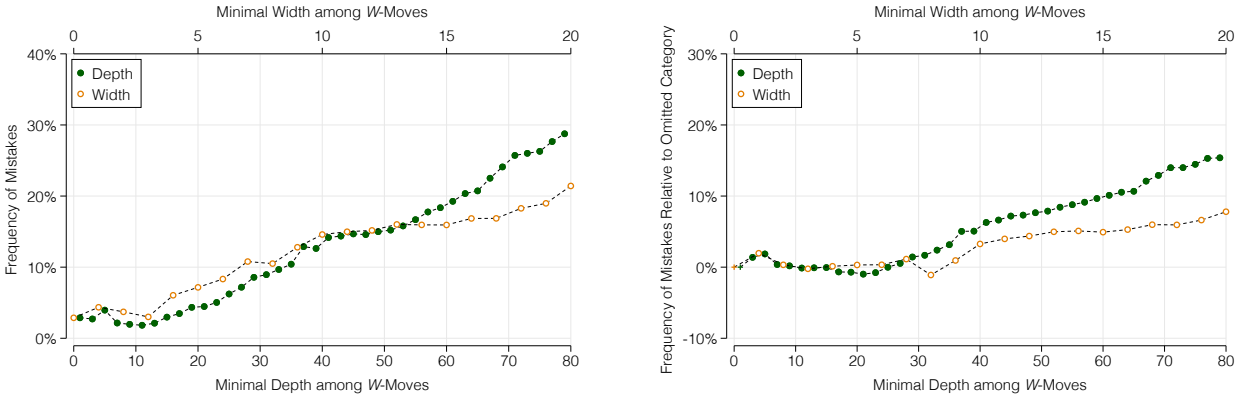


Notes: Panel (a) presents a histogram of moves' depth, separately by type of move. Panel (b) does so for moves' width. Panels (c) and (d) respectively depict the distribution of the minimal depth and width among all W-moves in the choice set.

Figure 4: 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.

Table 2: Object Complexity and Length of Subsequent Play

	Number of Subsequent Moves					
	(1)	(2)	(3)	(4)	(5)	(6)
Depth	0.343			0.121		
	(0.000)			(0.002)		
Width		0.758	–0.093		–0.081	0.002
		(0.001)	(0.001)		(0.003)	(0.003)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Depth	No	No	Yes	No	No	Yes
Type of Chosen Move	W-Move	W-Move	W-Move	L-Move	L-Move	L-Move
Mean of LHS Variable	15.647	15.647	15.647	13.288	13.288	13.288
R^2	0.197	0.129	0.235	0.329	0.314	0.340
N	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 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 player and Endgame, and are shown in parentheses.

it seems unlikely that they are able to do so when the available alternatives are all very complex.

Second, and perhaps more importantly, we present a set of complementary results that exploit variation in width *conditional on depth*. To appreciate why conditioning on depth is helpful, consider Table 2. 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.[20] The coefficient in column (2) reveals that, for W-moves, width is also positively correlated with length of play. The third column of Table 2, however, demonstrates that the relationship between width and length of play reverses if we condition on the depth of the move. Columns (4)–(6) show results from analogous regressions for L-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 DTM. Choices, however, are noisy. Sometimes the losing player chooses a move that

[20]DTM 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 DTM. 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 DTM.

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 complementary results 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.

5. How Does Object Complexity Affect Choice Frequencies?

We now proceed to test Predictions 1 and 2. Prediction 1 relates moves' complexity to the frequency with which the respective moves are chosen, under the assumption that the DM satisfices. The effect of complexity on own choice frequencies 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 choice set. The sign of this effect allows us to distinguish between satisficing and maximization.

5.1. *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:

$$Choose_a = \beta\, Complexity_a + \chi_p + \phi_{A\setminus a} + \varepsilon_a. \tag{1}$$

Here, $Choose_a$ is an indicator for whether player p facing choice set A executed 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 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 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.

Table 3: 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. (1) 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.

The upper panel of Table 3 shows results from estimating the regression model in eq. (1) 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 frequencies

of L-moves increase in their complexity. The results in Table 3 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 frequency of about 13.9 p.p.; and a standard deviation increase in width is associated with a 3.9 p.p. decrease in the frequency with which the respective move is chosen. Our findings, therefore, suggest that object complexity is an empirically important determinant of choice.

5.2. *Tests of Prediction 2*

To pit satisficing against maximization we restrict attention to W-moves and modify the regression specification in eq. (1) by replacing the left-hand-side variable with an indicator for whether the player chose a W-move other than a. In symbols:

$$(2) \qquad Choose\ Other\ W\text{-}Move_a = \gamma\, Complexity_a + \chi_p + \phi_{A\setminus a} + \eta_a.$$

Table 4 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 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) none of the available W-moves are easily recognizable as good (because minimal depth exceeds fifty; cols. 4A and 4B), (ii) small choice sets (with ten or fewer moves; cols. 5A and 5B), and (iii) long time controls (so that each player has, in expectation, at least twenty-five minutes for deliberation per game; cols. 6A and 6B). 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 4: 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 (÷ 100)	1.497	1.524	0.324	0.911	1.635	1.782
	(0.003)	(0.010)	(0.002)	(0.007)	(0.004)	(0.021)
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 (%)	85.651	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

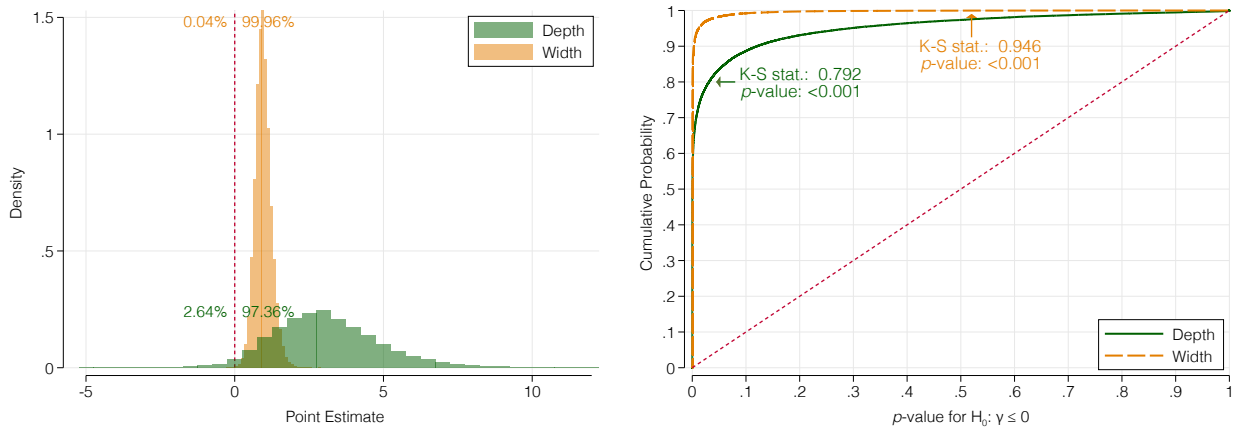
	Probability of Choosing Other W-Move					
	(1B)	(2B)	(3B)	(4B)	(5B)	(6B)
Width (÷ 100)	0.543	0.805	0.407	0.098	0.163	0.674
	(0.001)	(0.003)	(0.002)	(0.020)	(0.003)	(0.011)
Fixed Effects:						
Player	Yes	Yes	Yes	Yes	Yes	Yes
Number of W- × D- × L-Moves × Depth of Other Moves × Own Depth	Yes	Yes	Yes	Yes	Yes	Yes
Sample	Full	No D-Moves	Excl. Simplest Move	High Complexity	Small Choice Sets	Long Time Controls
Mean of LHS Variable (%)	85.651	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. (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-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 5: Testing Prediction 2 at the Player Level

(a) Player-Level Estimates of γ in Eq. (2)

(b) CDF of p-values for Estimates in Panel (a)



Notes: Figure presents player-level tests of Prediction 2. Panel (a) plots histograms of player-level estimates of γ in eq. (2), 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 4. 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. (2). We then plot the distribution of the resulting coefficients in the left panel of Figure 5. 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.

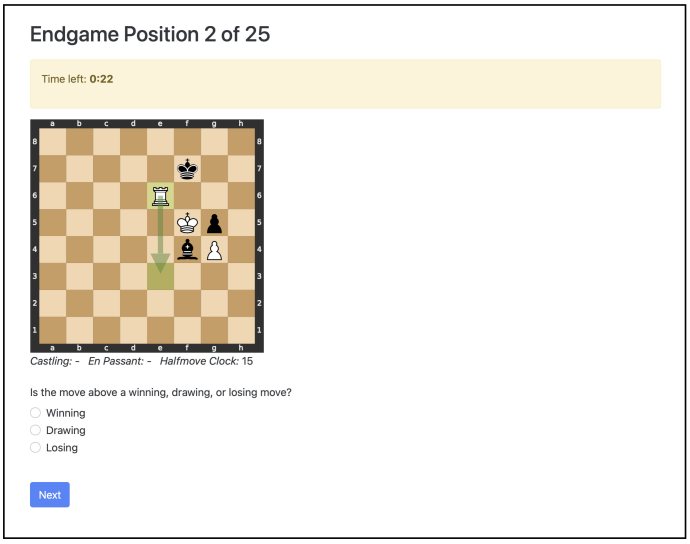
The right panel of Figure 5 shows the empirical CDFs of the one-sided p-values for our individual-level estimates. The relevant p-values are one-sided because the null hypothesis of maximization from consideration sets implies that $\gamma \leq 0$ (cf. Prediction 2). Under this null, the distribution of p-values should first-order stochastically dominate the uniform distribution.[21] 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 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.

6. Connecting Object Complexity and Evaluation Errors

The basic idea behind our approach is that higher object complexity leads to noisier perceptions of value (cf. Condition 1). Noisier evaluations should in turn increase the frequency with

[21] 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 6: 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.

which DMs misclassify moves, e.g., identify a W-move as a D- or an L-move.[22] To validate that our complexity measures do, indeed, capture how difficult it is to correctly classify a given move, we conducted an experiment with nearly four thousand online-chess players.

6.1. *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 6. 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.[23] We settled on this experimental task because it allows us

[22]Noisier evaluations imply more classification errors whenever DMs classify an alternative as a W-move if its score is high enough, as an L-move if its score is low enough, and as a D-move otherwise.

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

to relate moves' complexity to the accuracy of participants' evaluations. At the same time, it resembles the kind of puzzles that are popular among chess players.

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.[24] 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.[25] For additional details on the experimental setup, see Appendix E.[26]

6.2. *Experimental Results*

Figure 7 plots the raw frequency of incorrect evaluations against moves' complexity. In the left panel we use depth to measure complexity, while the right panel uses width instead. Since our analyses above 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.[27]

Consistent with the idea that object complexity injects noise into evaluations, Figure 7 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

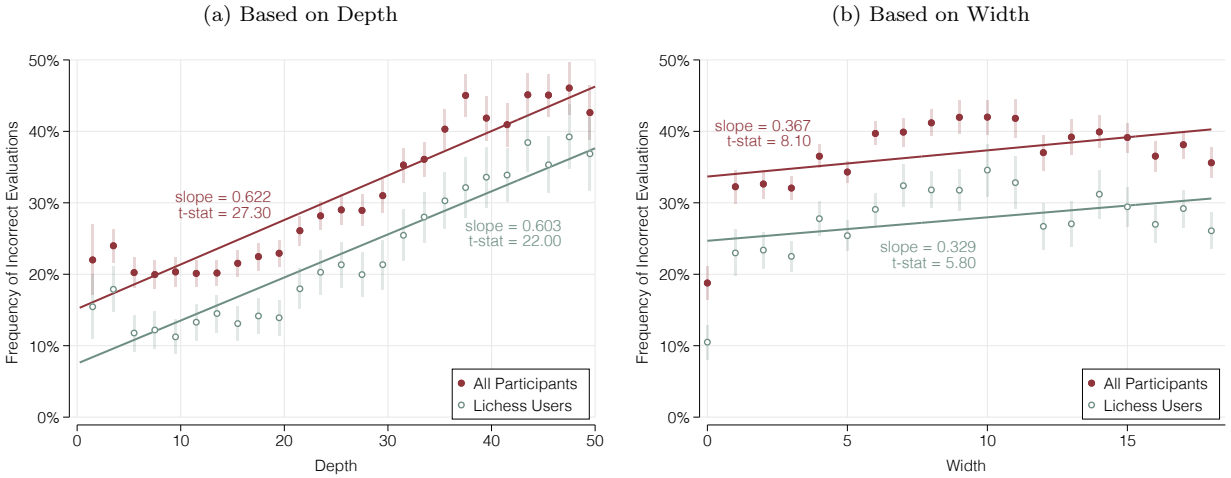
[24]The only constraint we imposed is that the depth and width of extracted moves do not exceed 50 and 18, respectively. Both numbers correspond roughly to the 95th percentiles of the respective marginal distributions.

[25]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 participants who did not complete the experiment.

[26]The pre-registration for the experiment is available at `https://osf.io/6zk9m`. We designed the experiment to test different hypotheses, including effects of time pressure and (lack of) gender differences. Below, we focus on H1 in the pre-analysis plan, leaving tests of other, unrelated hypotheses for future work. The empirical specifications that were preregistered correspond to columns (1A) and (2A) in Table 5 below. The remaining columns in Table 5 demonstrate that the results are qualitatively robust to restricting the sample to Lichess users and to limiting which types of comparisons identify the coefficient of interest.

[27]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 would be slightly stronger if we did not reweight observations to approximate the ratings distribution in the real-world data.

Figure 7: Object Complexity Predicts Incorrect Evaluations



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 4–5. Error bars correspond to 95%-confidence intervals, accounting for two-way clustering by participant and move.

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.[28] Among all players, a one standard deviation increase in depth is associated with a 9.0 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.[29] Consistent with the results in Figure 4, this suggests that depth might be a better proxy for object complexity than width.

Table 5 presents results from estimating variants of the following linear probability model:

$$Incorrect_a = \kappa\, Complexity_a + \psi_A + \xi_a, \tag{3}$$

where $Incorrect_a$ is an indicator for whether the subject made a mistake in classifying 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

[28] 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.

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

Table 5: 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. (3) 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 4–**??**. 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.

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 6 we might be comparing ♖e3 with ♖e7. The former is an L-move of depth 44, while the latter is an L-move whose depth is 40.

The first two columns of Table 5 reproduce the evidence in Figure 7. 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 5 show that either complexity measure is related to evaluation errors even after controlling for the other one, although the point estimates for width are only statistically significant in the pooled sample. Broadly summarizing, the experimental results support the fundamental

idea behind our notion of object complexity: Moves that, according to our measures, are inherently more complex are more difficult to evaluate.

7. Beyond Chess

Many objects that are of interest to economists are inherently complex. Our first contribution is to propose a notion of object complexity and to incorporate it into choice theory.

We associate object complexity with the size of the respective object and postulate that complexity induces noisier evaluations. Although we only validate the link between complexity and the accuracy of subjective evaluations in the context of chess, there are many other settings in which object complexity is likely relevant. For example, individuals may be flummoxed by contracts that span tens of pages with dozens of contingencies. Policymakers may struggle to evaluate long proposals with many clauses. Consumers may stumble when choosing among durable goods that differ along multiple attributes. A common thread in these and other examples is that the objects are large and are composed of smaller, more basic components, which together determine value.

This observation suggests a potential mechanism for why object complexity leads to noisier evaluations. Suppose that when evaluating alternatives, DMs break up large objects into smaller components, evaluate (perhaps some of) them, and then sum all of the "micro estimates" to form an overall assessment of value. If the micro estimates are subject to errors, then the overall evaluation of the object will be noisier for alternatives that consist of more components, i.e., larger objects.[30]

Thinking of complexity as the number of basic components is useful for measuring object complexity across a variety of settings. Take contracts, for example. On a basic theoretical level, a contract might be thought of as a set of clauses that map contingencies into actions and responsibilities (Battigalli and Maggi 2002).[31] Viewed through this lens, a natural measure of complexity is the number of contingencies or conditional clauses that are contained in the contract. Extracting conditional statements from text is a well-studied problem in computational linguistics and can be done at scale.[32] A more nuanced measure of contract complexity might also take into account that some clauses are more difficult to comprehend

[30]Specifically, assume that the object consists of a collection of k components and that the value of each component i is estimated with idiosyncratic noise according to a normal distribution $N(v_i, \sigma_i)$. Further suppose that an object's value, v, corresponds to the (weighted) sum of the values of the individual components. If the DM samples each component once and sums the estimated values to obtain an estimate of v, then that estimate is distributed according to the normal distribution $N(\sum_{j=1}^{k} v_j, \sqrt{\sum_{j=1}^{k} {\sigma_j}^2})$. Thus, as the number of components increases, overall evaluations become noisier.

[31]Similarly, a durable good might be thought of as a collection of attributes, and a policy proposal as a collection of propositions.

[32]See, e.g., Honnibal et al. (2020), Manning et al. (2014), or Bird and Loper (2004) for general-purpose natural language processing software. For a library specific to legal text, see Bommarito et al. (2021).

than others, which suggests weighting by linguistic attributes like length, lexical sophistication, density, and variability, or by syntactic difficulty.[33] Regardless of how researchers define a basic component in their application, we recommend that any particular measure be experimentally validated—as we did in Section 6. For the case of contracts, Besliu (2022) shows that participants in a carefully controlled laboratory experiment are more likely to select a dominated health insurance plan when it features more contingencies.

Our second contribution is to examine how DMs cope when choosing from a set of complex alternatives. We consider two leading mechanisms: satisficing and maximization. We develop a new empirical test that relies on variation in object complexity to distinguish between both mechanisms. Our data on endgame moves in chess are consistent with satisficing but not maximization.

While chess provides an almost ideal "proof of concept," our test is not specific to any particular environment. Following the approach in this paper, it is possible to pit satisficing against maximization in any setting that satisfies the following conditions. (i) Choices and choice sets are observable. (ii) The available alternatives can be ranked according to their value to the DM. (iii) It is possible to measure, or at least approximate, objects' complexity. If extended to other settings, the finding that many DMs satisfice might have important theoretical and applied implications.

To illustrate, consider an online marketplace. If consumers are satisficing rather than fully maximizing, then sellers have an incentive to influence the order in which products are being considered. To the extent that the ordering of products on the screen affects the order of consideration, we would expect sellers to compete for their products to be displayed more prominently. Consistent with this prediction, the Federal Trade Commission (FTC) recently alleged that Amazon.com extracts rents from third-party sellers by steering consumers towards products that are promoted via pay-to-play advertisements. Consumers are allegedly harmed because promoted products tend to be more expensive and of lower quality than similar items that are displayed below them. Per the FTC, 70% of Amazon's customers do not click past the first page of search results (see FTC 2023).

Moving beyond the ordering of products on the screen, our findings suggest that sellers also have an incentive to manipulate the complexity of product presentations. Suppose, for instance, that consumers are only willing to make a purchase if the perceived quality of a good exceeds their aspiration level, $\bar{q}$. If consumers are satisificing, then the demand for products whose true quality, q, exceeds (falls short of) $\bar{q}$ decreases (increases) as consumers'

[33] All of these characteristics correlate with how hard it is to comprehend the respective text. They can be quantified using tools from computational linguistics (see, e.g., Chen and Zechner 2011; Lu 2014; Kyle et al. 2017).

evaluations of the respective goods become noisier. Sellers are, therefore, expected to simplify the presentation of products with quality above $\bar{q}$. By contrast, sellers should strategically "complexify" the presentation of alternatives with quality below $\bar{q}$. Similar incentives to leverage object complexity are likely present in the design and presentation of policy proposals, in developing financial products, and in various other settings.

Supplementary Data

Supplementary data are available at *Review of Economic Studies* online.

Data Availability Statement (DAS)

Replication code as well as instructions for how to obtain the raw data that is used in this research are available on Zenodo at https://dx.doi.org/10.5281/zenodo.14393284.

References

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

Battigalli, P. and G. Maggi (2002). Rigidity, Discretion, and the Costs of Writing Contracts. *American Economic Review 92*(4), 798–817.

Besliu, C. (2022). Complexity in Insurance Selection: Cross-Classified Multilevel Analysis of Experimental Data. *Journal of Behavioral and Experimental Finance 35*, 100713.

Bird, S. and E. Loper (2004). NLTK: The Natural Language Toolkit. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, pp. 214–217. Association for Computational Linguistics.

Bommarito, M. J., D. M. Katz, and E. M. Detterman (2021). LexNLP: Natural Language Processing and Information Extraction for Legal and Regulatory Texts. In R. Vogl (Ed.), *Research Handbook on Big Data Law*, pp. 216–227. Cheltenham, UK: Edward Elgar.

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.

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.

Chen, M. and K. Zechner (2011). Computing and Evaluating Syntactic Complexity Features for Automated Scoring of Spontaneous Non-Native Speech. In *Proceedings of the 49th Annual Meeting*

of the Association for Computational Linguistics: Human Language Technologies – Volume 1, pp. 722–731. Association for Computational Linguistics.

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. and T. Graeber (2023). Cognitive Uncertainty. *Quarterly Journal of Economics 138*(4), 2021–2067.

FTC (2023). Federal Trade Commission v. Amazon.com, Inc. Case No.: 2:23-cv-01495-JHC.

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.

Honnibal, M., I. Montani, S. Van Landeghem, and A. Boyd (2020). spaCy: Industrial-strength Natural Language Processing in Python.

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.

Iyengar, S. and M. Lepper (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology 79*(6), 995–1006.

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.

Kyle, K., S. Crossley, and C. Berger (2017). The Tool for the Automatic Analysis of Lexical Sophistication (TAALES): Version 2.0. *Behavior Research Methods 50*, 1030–1046.

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.

Lu, X. (2014). *Computational Methods for Corpus Annotation and Analysis*. New York: Springer.

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

Manning, C. D., M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky (2014). The Stanford CoreNLP Natural Language Processing Toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pp. 55–60.

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

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.

Salant, Y. and J. Spenkuch (2022). Complexity and Choice. NBER Working Paper No. 30002.

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.

Woodford, M. (2020). Modeling Imprecision in Perception, Valuation, and Choice. *Annual Review of Economics 12*, 579–601.

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.