CESifo WORKING PAPERS

9239
2021

August 2021

Complexity and Choice

Yuval Salant, Jörg L. Spenkuch

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9239*

Complexity and Choice

Abstract

We study two dimensions of complexity that may interfere with individual choice. The first one is object complexity, which corresponds to the difficulty in evaluating any given alternative in a choice set. The second dimension is composition complexity, which increases when suboptimal alternatives become more similar to optimal ones. We develop a satisficing-with-evaluation-errors theory that incorporates both dimensions and delivers sharp empirical predictions about their effect on choice behavior. We confirm these predictions in a novel data set with information on hundreds of millions of decisions in chess endgames. First, as the object complexity of an optimal (suboptimal) alternative increases, it becomes less (more) likely to be chosen. Second, even highly experienced decision-makers are more likely to make mistakes when choosing from sets with higher composition complexity. These findings help to shed some of the first light on the effect of complexity on choice behavior outside of the laboratory.

JEL-Codes: D910.

Keywords: complexity, choice, satisficing, bounded rationality.

Yuval Salant
MEDS Department, Kellogg School of Management, Northwestern University
2211 Campus Dr.
USA – Evanston, IL 60208
y-salant@kellogg.northwestern.edu

Jörg L. Spenkuch
MEDS Department, Kellogg School of Management, Northwestern University
2211 Campus Dr.
USA – Evanston, IL 60208
j-spenkuch@kellogg.northwestern.edu

July 2021
We are grateful to Sandeep Baliga, Tim Feddersen, Peter Klibanoff, Daniel Martin, and Pablo Montagnes as well as seminar participants at Northwestern and UC Santa Barbara for helpful suggestions. This research was supported in part through the computational resources provided by the Quest high-performance computing facility at Northwestern University.

1. Introduction

Individuals often invest significant cognitive resources into making decisions. Think, for example, about choosing an insurance plan or buying a house. At least two dimensions of the environment complicate choice in these and many other settings. The first one is that any given alternative may be intrinsically complex and therefore difficult to evaluate. An insurance plan, for instance, might include tens or even hundreds of clauses and contingencies, and many different attributes are relevant when assessing the value of a particular house. The second dimension is that the choice set itself may be complex. For example, the collection of available insurance plans could be quite large, and various options may appear to be very similar to one another. In this paper, we ask how object complexity, which refers to complexity at the alternative level, and composition complexity, which refers to complexity at the set level, affect real-world decision-making.

Studying this question requires an environment with observable variation in both sources of complexity. To examine the effect of complexity on the overall quality of choice, it is also useful for the environment to support an objective measure of the alternatives' values. We turn to endgames in chess because they satisfy both requirements.

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 move is of one of three types: it either allows White to subsequently force a win, Black to force a win, or both sides to achieve a draw. Although computing these types is generally infeasible in the opening and middlegame phases, endgames with up to six chess pieces have been definitively solved by modern computers. We can therefore assign values to all endgame moves and identify suboptimal decisions.

Different moves in chess correspond to different subgames of possible future moves. Some of these subgames are short and straightforward to analyze, whereas others are long and complicated. To study the consequences of object complexity, we determine the "minimax length" of equilibrium play in a particular subgame and use it as a proxy for the difficulty in evaluating the associated move. At the choice-set level, there are many millions of board configurations that vary in the number and types of legal moves. We rely on this variation to study composition complexity.

Our analysis begins by developing an empirically testable theory of complexity and choice. The theory has two key ingredients. The first one adapts the familiar random-utility framework to study complexity. In our model, the decision maker does not know the values of the alternatives. And, due to complexity, she is only able to imperfectly evaluate them. Specifically, each alternative x is characterized by its true value v_x as well as its inherent complexity σ_x. The latter can be thought of as any dimension of the alternative that complicates the ability of the decision maker to assess its value. Knowing neither v_x nor σ_x, the decision maker relies on her skill S (i.e., cognitive ability, expertise, etc.) to break down complexity and evaluate x. The outcome of the evaluation process is a noisy estimate $u_x = v_x + \epsilon_x$, where ϵ_x denotes a normally distributed error term with mean zero and standard deviation $\max\{0, \sigma_x - S\}$.

Object complexity in this formulation affects choice behavior through σ. If $S \geq \sigma_x$, the decision maker evaluates x accurately. Otherwise, she only obtains an estimate of v_x, whose accuracy depends on σ_x. In other words, object complexity manifests itself as a greater amount of noise in the evaluation of individual alternatives; although all evaluations are correct on average.

Composition complexity operates at the set level and depends only on the true values of the available alternatives. We say choice set A has higher composition complexity than choice set B (with the same support of values) if it is possible to obtain A from B by a series of any number of the following "basic" operations: *(i)* removing an optimal alternative, *(ii)* adding a suboptimal alternative, or *(iii)* replacing a suboptimal alternative with another one that is closer in value to being optimal. Because set A may be larger or smaller than set B, composition complexity is related to but conceptually distinct from elementary notions of choice overload, according to which larger choice sets are *per se* more complex (see Iyengar and Lepper 2000).

The second ingredient of our theory incorporates the noisy evaluation process above into a model of choice. Here, we build on Simon's (1955) seminal work on bounded rationality and satisficing, which he motivated, in part, by appealing to chess. According to Simon (1972, p. 166), "chess players do not consider all possible strategies and pick the best, but generate and examine a rather small number, making a choice as soon as they discover one that they regard as satisfactory." In our model, the decision maker has in mind an aspiration level, T, that she wishes to exceed. She lists all available alternatives in random order, and sequentially conducts noisy evaluations of each of them until she encounters one for which $u_x \geq T$. This is the alternative she chooses.

After developing the model's predictions, we empirically test them in a novel data set on chess endgames. The data come from two sources. The first one is `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.[1] We use this information to reconstruct all choice sets that players faced in endgames with six or fewer chess pieces. Our analysis focuses on the choices of nearly a quarter million highly

[1] 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.

experienced users, ranging from seasoned hobbyists to the best players in the world. In total, we observe about 227 million choices from sets with approximately 4.6 billion alternatives.

We augment these data with information from the Syzygy and Nalimov endgame tablebases (Nalimov et al. 2000; de Man 2013). These databases record the type of every available move in all non-trivial endgame positions with up to six pieces. As in Zermelo's Theorem, each move is assigned one of three types. A "winning" move allows the player to subsequently force a win. A "losing" move enables her opponent to guarantee himself a win, whereas a "drawing" move lets both players force a draw. The relevant classifications are based on exhaustive brute-force computational methods, assuming subsequent best responses.

We also query the Nalimov database for information on depth to mate (DTM), our measure of object complexity. DTM is a metric of how fast a checkmate can be forced. More formally, the DTM of a winning move is the length of the equilibrium path when the moving player wishes to minimize the number of moves to mate while her opponent tries to maximize it.[2] DTM can thus be thought of as the "minimax length" of equilibrium play when both players entertain the possibility that, at any stage, the dominant player (who has a winning strategy) may tremble and pick a drawing or losing move with some small probability. By expediting the end of the game, the dominant player tries to minimize this risk, whereas her opponent attempts to maximize it by holding out as long as possible. The outcome is an equilibrium path whose length is equal to DTM.

In the raw data, greater complexity coincides with a higher frequency of mistakes. A player makes a mistake when she chooses a drawing or losing move in the presence of a winning alternative. Such mistakes depend nontrivially on the composition of the choice set. For example, when the share of winning moves is less than 10%, an increase in the share of drawing moves from less than 10% to more than 90% (with a respective decrease in the fraction of losing moves) is associated with a more than 15 percentage points (p.p.) higher frequency of mistakes. Mistakes also increase with DTM. For instance, when a player can force checkmate within fifteen moves or less, mistake are relatively rare—about 3%. If, however, the shortest DTM among winning moves exceeds sixty, mistake frequencies increase to more than 20%.

Going beyond simple descriptive statistics, we find both object and composition complexity affect choice behavior as predicted by the theory. At the object level, we observe that, within the same the choice set, winning moves with a low DTM are more often selected than their counterparts with higher object complexity. The reverse holds for losing moves. Furthermore, comparing the choice frequencies of two moves of the same type across otherwise similar

[2] An analogous definition applies to losing moves. We refrain from measuring object complexity for drawing moves. See the discussion in Section 3 for details.

choice sets, we find that a winning (losing) move becomes less (more) likely to be chosen as its DTM increases. To see the economic significance of this result, consider a two-node increase in the DTM of a winning move. According to our estimates, such an increase corresponds to 1.0–2.3 p.p. decrease in the choice frequency of the corresponding move. This translates to a 10%–24% reduction relative to the mean choice frequency of winning moves.

At the choice-set level, what matters for the quality of decision-making is not necessarily the number of available alternatives but their "mix." Specifically, we show that the frequency of mistakes increases when a winning move is removed from the choice set, a drawing or losing move is added, or a losing move is replaced with a drawing alternative. To illustrate how compositional changes affect behavior, consider replacing a single losing move in the choice set with a drawing alternative. Based on our estimates, such a change increases the frequency of mistakes on average by about .84 p.p.—or 14% relative to the mean frequency of mistakes.

Finally, we probe the theory's predictions about object complexity and skill. Given that our data contain approximately 3.4 million moves by chess masters, of which nearly 350,000 were made by grandmasters, we can ask whether highly skilled players are less likely to make mistakes when object complexity is large. As expected, chess masters err significantly less than other experienced players, and, taking the point estimates at face value, grandmasters are even less likely to do so than other masters. Consistent with the theory, we only detect consistently large differences between titled and untitled players when the minimal DTM among winning moves is relatively high.

The analysis in this paper complements two literatures. The first is 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). A high-level takeaway is that computational constraints can greatly affect both individual and strategic outcomes.

Our contribution relative to extant theoretical work is twofold. First, we develop a theory that highlights object and composition complexity as two distinct sources of complexity. While object complexity relies on the idea that agents face cognitive limitations, composition complexity is conceptually closer to psychological accounts of choice overload (see, e.g., Iyengar and Lepper 2000; Schwartz 2004; Chernev et al. 2015). Second, we provide evidence on the empirical relevance of complexity considerations in general, and object and composition complexity in particular.

Our findings also complement a growing experimental literature on complexity in decision-making (see, e.g., Huck and Weizsäcker 1999; Gabaix et al. 2006; Abeler and Jäger 2015;

Bossaerts and Murawski 2017; Enke and Graeber 2021). Rubinstein (2007, 2016), for instance, shows that decisions that require explicit cognitive reasoning take far longer to complete than those that do not. Caplin et al. (2011) provide evidence that individuals rely on satisficing in choice environments in which evaluating each option takes time and effort. Oprea (2020) develops a revealed-preference methodology to measure the cost of complexity. He finds subjects are willing to pay significant amounts to avoid tasks that are inherently complex. Overall, laboratory experiments support the idea that complexity can affect decision-making.

Outside of the laboratory, tests of fundamental decision-theoretic concepts remain rare.[3] As Chiappori et al. (2002) note, nonexperimental settings may be 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 in real-world environments.[4]

Chess endgames provide an almost ideal empirical setting to study this question. Beyond yielding observable variation in complexity and admitting an ordinal measure of alternatives' value, chess possesses at least four 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. Yet, evaluating different moves often strains the bounds of human cognition. Second, data on chess games are abundant. We therefore have enough statistical power to test even subtle theoretical predictions. Third, there are large differences in chess players' skill, which allows us to analyze how complexity and skill interact. Fourth, 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.[5]

Finally, scholars have long been intrigued by chess as a game that is theoretically trivial but practically intractable. Chess is a finite, two-player, zero-sum game with perfect information (von Neumann 1928).[6] It is well known that, in any board configuration, either White has

[3]A related literature asks whether some of the basic tenets of game theory are consistent with observed behavior in different real-world environments. Walker and Wooders (2001), Chiappori 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 in the U.S. Senate. On the whole, the evidence from these settings corroborates theory more closely than one might have guessed based on an abundance of negative findings from the laboratory (see, e.g., Camerer 2003 for a review).

[4]Levitt and List (2007) argue that even if individuals consistently make mistakes in the laboratory, competitive forces and experience may limit such behavior in the real world.

[5]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).

[6]There is a question as to whether chess should, in fact, be considered a finite game (see, e.g., the discussion

a winning strategy, Black possesses such a strategy, or both sides can guarantee themselves a draw (Zermelo 1913; von Neumann and Morgenstern 1944). In the words of Osborne and Rubinstein (1994, p. 6):

> The existence of such strategies suggests that chess is uninteresting because it has only one possible outcome. Nevertheless, chess remains a very popular and interesting game. [...] Even if White, for example, is shown one day to have a winning strategy, it may not be possible for a human being to implement that strategy. Thus while the abstract model of chess allows us to deduce a significant fact about the game, at the same time it omits the most important determinant of the outcome of an actual play of chess: the players' "abilities." Modeling asymmetries in abilities and in perceptions of a situation by different players is a fascinating challenge for future research, which models of "bounded rationality" have begun to tackle.

To this we only add that it is also important to *empirically* explore deviations from optimal play; and our study can be understood as taking a first step in this direction.[7]

The remainder of this paper proceeds as follows. Section 2 develops the theory that guides our empirical analysis. Section 3 explains how we measure complexity and the quality of decision-making in chess endgames. Section 4 introduces our data, followed by a brief description of some of the first-order empirical regularities in Section 5. Section 6 takes the specific predictions of our theory to the data, and Section 7 concludes. Proofs and robustness checks appear in the appendix.

2. Theory

Our analysis begins by developing a general model of complexity and choice, which we then specialize to the case of chess. After deriving the model's empirical predictions, we provide a critical discussion of its key ingredients.

2.1. *General Setup*

A decision maker (DM) chooses a single object from every choice set $A \subseteq X$, where X is the grand set of all feasible alternatives. Each object $x \in X$ is characterized by a pair (v_x, σ_x), where v_x denotes the true value of x to the DM and σ_x corresponds to its inherent complexity. The latter can be thought of as any dimension of the object that complicates the ability of the DM to recognize its value.

---

in Osborne and Rubinstein 1994, p. 100). At least since 2014, the FIDE rule book specifies that a game is *automatically* drawn if the same position has occurred five times (cf. Article 9.6.1). Under official rules, chess is thus finite.

[7]One previous paper analyzing performance in real-world chess matches is Gonzalez-Diaz and Palacios-Huerta (2016). They demonstrate that a player who is randomly selected to play White in the first of an even number of chess games against the same opponent wins about 60% of these matches. The authors attribute this striking advantage to psychological elements that affect cognitive performance. Another related set of papers uses professional chess players as subjects in standard laboratory games (e.g., Palacios-Huerta and Volij 2009; Levitt et al. 2011).

The DM knows neither v_x nor σ_x when she encounters x. Rather, she relies on her skill, S, to estimate v_x. By skill, we refer to the ability of the DM to break down complexity, i.e., her cognitive resources, experience, and so on. The outcome of the evaluation process is a noisy estimate:

$$(1) \qquad u_x = v_x + \epsilon_x,$$

where the evaluation error ϵ_x is distributed normally with mean zero and standard deviation $\max\{0, \sigma_x - S\}$. Thus, when an object's inherent complexity is low, the DM can accurately assess its true value. As object complexity increases, however, the DM needs to contend with more noise in the evaluation process. Her assessments are only accurate in expectation.

In choosing from a choice set A, the DM wishes to pick a high-value object while economizing on time and cognitive effort. To this end, she follows a satisficing procedure whereby she chooses the first "satisfactory" alternative she encounters. Formally, the DM has in mind an aspiration level T. She randomly draws an object from A and evaluates it. The object is chosen if its estimated value exceeds T. Otherwise, the DM discards the object as non-satisfactory and draws the next one (without replacement). She continues in this fashion until she finds a satisfactory object or until she exhausts all alternatives in A without making a choice. In the latter case, the DM chooses the last object that she examined.[8]

The outcome of this choice procedure can be summarized by a random choice function C that maps every choice set A and every $x \in A$ to the probability $C(x, A)$ of choosing x from the set A. Choice behavior is stochastic because object evaluations are noisy and because the evaluation order is random.[9]

Our empirical analysis studies how choice probabilities depend on *(i)* objects' inherent complexity, and *(ii)* composition complexity, which we now define. For a choice set A, let $V(A)$ be the set of values in A, i.e., $V(A) = \{v_x \mid x \in A\}$, let $\bar{v}(A)$ be the largest value in $V(A)$, and let $N(v, A)$ denote the number of objects in A that have value v. With this

[8]If the DM recalls the entire history of u's then she may choose the u-maximal object. This form of recall complicates the analysis to follow without adding substantive insights. Intuitively, this is because in the large sets we consider and assuming the threshold is not too high, exhausting all alternatives without exceeding T is a low probability event with a second-order effect on the theory's empirical predictions.

[9]Note that the evaluation order does not favor any of the alternatives. Conditional on not stopping and not having been evaluated yet, every alternative x is equally likely to be examined in the next stage of the choice process, and its conditional choice probability is determined by v_x and σ_x. If the evaluation order is allowed to favor some alternatives over others, then *any* random choice function C can be rationalized by a choice procedure that *(i)* draws the first alternative to be evaluated with its choice probability, and *(ii)* always chooses the alternative that is evaluated.

notation in hand, the composition complexity of A is:

$$(2) \qquad \Gamma(A) \equiv -N(\bar{v}(A), A) + \sum_{v \in V(A) \setminus \bar{v}(A)} \alpha(\bar{v}(A) - v) N(v, A),$$

where $\alpha(\cdot)$ is a positive and monotonically decreasing function.

The minus sign on the first term in the definition above captures the idea that adding more objects with the maximal value to a choice set simplifies the task of choosing optimally. The function $\alpha(\cdot)$ in the second term aims to capture two features of choice sets that complicate choice. Positivity of $\alpha(\cdot)$ means adding non-maximal objects increases complexity, and monotonicity implies that as the values of maximal and non-maximal objects become more similar, identifying a maximal alternative gets harder. Note that the exact cardinal specification of $\alpha(\cdot)$ determines, together with the composition of the choice set, how composition complexity changes when an object with value above $\bar{v}(A)$ is added to A.

2.2. *Application to Chess*

In chess, the set X is the set of all possible legal moves of all pieces, and a choice set corresponds to all legal moves in some board configuration. Players' choice sets can thus be directly mapped to board configurations.

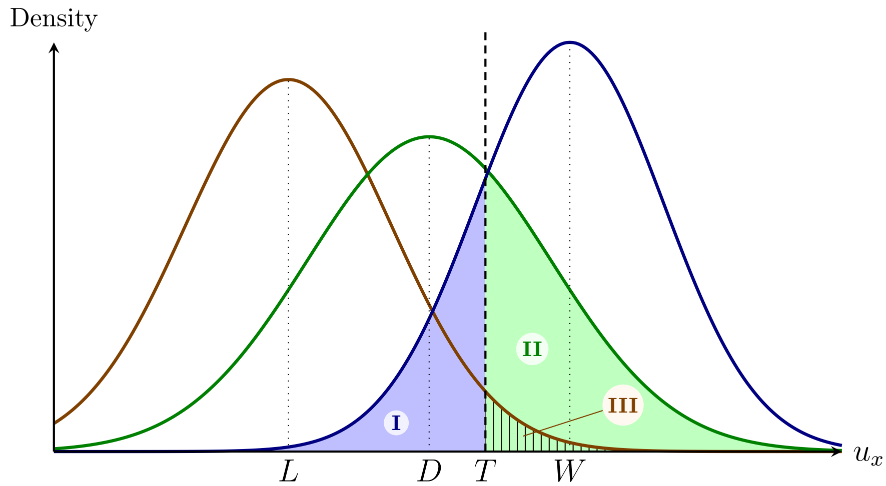
By Zermelo's Theorem, starting from any given configuration, either White can force a win, Black can force a win, or both sides can guarantee themselves a draw. It is therefore possible to associate every move in any board configuration with the ultimate outcome of the game under subsequent optimal play. A move that allows the DM to force a win yields positive payoff W, whereas a move that enables her opponent to do so produces a negative payoff $L = -W$. Moves that lead to draws generate a payoff of $D = 0$.[10] Hence, $v_x \in \{W, D, L\}$ for every move x. As a matter of terminology, we say move x in some choice set is of type W, or simply a W-move, if $v_x = W$, with analogous designations for moves that yield payoffs of D and L.

Our analysis focuses on choice sets that include at least one W- and at least one D- or L-move. In these sets, it is clear what it means to choose suboptimally. Moreover, the value of the maximal alternative remains unchanged as moves are added to the choice set. Both of these features are helpful in the empirical analysis.

We make two further assumptions. The first assumption plays a role throughout our analysis. It guarantees that the DM stops and makes a choice whenever she encounters a W-move

[10]The assumed payoff structure reflects the fact that chess is usually considered a zero-sum symmetric game. This payoff structure is not critical for our analysis. For the predictions below, it suffices to assume that $W > D > L$ and modify Assumption 2 appropriately.

Figure 1: Stochastic Evaluation of Moves



Notes: Figure illustrates the evaluation process in our model, as applied to chess.

that she can accurately evaluate.

ASSUMPTION 1: *The threshold T is between D and W.*

The second assumption is only relevant for Predictions 1 and 4 below. It states that the inherent complexity of a D-move is not "too small" relative to that of L-moves. In Section 3, we argue that this assumption is plausible in the context of chess.

ASSUMPTION 2: *In any choice set, the inherent complexity of any D-move is at least half of the inherent complexity of any L-move.*

Under these assumptions, the model predicts that the choice probabilities of individual moves are monotonically increasing in their values.

PREDICTION 1: *The choice probability of every W-move in a choice set exceeds the choice probability of every D-move in the same set, which in turn is larger than the choice probability of every L-move in the set.*

Figure 1 provides a graphical intuition for this and subsequent predictions. Area I in this figure corresponds to the probability that the DM accidentally discards a W-move, conditional on encountering one. Areas II and III correspond to the probability of mistakenly accepting a D- or an L-move instead. Given that $W > T > D$ and in light of the symmetric distribution of evaluation errors, areas I and II must each be smaller than one half. This observation, combined with random sampling of alternatives, means W-moves are accepted more frequently than D-moves. Figure 1 also makes clear why the inherent complexity of

D-moves cannot be "too small" relative to those of type L for Prediction 1 to hold. If it were, then area III may be larger than area II, which would imply that the choice probability of the corresponding L-move is larger than that of the corresponding D-move.

2.3. *Object Complexity*

The model makes two predictions regarding the connection between object complexity and choice behavior. The first prediction compares the choice probabilities of two moves of the same type and different object complexity.

PREDICTION 2: *Fix a choice set A and two moves $x, y \in A$ with $\sigma_x < \sigma_y$. If both are W-moves then the choice probability of x is larger than that of y. However, if both are D- or L-moves then the choice probability of x is smaller than that of y.*

To build intuition, recall that higher object complexity corresponds to noisier evaluations of the respective moves and thus to "wider" distributions in Figure 1. This, in turn, leads to more probability mass in areas I, II, and III. As a result, higher object complexity affects the choice probabilities of moves above and below the threshold in opposite directions.

The second prediction pertains to changes in the complexity of a single move, holding all else equal.

PREDICTION 3: *As the complexity of a W-move increases, its choice probability decreases and that of all other available moves increases. And as the complexity of an L- or D-move increases, its choice probability increases and all other choice probabilities decrease.*

The intuition behind the effect of object complexity on own choice probabilities is the same as that behind Prediction 2. As for other moves' choice probabilities, they move in the opposite direction because the evaluation order is random, and in any given order, the choice probabilities of all moves that follow the alternative whose complexity changes must offset the change in the probability of choosing that move.

An immediate consequence of Prediction 3 is that mistakes become more likely as the complexity of any move in the choice set increases. By mistake, we mean that the DM picks a D- or an L-move in the presence of at least one W-alternative. Since we model the skill of the DM as equivalent to a reduction in the complexity of moves, we also obtain the following corollary.

COROLLARY 1: *Mistake probabilities decrease in the skill of the DM whenever her skill is lower than the maximal object complexity in the choice set.*

2.4. *Composition Complexity*

To study the effect of composition complexity on behavior, we ask how "basic" changes to the composition of a choice set affect mistake probabilities. By "basic" changes, we mean adding or removing a move from the choice set, or replacing one move with another one of a different type.

Consider first the addition of a move to the choice set. Random sampling together with noisy evaluations imply such an addition leads to choice probability shifting away from all previously available moves toward the new one. Thus, if the added move is an L- or D-move, the probability of making a mistake increases, and so does composition complexity according to eq. (2). If, however, the added move is a W-move then both composition complexity and the probability of a mistake decrease. In other words, adding a move to a choice set either decreases or increases both composition complexity and mistake probabilities.

Comparing the effect of adding a D- to that of adding an L-move, monotonicity of the function $\alpha(\cdot)$ in eq. (2) implies composition complexity increases more when a D-move is added to the set. And the model predicts that mistake probabilities change in the same direction.

PREDICTION 4: *For any choice set, adding a D-move to the set increases the probability of a mistake by more than adding an L-move.*

Returning to Figure 1, the intuition for Prediction 4 is that, given any evaluation order, moving from area III to II results in a greater probability of accepting a suboptimal move.

An implication of the above discussion is that both composition complexity and mistake probabilities change in the same direction when one move in the choice set is replaced by another move of a different type. Replacing a W-move with an L- or D-alternative increases both composition complexity and mistake probabilities, because this replacement can be accomplished by first adding an L- or D-move to the set (which increases both quantities) and then removing the W-move (which, again, increases composition complexity and mistake probabilities). Replacing an L-move with a D-move also increases both composition complexity and mistake probabilities. The increase in composition complexity follows from monotonicity of $\alpha(\cdot)$, whereas the increase in mistake probabilities is an implication of Prediction 4.

2.5. *Discussion*

Our satisficing-with-evaluation errors model combines two related but conceptually distinct ideas. Drawing directly on Simon's (1955; 1972) discussion of bounded rationality and satisficing in chess, we model the DM as searching for and accepting a "good enough" rather than the best available move. According to Simon (1955, p. 100), "in actual human decision-making, alternatives are often examined sequentially. [...] If a chess player finds a forced mate for his opponent, he generally adopts this alternative without worrying whether another alternative also leads to a forced mate."[11]

We formalize this choice procedure and combine it with a noisy evaluation process that resembles the familiar random-utility framework in discrete-choice models (Luce 1959; Marschak 1960; McFadden 1974). An important interpretative difference between the evaluation process in our model and the standard discrete-choice setup is that ϵ_x in eq. (1) corresponds to pure evaluation error. Moreover, the variance of the error term depends directly on the object's inherent complexity and the DM's skill. This feature of our model is critical for it to generate predictions on how object complexity affects decision-making.

Given that the model combines satisficing and noisy evaluation, it is natural to ask whether either of these ingredients alone would suffice to generate similar predictions and thus rationalize the empirical patterns below. The answer turns out to be "no"—not without complicating the model along other dimensions.

As long as Assumption 1 holds, satisficing without noisy evaluations would imply the DM always chooses a W-move if the choice set includes one. Imperfect evaluation of the alternatives is thus a necessary ingredient to explain mistakes. Even when coupled with i.i.d. errors in the evaluation process, satisficing implies equal choice probabilities for all moves with the same payoff—in contrast to Prediction 2. Hence, to the extent that we do observe systematic differences in choice probabilities for moves of the same type in the same choice set, they must be rooted in the particulars of the evaluation process.

Conditional on the noisy evaluation process in eq. (1), would the model yield similar predictions if it featured a canonical maximization procedure rather than satisficing? By canonical maximization, we mean the following:

- The DM considers all moves in the choice set, estimates their u's, and forms, for each move, a posterior belief about its value;
- the ranking of moves' expected values according to the DM's posterior coincides with the ranking of the estimates; and
- the DM chooses the move with the highest posterior expected value.

For a such a maximization procedure, which includes both maximum likelihood estimation and Bayesian updating from an identical prior as special cases, the following result contrasts with Prediction 2:

PROPOSITION 1: *Assume canonical maximization as choice procedure, and consider any*

[11] In a similar vein, in his well-known guide to chess endgames, IM Jeremy Silman (2007, p. 6) reassures readers, "Trust me, nobody cares if you mate in three or five, as long as you succeed every time!"

choice set A with two W-moves $x, y \in A$ such that $\sigma_x < \sigma_y$. The choice probability of x is smaller than that of y.

The intuition for Proposition 1 is as follows. Although the distribution of the difference $u_y - u_x$ is symmetric with mean zero, the distribution of u_y has "thicker" tails and, most importantly, a thicker right tail than that of u_x. The thicker right tail implies a positive probability of u_y being greater than u_x conditional on both of them exceeding "large" thresholds. Under maximization, the threshold they need to exceed is the maximum among all other moves and so, in expectation, it is large.

Note, canonical maximization encompasses a large class of models. Perhaps most importantly, it includes the standard "$max_x u_x$"–approach in the discrete-choice literature. It also includes some newer models of noisy cognition, in which agents choose the alternative with the highest Bayesian posterior mean (see, e.g., Woodford 2020 for a review). By rejecting either Proposition 1 or Prediction 2, we can, under the maintained assumption of noisy evaluations, empirically discriminate between such maximizing behavior and satisficing.

3. Measurement Approach

In bringing the model to data, we need to assign a value of L, D, or W to every move in every board configuration. Although Zermelo's Theorem tells us that *any* chess move can be assigned one of these three values, this assignment is well known to be computationally infeasible in board configurations with sufficiently many pieces. Our empirical work thus focuses on configurations with up to six pieces, which comprise only a small fraction of all board configurations but have been definitively solved by computer algorithms. We rely on these solutions to assign every move in our data the correct type.[12]

Our empirical analysis examines the effect of composition and object complexity on choice probabilities. Guided by the discussion in Section 2.4, we refrain from committing to a particular specification for $\alpha(\cdot)$ in eq. (2) when studying composition complexity. Instead, we directly compare choice probabilities across choice sets that differ by one of the basic compositional changes outlined in Section 2.4. Because we can determine the type of each available endgame move, implementing these comparisons is straightforward.

As explained in the introduction, our proxy for the object complexity of W- and L-moves is depth to mate (DTM). A move's DTM corresponds to the "minimax length" of equilibrium

[12] Although board configurations with seven pieces have, in principle, been solved as well, we do not use them in our analysis, because doing so would be impractical. Specifically, the open-source Syzygy tablebases only contain information on depth to zero (DTZ) rather than depth to mate (DTM). While DTZ and DTM are related, they are not the same, and DTM is a critical component of our measurement approach. The commercially available seven-piece Lomonosov tablebases do contain information on DTM but require about 140TB of storage. They are thus too large to be usable in most computing environments.

play following that move. By "minimax length," we mean the number of nodes along the equilibrium path on which the dominant player (who has a winning strategy) minimizes the number of moves to mate while her opponent resists as long as possible.[13] DTM is thus an intuitive but coarse measure of how deep the DM needs to look down the game tree—and therefore how cognitively difficult it is—to evaluate the current move. If the DM is not skilled enough to explore the game tree to sufficient depth, then she cannot assess the move's value with certainty. Her assessment remains noisy, and the amount of noise increases in the number of nodes she was unable to reach.

Measuring the complexity of D-moves is more challenging because, to the best of our knowledge, there exists no general algorithm to even identify D-moves by means other than elimination, which is how computer algorithms proceed. Computer algorithms recursively classify all moves, and if they cannot classify the move as an L- or W-move, they conclude it must be one of type D (see Section 4 for a high-level description of these algorithms). Such a classification-by-elimination approach does not lend itself to a natural measure of complexity for D-moves—other than suggesting it may be large relative to that of other alternatives in the choice set.

For this reason, we refrain from quantifying object complexity for D-alternatives, and restrict our empirical tests of Predictions 2 and 3 to W- and L-moves only. As far as Predictions 1 and 4 are concerned, they rely on Assumption 2, which intuitively says identifying a D-move with certainty is at least half as difficult as identifying L-moves. Per the discussion above, this assumption may be reasonable. It is also reasonable to interpret our empirical tests of Predictions 1 and 4 as joint tests of the respective prediction and Assumption 2.

4. Data

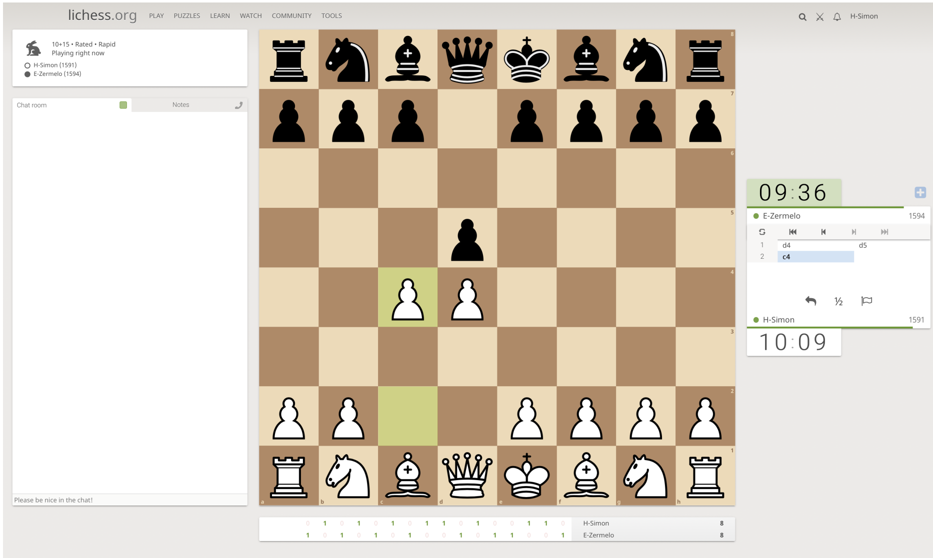
Our data on endgame moves come from `lichess.org`, one of the most popular online chess platforms. Funded by donations, Lichess is ad free and allows anyone to play live chess games at no cost through a high-quality graphical user interface.[14] Figure 2 shows a screenshot of a typical game.

Although Lichess offers a choice between many different time limits, the vast majority of games that are actually hosted on the platform can be broadly classified as "speed chess," i.e., games in which players have significantly less time to complete their moves than under classical time controls. Lichess further distinguishes between casual and rated games. The latter

[13] By this definition, a W-move's DTM is always an odd number, whereas that of L-moves is even. Sometimes DTM is expressed in a different but substantively equivalent way by ignoring the opponent's moves. There is also some arbitrariness in whether to count the current move. Thinking about moves as decision nodes in a game tree, we prefer to count all of them, including the current move.

[14] According to internet traffic data by `Alexa.com`, most users are located in North America, Europe, and India.

Figure 2: Screenshot of 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.

determine player ratings and are therefore only available to registered users. In a nutshell, a player's strength rating increases (decreases) whenever she wins (loses) a rated game, and it increases (decreases) by more the stronger (weaker) her opponent was.[15] Anecdotal evidence from online messaging boards suggests users often care intensely about their rating. Since high ratings tend to be a source of pride among chess players, Lichess has a strict policy against computer-assisted play. Enforcement of this policy relies on a variety of methods, including community reporting of suspected offenders and automatic detection algorithms.

We have data on the universe of rated games between human players from January 2013 through August 2020. The available information includes players' usernames, ratings and real-world titles, if any (i.e., candidate master, national master, international master, grandmaster, etc.), the date and start time of the game, its outcome, the precise sequence of moves, and the clock reading at the end of every move.[16] We can therefore reconstruct all choice sets a player faced, which moves she chose, how long she took to make a decision, and how much time she had left on the clock when she made a choice.

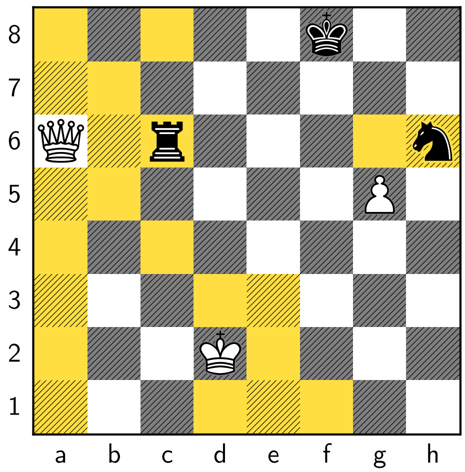
Our analysis focuses on choice sets that correspond to board configurations with six or fewer pieces. For these choice sets, the true value, or type, of essentially any legal move is known with certainty, as is the DTM of W- and L-moves.[17] The relevant information is stored in so-

[15]More specifically, in calculating rating scores, Lichess implements the Glicko-2 algorithm, which extends the better-known ELO rating system by incorporating a measure of uncertainty in approximately Bayesian fashion (Glickman 1999, 2000).

[16]Lichess verifies that users do, indeed, hold the real-life titles they claim.

[17]The only exceptions are positions with castling rights and configurations in which a lone king faces five

Figure 3: Example of an Endgame Table



Move	Evaluation	Move	Evaluation
♕c6	W in 19	♕b7	D
♕a8	W in 21	♙h6	L in 34
♕a3	W in 47	♕c4	L in 34
♕a7	D	♕c8	L in 34
♕a5	D	♔e2	L in 32
♕a4	D	♔e3	L in 32
♕a2	D	♔d3	L in 32
♕a1	D	♕b6	L in 32
♕b5	D	♔e1	L in 30
♕d3	D	♔d1	L in 28
♕e2	D	♙g6	L in 28
♕f1	D		

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

called endgame tables, some of which are in the public domain. Figure 3 provides a concrete example of endgame databases' content. The left panel depicts the board configuration that is being considered, with the data for each available legal move shown on the right. The assessment of a move consists of two components: its type (i.e., W, D, or L), and, for W- and L-moves, their DTM.

We augment the Lichess data with these two key pieces of information, which we retrieve by running several billion queries against the Syzygy and Nalimov endgame databases (Nalimov et al. 2000; de Man 2013).[18] For every legal move in every choice set, we record its type and DTM, if applicable. And for each decision problem, we record which of the feasible moves was chosen, as well as basic information on the composition of the choice set, such as the number of legal moves of each type, and the distribution of DTM values.

To provide some insight into how the information in endgame databases is calculated, consider the following high-level description of the underlying algorithm.[19] The algorithm begins by constructing an exhaustive list of all possible (up to symmetry) legal board configurations

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 analysis excludes all board configurations for which information on DTM is not available.

[18] As a technical side note, although the Syzygy tablebases are available for board configurations with seven pieces, they do not contain information on DTM. In contrast to the Nalimov tables, they do, however, take into account the fifty-move stalemate rule. In rare instances, the fifty-move rule matters for correctly determining whether one player can unilaterally invoke a draw. We therefore rely on win-draw-loss information from the Syzygy database, while information on DTM comes from the Nalimov database, which only contains board configurations with at most six pieces.

[19] For additional information, see, e.g., Thompson (1986).

with three chess pieces.[20] Every configuration is examined, and the ones in which the player to move is in checkmate are stored as "mated in 0." Next, all configurations with the other side to move are evaluated. If one of them can reach a configuration that has previously been determined to be "mated in 0" by executing a legal move, then it is stored as "mate in 1." To find the set of configurations that are "mated in 2," the algorithm looks for configurations from which *all* possible legal moves lead to "mate in 1" configurations; and to determine configurations that are "mate in 3," it subsequently checks for configurations from which it is possible to directly reach a configuration that is known to be "mated in 2." Proceeding recursively, a configuration is classified as "mated in l" if and only if it is impossible to avoid moving to configurations that are "mate in w," with $w = l - 1$ for at least one available legal move and $w \leq l - 1$ for all others. By contrast, a configuration is marked as "mate in w" if it is possible to move to another one that is "mated in $w - 1$."[21] This procedure continues until no further progress at classifying configurations is made, at which point all remaining configurations with three chess pieces are designated as "drawn." Essentially the same algorithm is next applied to board configurations with four pieces, then five, and then six—the only important difference being that the procedure must now also account for the possibility of reaching configurations with fewer pieces. The end result is a database in which *board configurations* are classified as either "drawn," "mated in l," or "mate in w."

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

Our final sample contains nearly 227 million decision problems with a total of over 4.6 billion legal moves. There are four distinct sources of selection into this sample. First, we restrict attention to choice sets that contain at least one W- and at least one D- or L-move. We impose this restriction for expositional convenience. When a board configuration admits at least one W-move, any choice of a D- or L-alternative is unambiguously a mistake. In addition, we can distinguish "small" mistakes from "large" ones. The former correspond to picking a D-move in the presence of one or more of type W, while the latter consist of choosing an L-move instead.

[20] It is not necessary to consider configurations with two lone kings, as these are automatically drawn.

[21] In checking reachable board configurations it is necessary to account for the possibility of converting pawns into other pieces.

A disadvantage of this restriction is that the choice sets in our final sample are, by construction, not representative. If one player has one or more W-moves, then the game is typically tilted in her favor. Moreover, if she chooses one of the W-moves, her opponent's subsequent choice set contains *only* L-moves and is therefore not in the sample. Because the opponent has no scope to make a mistake when deciding only among L-moves, our sample will not include any of his subsequent decisions until the dominant player blunders. As a consequence, the choice sets in our sample contain more W- and fewer L-moves than the average choice set. They do contain a non-negligible number of D- and L-moves, however, because we are focusing on positions in which the dominant player can still make a mistake and thereby lose her advantage. In our view, these are the most interesting choice sets in the data.[22]

The second and related source of selection pertains to how often different individuals reach 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. In what follows, we address this issue in two ways. First, whenever appropriate, we control for player fixed effects. Second, we reweight all observations by the inverse number of decision problems that we observe for a particular user. This ensures that all players receive equal weight in our analysis. The results below should hence be interpreted as referring to a typical decision by the average player in our sample.[23]

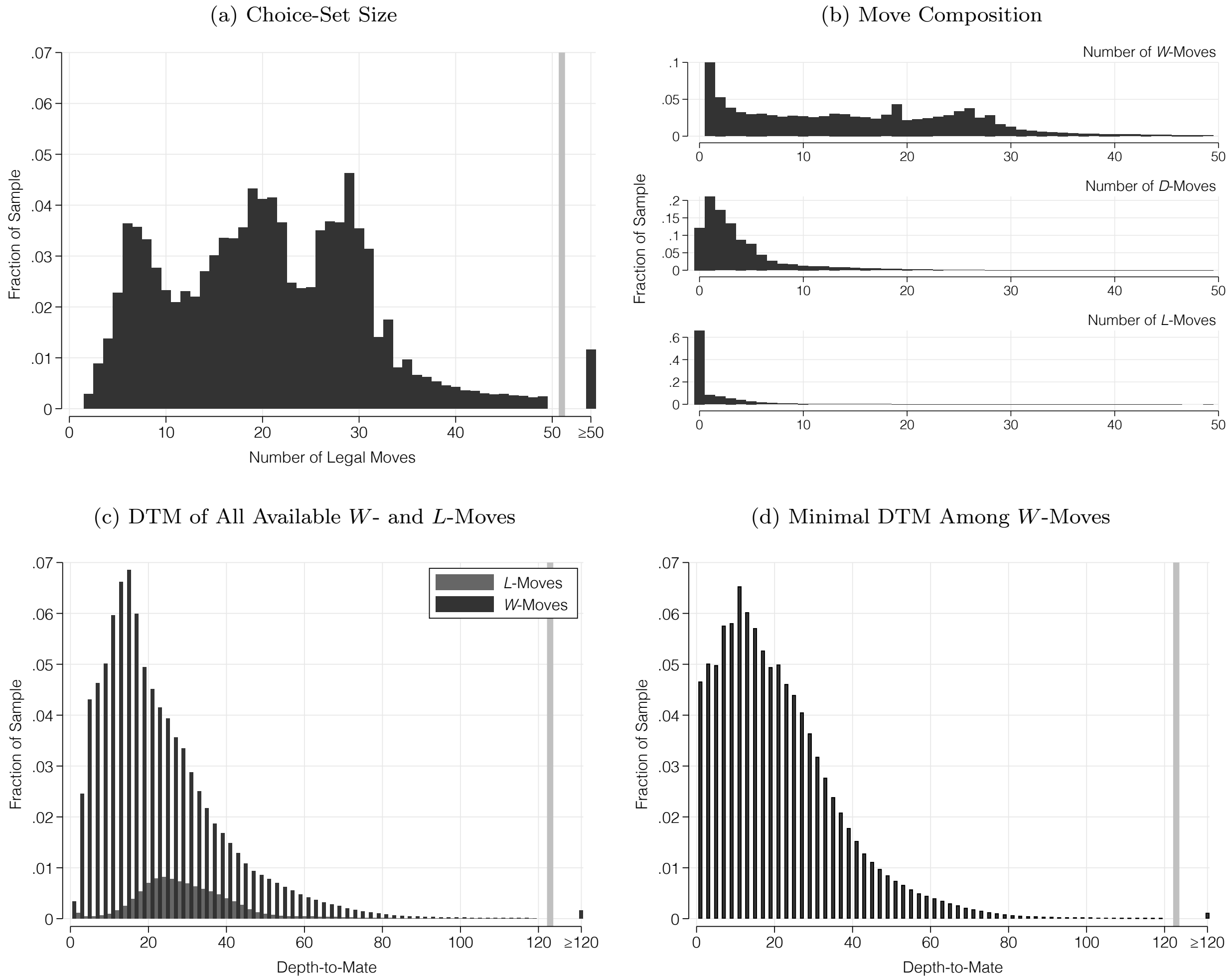
Third, to minimize the risk that our findings are due to an unfamiliar setting or a lack of experience with similar decision problems, we exclude the first one thousand endgame moves of every player in the data. We are thus left with highly experienced chess players who are very familiar with the task at hand. By design, they are not representative of the whole population. Nonetheless, if even experts' choice quality deteriorates as complexity increases, then we may expect a similar, or perhaps even stronger, effect of complexity on less experienced DMs.

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 *This Week in Chess* (TWIC), which covers "all the latest news and games from international chess." The main disadvantage of this alternative data set is that there is less variation in the skill of players, and that it is more

[22]To be clear, our sample is *not* selected on the outcome, i.e., whether or not a player makes a mistake. It is selected based on whether the DM has the *opportunity* to make a mistake in a particular decision problem.

[23]For robustness checks in which every individual observation receives the same weight, see Appendix C.1. The results therein are qualitatively similar to those below.

Figure 4: Distributions of Key Variables



Notes: Panel (a) presents the distribution of the number of available moves, with panel (b) distinguishing between W-, D-, and L-moves. Panel (c) shows the distribution of DTM among W- and L-moves, and panel (d) depicts the distribution of the smallest DTM among all W-moves in the choice set.

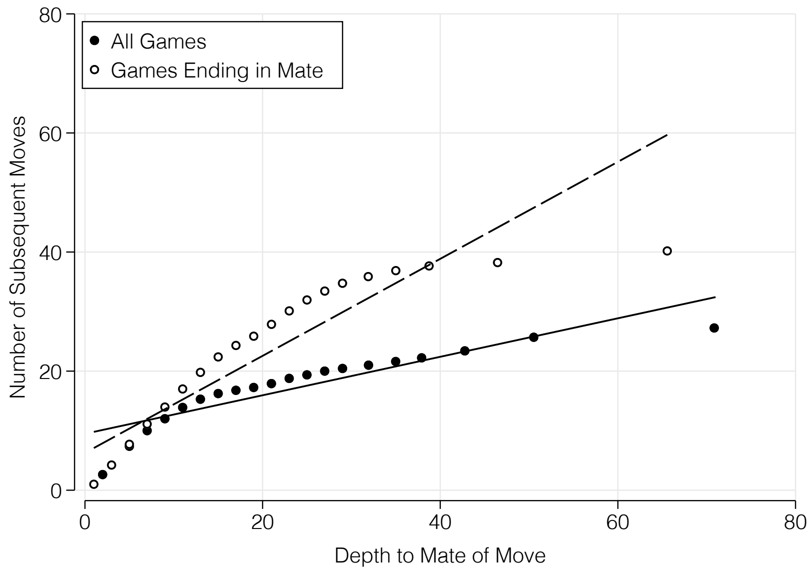
than two orders of magnitude smaller. These limitations notwithstanding, the TWIC data yield qualitatively similar conclusions (cf. Appendix D).

5. A First Look at the Data

5.1. *Descriptive Statistics*

We now describe some of the basic features of the data, starting with the distributions of key variables. The upper two panels in Figure 4 display histograms for the total number of available legal moves in the DM's choice set (left) and for the number of W-, D-, and L-moves (right). Choice sets contain, on average, about 20.6 moves, of which 15.5 are W-moves. Important for our purposes, there is significant variation in choice-set size ranging from less than a handful of moves to more than thirty, or even fifty alternatives. Most, but

Figure 5: Greater Depth to Mate is Associated with Longer Realized Game Paths



Notes: Figure shows binned scatter plots of the relationship between chosen moves' DTM (x-axis) and the total number of subsequent moves in the same game (y-axis). Solid circles correspond to within-bin averages for all games, while hollow circles restrict attention to games in which one of the players is eventually mated. The solid and dashed lines show the lines of best fit in the respective underlying move-level data.

by no means all, of this variation is due to differences in the availability of W-moves. The data therefore allow us to study choices from small, medium, and large sets, with various degrees of composition complexity.

The lower two panels of Figure 4 demonstrate that there is also significant variation in our measure of object complexity. The lower-left panel depicts a histogram of the DTM of all available W- and L-moves. The modal DTM is 15 among the former and 24 among the latter. Given the size of the sample, there are millions of legal moves with a DTM of five or less, and millions of moves with DTM of more than fifty. The lower-right panel plots the distribution of the minimal DTM among W-moves at the choice-set level. The histogram in this panel indicates that the data include choice sets in which evaluating some of the moves is straightforward, others where accurately classifying optimal moves likely exceeds the bounds of human cognition, and a great range of intermediate cases.

To verify that moves with greater DTM are, indeed, associated with longer subgames, Figure 5 presents a binscatter plot of the relationship between moves' (theoretical) DTM and the total number of subsequent moves in the same game. If the dominant player succeeds in mating her opponent but does not take the shortest path to victory, the total number of subsequent moves may exceed the initial move's DTM. If, however, the losing player does not hold out as long as possible, there will be fewer subsequent moves than implied by DTM.

If the dominant player eventually blunders and thus fails to mate her opponent, the actual path may be shorter or longer than the DTM of the original move. Importantly for our purposes, we observe a monotonically increasing relationship between moves' DTM and the total number of subsequent moves, which suggests higher-DTM moves are, indeed, associated with more complex subgames.

Some of our analysis focuses on the quality of choice as reflected by mistakes. As shown in Table 1, about 6% of observed moves are mistakes, with "small" mistakes (i.e., choosing a D-move) occurring roughly five times as frequently as "large" ones (i.e., picking an L-alternative). Given that, on average, W-moves account for about 75% of all available moves in a choice set, mistakes occur at nearly one quarter the rate one would expect if DMs were choosing at random. The descriptive statistics in Table 1 thus provide evidence of significant skill among the players in our data. Yet, the raw data also imply that mistakes occur with a certain regularity. They are not rare events.

Moreover, mistakes tend to be consequential. A player whose current move is a mistake is about 43 p.p.—or roughly 58%—less likely to ultimately win the game than one who chooses a W-move. At the same time, the probability of a loss more than doubles.[24]

Table 1 further shows that, in this environment, decisions are made quickly. On average, the experts in our data take 1.7 seconds to execute a move. Only about 5% of moves take longer than 5 seconds. As most games hosted on Lichess are best thought of as "speed chess," the rapid pace of decision-making is not surprising. Given that the timing of decisions is endogenous, we do not control for it in our main analysis. In robustness checks, we have estimated alternative specifications that do control for how much time a player had left on the clock. We have also estimated specifications that restrict attention to chess games with classical time controls, in which players face significantly less time pressure. These robustness checks yield results that are qualitatively equivalent to those below (cf. Appendix C.2).

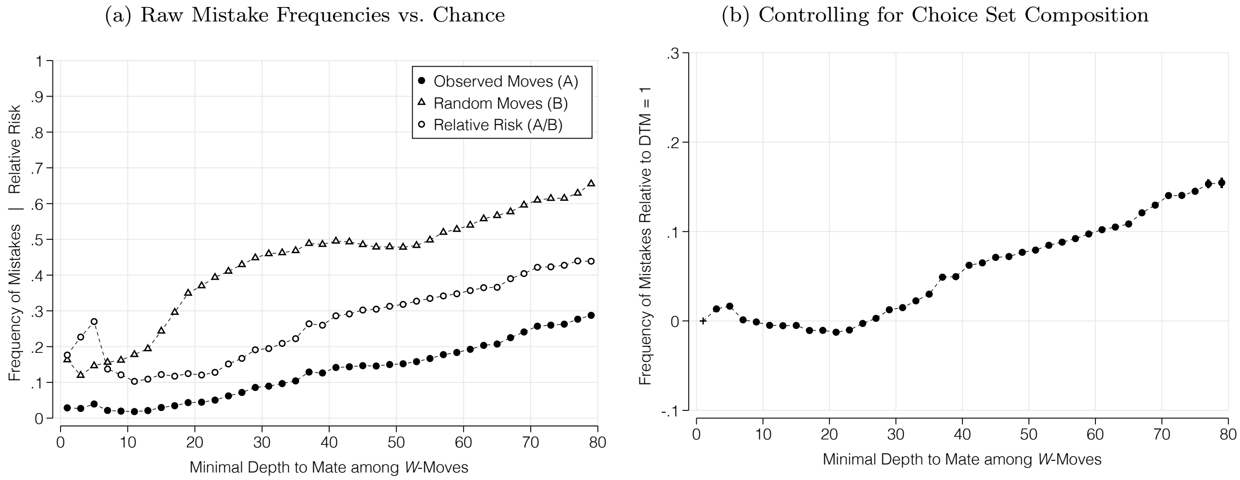
[24]Mistakes do not always result in forgone wins, because the player's opponent may subsequently also choose a suboptimal move. Similarly, due to the potential for mistakes in future moves, choosing a W-move in a particular instance does not automatically mean the player will ultimately win the game.

Table 1: Descriptive 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
DTM:							
W-Moves	23.15	16.35	11	19	31	55	3,457,878,398
L-Moves	30.49	14.86	22	28	36	54	296,522,573
B. Choice-Set Characteristics							
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
DTM of W-Moves:							
Shortest	20.91	16.18	9	17	29	51	226,955,095
Median	25.81	17.34	13	22	33	59	226,955,095
Longest	31.62	20.05	17	27	39	69	226,955,095
DTM of L-Moves:							
Shortest	26.95	11.39	20	26	34	44	77,045,730
Median	30.19	12.50	22	28	36	48	77,045,730
Longest	34.27	16.31	24	32	40	60	77,045,730
C. Outcomes							
Mistakes:							
Any Type of Error	0.06	0.24					226,955,095
Small Mistake	0.05	0.23					226,955,095
Large Mistake	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 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 receive equal weight. The number of observations related to the timing of moves is smaller because the raw data do not include this information for games that were played prior to April 2017.

Figure 6: Mistakes Increase in the Minimal DTM among W-Moves



Notes: Conditioning on the minimal DTM among all W-moves in the choice set, panel (a) compares the observed mean rate of mistakes with their expected frequency if players were to choose at random. Panel (b) shows point estimates and 95%-confidence intervals for γ_i in eq. (3). Confidence intervals account for two-way clustering by player and game. Since $DTM = 1$ is the omitted category, all coefficients denote the difference in mistake rates relative to the shortest winning path being as short as possible.

5.2. *Preliminary Evidence on Complexity and Mistakes*

Figures 6 and 7 present raw data hinting at a connection between our two notions of complexity and choice behavior. Although these patterns are consistent with our theory, this part of the analysis should not be interpreted as testing specific model predictions. We defer such tests to Section 6.

Figure 6 studies how mistake frequencies change as W-moves in the choice set become harder to evaluate, i.e., as the minimal DTM among available W-moves increases. The panel on the left depicts the bivariate relationship in the raw data as well as the mechanical baseline risk of making a mistake when choosing at random. Based on the raw data alone it is unclear whether mistake rates increase because W-moves are becoming inherently more complex or because of changes to the composition of the choice set.

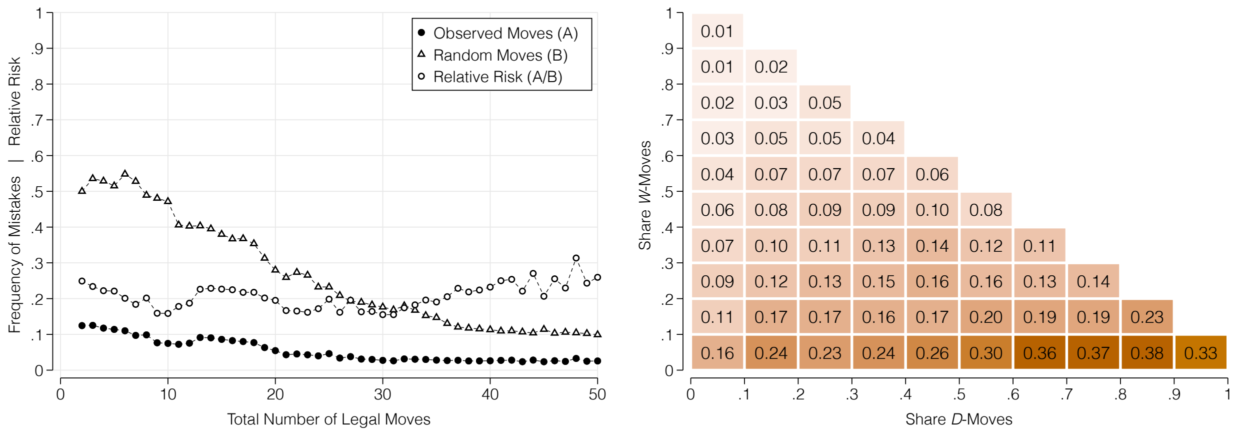
The right panel in Figure 6 addresses this issue econometrically. The estimates therein correspond to $\{\gamma_i\}$ in the following linear probability model:

$$(3) \qquad Error_d = \sum_{i=2}^{80} \gamma_i \lfloor DTM \rfloor_{i,d} + \theta_d^{WDL} + \mu_p + \varepsilon_d,$$

where $Error_d$ is an indicator variable for whether decision d of player p is a mistake, and μ_p denotes a player fixed effect. Each $\lfloor DTM \rfloor_{i,d}$ is an indicator for whether the minimal DTM among W-moves in the respective choice set is equal to i. We control for the composition of the choice set by including θ_d^{WDL}, a fixed effect for every combination of the number of

Figure 7: Mistakes as Function of the Number and Type of Available Moves

(a) Mistakes vs. Choice-Set Size (b) Mistake Frequencies and Choice-Set Composition



Notes: Panel (a) compares observed mistake rates for choice sets of a particular size with the frequency of mistakes that would be expected if players chose a move at random. Panel (b) shows mistake rates for different shares of W-, D-, and L-moves.

available W-, D-, and L-moves. The specification above thus accounts nonparametrically for any (potentially mechanical) relationship between mistakes and the composition of the choice set.

Reassuringly, the pattern in the right panel is broadly similar to that in the left one. This suggests mistakes depend on how difficult it is to evaluate W-moves, even after accounting for their availability. Interestingly, we see little change in mistake frequencies for minimal DTM values below twenty. For higher values, however, the gradient is relatively steep. This observation is consistent with the view that even experienced chess players face cognitive limitations, which begin to constrain their decision-making at higher levels of object complexity.

Moving to complexity at the set level, the left panel of Figure 7 depicts the relationship between mistakes and the size of the choice set. It also contrasts observed mistake frequencies with the probability of making a mistake when choosing a move at random. There is a clear negative association between the frequency of mistakes and choice-set size. However, because larger choice sets turn out to contain, on average, a greater share of W-moves, the observed decline in the frequency of mistakes might be mechanical. Measured as a fraction of the mechanical baseline risk of making a mistake when choosing at random, error rates remain nearly constant over approximately 85% of our sample. Only for choice sets with more than thirty moves do we observe some increase in relative risk. Taken at face value, the evidence in this panel is difficult to reconcile with elementary notions of choice overload, according to which larger choice sets are *per se* more complex.

What appears to affect mistake frequencies is the "mix" of moves in the choice set. The right panel of Figure 7 illustrates this point by examining how the rate of mistakes varies with the relative frequency of each type of move in the set. The entry on the bottom left, for instance, indicates that when the fraction of W- and D-moves ranges from zero to ten percent each, about 16% of choices end up being mistakes. Comparing rows within the same column, we see mistakes become more frequent as W-moves become relatively more scarce. And reading across columns of the same row, mistakes also tend to increase as D-moves become more frequent. In the bottom row, for instance, the frequency of mistakes increases from 16% to more than 30%. Since comparisons within the same row hold the fraction of W-moves—and hence the mechanical probability of making a mistake—approximately fixed, the raw data suggest the quality of decision-making correlates with the *composition* of the remaining, inferior alternatives.

6. Tests of Model Predictions

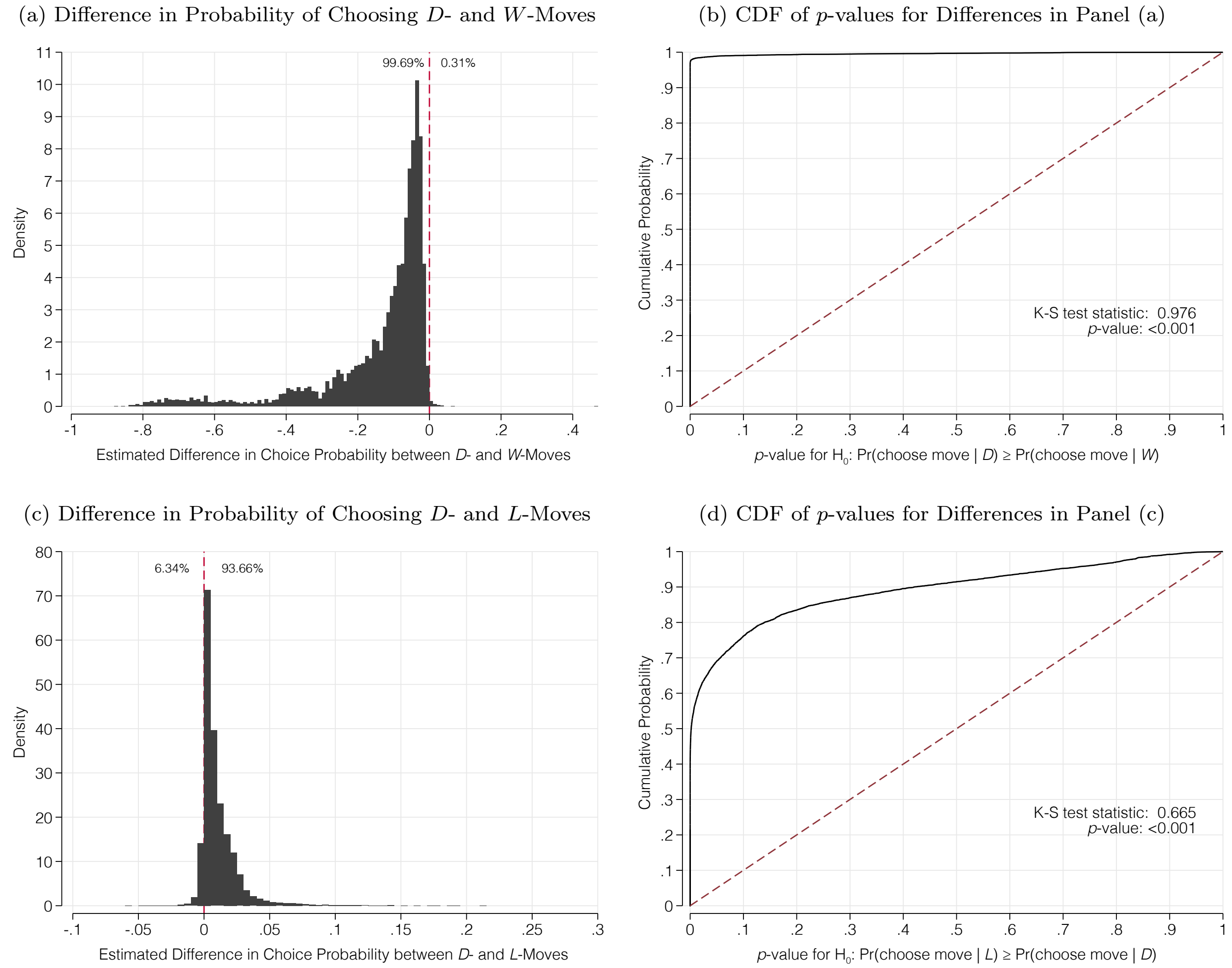
6.1. *Ranking of Choice Probabilities*

Having provided a first look at the broad patterns in the data, we now proceed to test the specific predictions of our model. Prediction 1 states that, within every choice set, choice probabilities increase in moves' values, i.e., W-moves are more likely to be chosen than D-moves, which in turn are more likely to be chosen than L-alternatives. The summary statistics in Table 1 indicate this is true in the aggregate. To test Prediction 1 at a more granular level, we partition the data into "classes" of choice sets so that two sets are in the same class if and only if they have the same number of W-, D-, and L-moves. In the data, we observe more than a hundred choices from 9,701 of such classes. For each of them, we calculate the empirical frequency with which the available legal moves of a particular type are chosen, and plot the observed differences in Figure 8.

The top-left panel shows a histogram of differences in choice probability between D- and W-moves—one estimate for each class—and the bottom-left panel does so for the difference between D- and L-alternatives. To the right of each histogram, we show the CDF of the associated p-values. Under the null hypothesis that choice probabilities do not depend on moves' types, the p-values of the estimated differences should be approximately uniformly distributed over the unit interval.

This is clearly not the case. In almost all classes, different types of moves are chosen with frequencies that are statistically distinguishable at conventional significance levels. More importantly, the signs of the observed differences are generally as predicted by the model. W-moves are more likely to be chosen than D-moves in more than 99% of classes, and D-moves are more likely to be chosen than L-moves in more than 93% of classes. The evidence in Figure 8 thus supports Prediction 1.

Figure 8: Comparing Choice Probabilities of W-, D-, and L-Moves

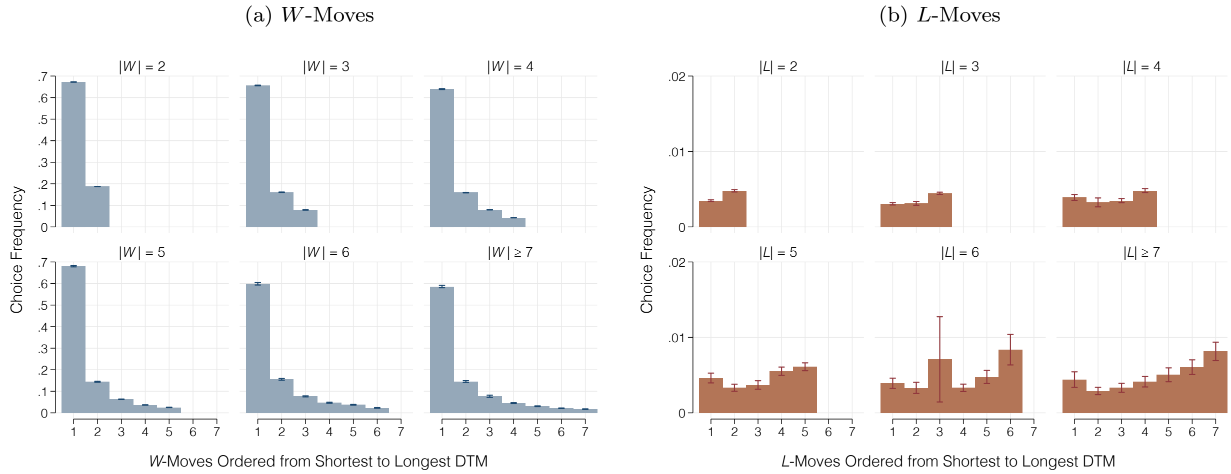


Notes: Figure compares choice probabilities of individual W-, D-, and L-moves. Panel (a) plots the distribution of estimated mean differences in the choice probabilities of D- and W-moves in choice sets with identical composition, conditional on observing at least 100 choices from sets with that makeup. Negative estimates imply that, in the respective class of choice sets, D-moves are, on average, less frequently chosen than W-alternatives. Panel (b) depicts the empirical cumulative distribution function of the one-sided p-values associated with the estimated differences in panel (a). It also shows results from a Kolmogorov-Smirnov test against the null hypothesis of a uniform distribution of p-values. Panel (c) mirrors panel (a) but focuses on the difference in choice probabilities between D- and L-moves. Positive estimates imply that, in the respective classes of choice sets, D-moves are, on average, more frequently chosen than L-alternatives. Panel (d) presents the empirical cumulative distribution function of p-values for the null hypothesis that, in choice sets with a particular composition, L-moves are at least as frequently chosen as D-moves. All p-values account for two-way clustering by player and game.

6.2. *Object Complexity*

Predictions 2 and 3 relate object complexity to choice probabilities. Prediction 3 is concerned with how choice probabilities change as, all else equal, evaluating a particular legal move becomes more difficult. Because every board configuration is associated with a fixed choice set, the relevant empirical comparisons are necessarily across decision problems. By contrast, Prediction 2 compares the choice probabilities of different moves of the same type within the same choice set. According to Prediction 2, we should see that the W-move with the

Figure 9: Choice Frequencies, by Relative Object Complexity



Notes: Figure compares choice frequencies of W- and L-moves according to the ordering of their DTMs relative to other moves of the same type in the same choice set. Panel (a) does so for W-moves, and panel (b) focuses on L-moves. Each plot within either panel restricts attention to choice sets that contain a particular number of W- and L-moves, respectively. Error bars correspond to 95%-confidence intervals and account for two-way clustering by player and game.

smallest DTM is chosen more frequently than the W-alternative with the second-smallest DTM value, which, in turn, is picked more often than that with the third-smallest value, and so on. For L-moves, the opposite pattern should emerge.

To test this prediction, we condition on the number of available moves of a particular type, and compute the empirical frequency with which DMs in the data pick the alternative that corresponds to the shortest, second-shortest, third-shortest, and so on path of that type. Figure 9 presents the results. For W-moves, we observe large and precisely estimated differences in choice probabilities. In line with Prediction 2, these probabilities are decreasing in moves' relative object complexity. The picture is less clear—but still broadly consistent with the theory—when it comes to L-moves, for which choice probabilities are smaller and less precisely estimated. On the whole, we conclude the evidence in Figure 9 supports Prediction 2.

To investigate Prediction 3, we restrict attention to W- and L-moves and estimate the specification

$$(4) \qquad Choose_a = \omega DTM_a \times W_a + \iota DTM_a \times L_a + \pi W_a + \phi_{A\setminus a} + \mu_p + \varepsilon_a.$$

Here, $Choose_a$ is an indicator for whether player p facing choice set A picked move a, DTM_a denotes that move's depth to mate, W_a and L_a are indicators for its type, and μ_p is a player fixed effect. We also include $\phi_{A\setminus a}$, a fixed effect for the other moves in the choice set. In constructing this fixed effect, we assume that, in line with the theory, moves can be reduced

Table 2: Choices as a Function of Depth to Mate

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.140		–0.974	–0.521		–0.333
	(0.002)		(0.002)	(0.004)		(0.003)
DTM × L-Move (÷ 100)		0.034	0.007		0.066	0.054
		(0.001)	(0.001)		(0.002)	(0.002)
W-Move (÷ 100)			63.450			41.857
			(0.069)			(0.114)
Fixed Effects:						
Composition × Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid= 0$	$\mid D \mid= 0$	$\mid D \mid= 0$
Mean of LHS Variable (%)	9.397	0.785	8.495	10.881	1.183	8.078
R^2	0.372	0.209	0.368	0.513	0.209	0.494
N	3,457,878,398	296,522,573	3,754,400,971	398,856,135	111,905,262	510,761,397

Notes: Entries are coefficients and standard errors from estimating ω, ι, and π in variants of eq. (4) by ordinary least squares. All regressions control for player fixed effects as well as a fixed effect for the combination of the number of W-, D-, and L-moves and the vector of DTMs of all other W- and L-moves in the same choice set, as explained in the text. The unit of observation in each regression is an available legal W- or L-move, and observations are reweighted so that all decision problems for a particular player and all players receive equal total weight. The sample in cols. (1)–(3) includes all board configurations in our data, whereas cols. (4)–(6) 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.

to their types and inherent complexity. Since we do not measure object complexity for D-moves, $\phi_{A\backslash a}$ conditions (only) on the vector of DTM values for W- and L-moves and the composition of the choice set, i.e., the number of W-, D-, and L-alternatives. By including $\phi_{A\backslash a}$, we aim to approximate the thought experiment in which we vary the object complexity of one particular alternative, holding composition complexity and the object complexity of all other moves fixed.

Table 2 shows results from estimating variants of the model in eq. (4). In the first three columns, we study W- and L-moves from all board configurations in our data. In the last three columns, we restrict attention to choice sets that do not contain any D-moves. The idea that our fixed effects appropriately control for the object complexity of all other moves is most plausible in the latter set of specifications. Reassuringly, the results in both sets of columns are qualitatively similar. That is, regardless of which sample we consider, we find individual W-moves are significantly *less* likely to be chosen as object complexity increases. By contrast, the choice probabilities of L-moves *increase* in their complexity.

The estimates in Table 2 are not only directionally consistent with Prediction 3, but are also economically large. According to the coefficients in cols. (1) and (4), an increase in object complexity equivalent to two extra nodes on the minimax equilibrium path decreases the choice probability of the corresponding W-move by 1.0–2.3 p.p. This decrease translates to a 10%–24% reduction relative to the mean choice probability of W-moves. The evidence in this table therefore suggests that object complexity is an empirically meaningful determinant of choice.

6.3. *Composition Complexity*

Next, we consider the theory's predictions regarding composition complexity. To this end, Table 3 presents estimates of η in the following econometric model:

$$Error_d = \eta Number\ Moves_A^t + \kappa_A^{\lfloor DTM \rfloor} + \theta_A^{-t} + \mu_p + \varepsilon_d. \tag{5}$$

In this specification, $Error_d$ is an indicator for whether decision d of player p choosing from choice set A is a mistake, and $Number\ Moves_A^t$ denotes the number of available legal moves of type $t \in \{W, D, L\}$. Although not required by the theory, we account (in a low-dimensional way) for the contribution of object complexity to mistake frequencies by including $\kappa_A^{\lfloor DTM \rfloor}$, a fixed effect for the minimal DTM among all available W-moves.[25] The fixed effect θ_A^{-t} controls in different ways for the number of moves in the same choice set that are not of type t. As we vary θ_A^{-t} across columns in Table 3, the interpretation of η changes to closely align with the different changes to choice-set composition that we discussed in Section 2.4. For example, col. (6) conditions on the choice-set size and the number of W-moves, so that $\hat{\eta}$ should be interpreted as the change in mistake probabilities when an L-move is replaced with one of type D. Finally, by including μ_p, the model in eq. (5) also controls for player fixed effects.

Col. (1) in Table 3 conditions on the number of D- and L-moves. The point estimate in this column therefore indicates that the frequency of mistakes decreases, on average, by .279 p.p. for every W-move that is added to the choice set. This corresponds to almost 5% of the mean frequency of mistakes. In analogous fashion, the coefficients in cols. (2) and (3) imply that mistake frequencies respectively increase by .810 and .247 p.p. as D- and L-moves are added to the choice set. The evidence in the first three columns of Table 3 thus shows that whether mistakes become more or less frequent as the choice set expands depends on the type of move that is being added. This pattern helps explain why the raw data yield little

[25] After conditioning on the minimal DTM among W-moves, other low-dimensional proxies for object complexity turn out to be only weakly correlated with mistakes, which suggests $\kappa_A^{\lfloor DTM \rfloor}$ captures much of the relevant variation. In Appendix Table AT.1, we show that controlling for object complexity is not necessary to obtain results that are substantively equivalent to those in Table 3. After all, Prediction 2 holds for any basic change to the choice set, regardless of objects' inherent complexity.

Table 3: Mistakes as a Function of Composition Complexity

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves (÷ 100)	–0.279			–1.242	–0.988	
	(0.001)			(0.002)	(0.003)	
Number of D-Moves (÷ 100)		0.810				0.841
		(0.002)				(0.003)
Number of L-Moves (÷ 100)			0.247			
			(0.003)			
Fixed Effects:						
Number of D- × L-Moves	Yes	No	No	No	No	No
Number of W- × L-Moves	No	Yes	No	No	No	No
Number of W- × D-Moves	No	No	Yes	No	No	No
Total Moves × Number of L-Moves	No	No	No	Yes	No	No
Total Moves × Number of D-Moves	No	No	No	No	Yes	No
Total Moves × Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.119	0.144	0.153	0.120	0.143	0.151
N	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095

Notes: Entries are coefficients and standard errors from estimating η in variants of eq. (5) by ordinary least squares. All regressions control for player fixed effects as well as the minimal DTM among W-moves in the same choice set. Controls for the number and types of other moves vary across columns. The unit of observation in each regression is a decision, and individual observations are reweighted so that all players receive equal total weight. The sample in every column includes all board configurations in our data. All estimates are scaled to correspond to the percentage point change in the probability of an error 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.

evidence of choice overload *per se*, i.e., why mistake frequencies are not uniformly higher for larger choice sets.

Perhaps more importantly, the increase in the frequency of mistakes is significantly greater when a D- rather than an L-move is added to the choice set ($p < .001$). This observation is consistent with Prediction 4 and the key idea behind composition complexity: it is easier to choose from sets in which optimal and suboptimal alternatives are less similar.

Our theory also makes predictions for changes to the choice set that involve replacing one move with another one of a different type, holding the size of the set fixed. Estimates of the respective comparative statics are shown in cols. (4)–(6). The coefficient in col. (4) refers to a W-move being replaced by a D-alternative, whereas that in col. (5) refers to a switch between W- and L-moves. The point estimate in col. (6) tells us that, on average, the probability of a mistake increases by .841 p.p. when an L-move is replaced by a D-alternative.

In sum, the evidence in Table 3 is consistent with the theory's predictions regarding composition complexity. Moreover, the point estimates for all types of compositional change are nontrivial in size—ranging from 4% to 21% of the mean mistake rate in the data. This finding suggests that our notion of composition complexity is empirically relevant.

6.4. *Skill*

In our model, an increase in players' skill S is analogous to a decrease in object complexity for all moves. Less skilled DMs, i.e., those with a smaller S, are thus more likely to make mistakes (cf. Corollary 1).

We test this implication of the theory by distinguishing between titled and untitled players. In our sample about 1.5% of decisions are made by users who hold real-world chess titles, such as candidate master, national master, international master, or grandmaster. These are some of the best chess players in the world, many of whom are professionals. By contrast, untitled users in our sample are best thought of as experienced hobbyists. It is thus plausible that titled players are, on average, more skilled than untitled ones. Given the sheer size of our data, we observe enough decisions by either type of player to ask whether their mistake frequencies are systematically different, and, if so, when.

We study the first of these questions by estimating the specification

$$Error_d = \tau Titled_p + \xi_A + \varepsilon_d, \tag{6}$$

where $Error_d$ indicates whether decision d from choice set A by player p is a mistake, and $Titled_p$ indicates whether player p holds a title of candidate master or higher. To control for the complexity of the decision problem, we add a fixed effect, ξ_A, that holds constant the number of $L-$, D-, and W-moves as well as the vector of DTM values for all L- and W-moves.

Relying on the regression specification above, the results in Table 4 imply titled players are, on average, significantly less likely to make mistakes than untitled ones. Cols. (1) and (2) show fewer mistakes for all board configurations, while cols. (3) and (4) focus on choice sets that do not contain D-moves. The estimates in cols. (2) and (4) even suggest grandmasters are less likely to err than other titled players facing similarly complex choice sets. In chess, grandmasters are typically regarded as the "best of the best," but they only account for about .15% of decisions in our data, which may explain why the difference between them and other titled players is only statistically significant in col. (2). The latter fact notwithstanding, Corollary 1 is borne out in the data.

Figure 10 provides a partial answer to the question of *when* skill matters, i.e., in which settings the differences between titled and untitled players emerge. The estimates in this figure correspond to $\{\phi_i\}$ in

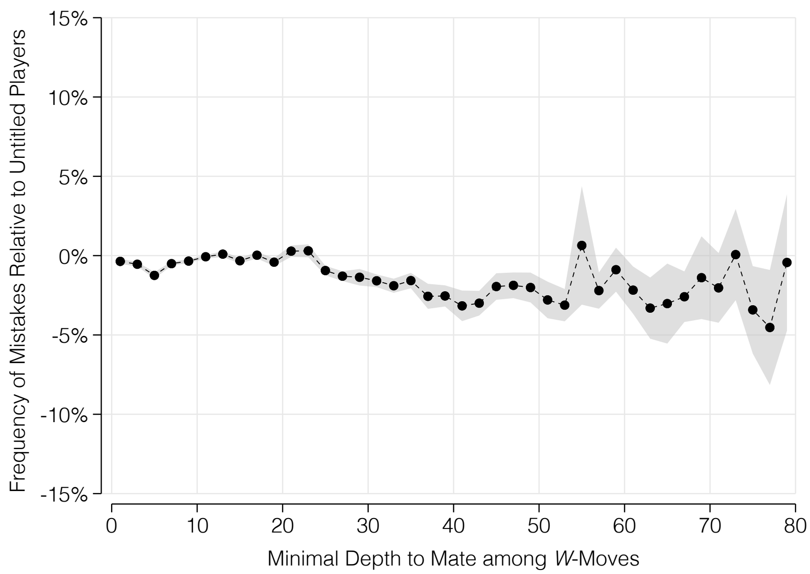
$$Error_d = \sum_{i=1}^{120} \phi_i \lfloor DTM \rfloor_{i,A} \times Titled_p + \xi_A + \varepsilon_d, \tag{7}$$

Table 4: Titled vs. Untitled Players' Frequency of Mistakes

	(1)	(2)	(3)	(4)
		Probability of Mistake		
Titled Player (÷ 100)	–0.813		–0.179	
	(0.048)		(0.061)	
Other Title (÷ 100)		–0.768		–0.155
		(0.051)		(0.064)
Grandmaster (÷ 100)		–1.134		–0.350
		(0.130)		(0.197)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	0.004	0.012
H_0: Grandmasters = Other Titled Players		0.009		0.344
Fixed Effects:				
Composition × Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	6.039	6.039	2.748	2.748
Board Configurations	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$
R^2	0.366	0.366	0.410	0.410
N	226,955,095	226,955,095	27,600,514	27,600,514

Notes: Entries are coefficients and standard errors from estimating τ in variants of eq. (6) by ordinary least squares. As explained in the text, all regressions include a fixed effect for the combination of the number of W-, D-, and L-moves and the vector of all DTM values of the W- and L-moves in same choice set. The unit of observation in each regression is a decision problem, and individual observations are reweighted so that all players receive equal total weight. The sample in cols. (1) and (2) includes all board configurations in our data, whereas cols. (3) and (4) 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 the probability of an error 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 10: Titled vs. Untitled Players' Frequency of Mistakes, by Minimal DTM among W-Moves



Notes: Figure shows point estimates and 95%-confidence intervals for ϕ_i in eq. (7), i.e., the difference in error rates between titled and untitled players when the length of the shortest winning path equals i. Negative point estimates indicate fewer mistakes among titled players. Confidence intervals account for two-way clustering by player and game.

where $Error_d$, $Titled_p$, and ξ_A are defined as in specification (6) above; and $\lfloor DTM \rfloor_{i,A}$ is an indicator equal to one if the minimal DTM among all W-moves in the choice set A is equal to i.

The estimates in Figure 10 imply small, if any, differences between both kinds of players when the minimal DTM among W-moves in the choice set is less than twenty-five. Only for choice situations in which evaluating W-moves is arguably difficult do we observe that titled players commit significantly fewer mistakes. This pattern is consistent with the theory if, around the "kink" point at a DTM of twenty-five, cognitive constraints begin to negatively affect untitled players' decision-making. More generally, the evidence in Figure 10 is suggestive of skill moderating the adverse impact of complexity on the quality of decision-making.

7. Conclusion

This paper studies two potential sources of complexity in decision-making. The first one is object complexity, which corresponds to the difficulty in evaluating any given object in a choice set. The second source is composition complexity, which refers to the difficulty of identifying the best object among similar ones.

In the first part of the paper, we develop a theory of satisficing with evaluation errors that delivers sharp predictions regarding the effect of these two sources of complexity on choice behavior. In the second part of the paper, we test these predictions in a large, novel data set on endgame moves in chess. Chess endgames provide a nearly ideal setting to study complexity in decision-making: they not only allow us to construct an objective, ordinal measure of decision quality, but also admit ample variation in object and composition complexity.

Broadly summarizing, the empirical evidence is remarkably consistent with the theory's predictions. Both object and composition complexity adversely affect the quality of decision-making—even among highly experienced players. Our results therefore help shed some of the first light on the role of complexity in decision-making outside of the laboratory.

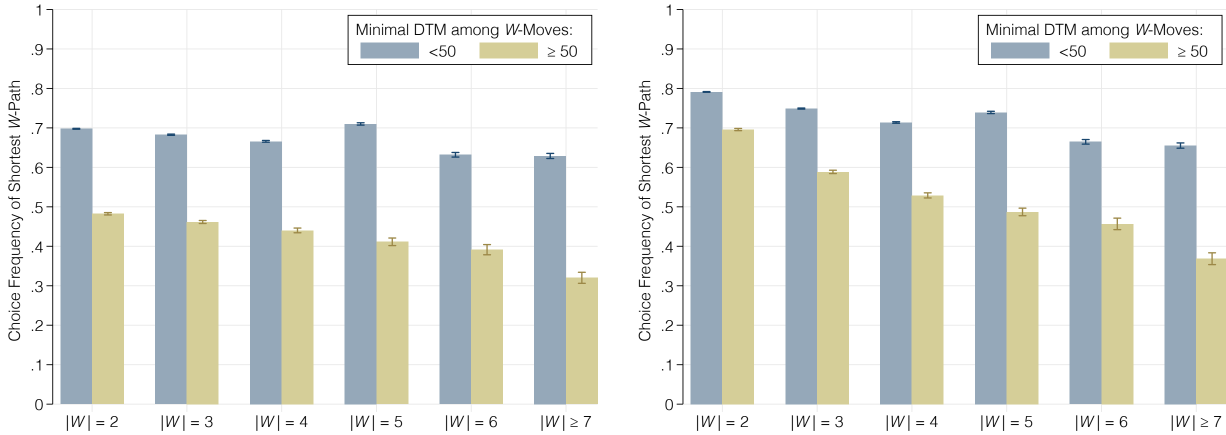
More tentatively, given that our empirical results are inconsistent with Proposition 1, the data suggest satisficing may provide a better description of actual behavior in this setting than canonical maximization. In particular, we can rule out the possibility that chess players always conduct noisy evaluations of all alternatives before choosing the one with the highest estimated value.

Of course, speed chess is an environment in which satisficing may be *a priori* more plausible than canonical maximization. It is therefore natural to wonder whether one would obtain similar results in settings where time pressure is less of an issue. The answer is "yes." In Appendix C.3, we repeat the analysis focusing on the approximately 6.5 million decisions in our data that were made subject to classical time controls. On Lichess, games with classical

Figure 11: Choice Frequency of Simplest W-Move, by Number of W-Moves and DTM

(a) All Decisions

(b) Conditional on Choosing a W-Move



Notes: Panel (a) shows the raw frequency with which the shortest available W-path is chosen, depending on the total number of W-moves in the choice set and the minimal DTM among them. The frequencies in panel (b) are conditional on any W-move having been selected. Error bars correspond to 95%-confidence intervals and account for two-way clustering by player and game.

time controls typically last more than 25 minutes. Reassuringly, the results from this analysis are qualitatively similar to those in the main text. We also observe that the gap in mistake frequencies between titled and untitled players increases under classical time controls. This finding suggests time pressure may affect expert players more than others. In other words, time and skill might be complements in decision-making.

More generally, chess is an environment in which complexity varies dramatically across board configurations. In some configurations, complexity may not be an issue. For example, if there is a W-move with DTM of ten or fifteen, players may instantaneously recognize it as “promising,” carefully evaluate it to verify their initial intuition, and choose it if the estimated value exceeds the threshold. One way to model such behavior is to assume that, instead of being random, the search order favors promising over non-promising alternatives. In other configurations, complexity may be binding in the sense that all moves have high DTM. Instantaneously recognizing some moves as promising becomes challenging, and players may resort to random search.

In nearly 13 million choice sets in the data, the minimal DTM among W-moves is at least fifty. Our results continue to hold in this environment, in which complexity is likely binding (cf. Appendix C.4). Interestingly, we observe a significant decline in the frequency of choosing the simplest W-move, as Figure 11 illustrates. Figure 11 also shows that the decline widens significantly as the number of W-moves in the choice set increases—even when conditioning on choosing some W-move. One possible reason is that the search order is indeed related to complexity: non-random search is more likely to be used in simple settings, whereas random search becomes more prevalent as complexity increases. Exploring this possibility remains an important challenge for future research.

References

Abeler, J. and S. Jäger (2015). Complex Tax Incentives. *American Economic Journal: Economic Policy 7*(3), 1–28.

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

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.

Chernev, A., U. Böckenholt, and J. Goodman (2015). Choice Overload: A Conceptual Review and Meta-Analysis. *Journal of Consumer Psychology 25*(2), 333–358.

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.

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

Enke, B. and T. Graeber (2021). Cognitive Uncertainty. NBER Working Paper No. 26518.

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.

Glickman, M. E. (1999). Parameter Estimation in Large Dynamic Paired Comparison Experiments. *Journal of the Royal Statistical Society: Series C 48*(3), 377–394.

Glickman, M. E. (2000). The Glicko-2 System for Rating Players in Head-to-Head Competition. mimeographed, Boston University.

Gonzalez-Diaz, J. and I. Palacios-Huerta (2016). Cognitive Performance in Competitive Environments: Evidence from a Natural Experiment. *Journal of Public Economics 139*, 40–52.

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.

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

Levitt, S. and J. List (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives 21*(2), 153–174.

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

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

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

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

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

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

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

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

Osborne, M. and A. Rubinstein (1994). *A Course in Game Theory.* Cambridge, MA: MIT Press.

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.

Schwartz, B. (2004). *The Paradox of Choice: Why More Is Less.* New York: HarperCollins.

Silman, J. (2007). *Silman's Complete Endgame Course.* Los Angeles: Siles Press.

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.

von Neumann, J. (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen 100*(1), 295–320.

von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

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.

Appendix

Contents

List of Figures

List of Tables

Appendix A: Proofs

A.1. *Proof of Prediction 1*

By Assumption 1, the choice probability of any W-move is larger than $1/2$, and the choice probability of any D- or L-move is smaller than $1/2$. It thus remains to be proven that the choice probability of any D-move is larger than that of any L-move. To do so, we first state the following lemma.

LEMMA 1: *Let $X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$. Then,*

$$Pr(X_1 \geq T) > Pr(X_2 \geq T) \iff T(\sigma_2 - \sigma_1) < \mu_1\sigma_2 - \mu_2\sigma_1.$$

PROOF: Let Φ denote the CDF of the standard normal distribution. Then,

$$Pr(X_i \geq T) = Pr\left(\frac{X_i - \mu_i}{\sigma_i} \geq \frac{T - \mu_i}{\sigma_i}\right) = 1 - \Phi\left(\frac{T - \mu_i}{\sigma_i}\right).$$

The required ranking of the probabilities holds if and only if

$$\Phi\left(\frac{T - \mu_2}{\sigma_2}\right) > \Phi\left(\frac{T - \mu_1}{\sigma_1}\right) \iff \frac{T - \mu_2}{\sigma_2} > \frac{T - \mu_1}{\sigma_1} \iff T(\sigma_2 - \sigma_1) < \mu_1\sigma_2 - \mu_2\sigma_1.$$

Q.E.D.

The condition $T(\sigma_2 - \sigma_1) < \mu_1\sigma_2 - \mu_2\sigma_1$ identified in Lemma 1 holds when the first random variable corresponds to a D-move x and the second corresponds to an L-move y, because in this case $\mu_1 = 0$, $\mu_2 = L = -W$, $W > T$ by Assumption 1, and $\sigma_1 \geq \frac{1}{2}\sigma_2$ by Assumption 2. Thus, $Pr(u_x \geq T) > Pr(u_y \geq T)$.

To complete the proof, fix a choice set and two evaluation orders of moves O_1 and O_2 that are identical except that x appears before y in O_1, and their locations are switched in O_2. The choice probability of x from these two orderings is

$$Pr(u_x \geq T) \times \Big(Pr(T \text{ not exceeded prior to } x \text{ in } O_1) + Pr(T \text{ not exceeded prior to } x \text{ in } O_2)\Big).$$

The choice probability of y is identical except for $Pr(u_y \geq T)$ replacing $Pr(u_x \geq T)$. The choice probability of x is larger than that of y because

(i) $Pr(u_x \geq T) > Pr(u_y \geq T)$ as we proved above,

(ii) $Pr(T$ not exceeded prior to x in $O_1) = Pr(T$ not exceeded prior to$y$ in $O_2)$ because the evaluation order prior to reaching the first of the two moves x and y is identical in O_1 and O_2, and

(iii) $Pr(T$ not exceeded prior to x in $O_2) > Pr(T$ not exceeded prior to y in $O_1)$ because the order prior to reaching the second of the two moves x and y is identical expect for x appearing in O_1 and being chosen with larger probability than y in O_2.

Since this ranking of the probabilities holds for every pair of evaluation orders in which the locations of x and y are switched, it holds for their choice probabilities from the choice set. *Q.E.D.*

A.2. *Proof of Prediction 2*

We prove the result for W-moves. The proof for D- and L-moves is analogous.

Assume to the contrary that the choice probability of x is weakly smaller than the choice probability of y. Now increase σ_x until it is equal to σ_y. By Prediction 3 (to be proved below), the choice probability of x decreases and the choice probability of y increases, implying the choice probability of x remains smaller than that of y. This is in contrast to the fact that two W-moves with the same object complexity should be chosen with the same probability. *Q.E.D.*

A.3. *Proof of Prediction 3*

We prove the result for a W-move x. The proof for other moves is analogous.

Fix the order in which the DM evaluates moves. If x is last in the order, its choice probability conditional on reaching it is 1 independently of its complexity σ_x. Otherwise, its choice probability conditional on the order is

$$Pr(T \text{ not exceeded prior to } x) \times Pr(u_x \geq T).$$

The first component in this expression is independent of σ_x, and the second decreases in σ_x because $T < W$. Consequently, the choice probability of any move that appears after x in the order increases in σ_x. Because this holds for any order in which x does not appear last, the result follows. *Q.E.D.*

A.4. *Proof of Prediction 4*

By Assumptions 1 and 2 and Lemma 1, we have that $Pr(u_x \geq T) > Pr(u_y \geq T)$ for a D-move x and an L-move y.

Fix a choice set A. Let $A_x = A \cup \{x\}$ and $A_y = A \cup \{y\}$. Fix an evaluation order O_1 of A_x in which x does not appear last, and an evaluation order O_2 of A_y, which is identical to O_1 except that y replaces x. The probability of making a mistake prior to x in O_1 and y in O_2 is identical. The conditional probability of making a mistake when evaluating x in O_1 is larger than when evaluating y in O_2. The probability of making a mistake conditional on $x < T$ in O_1 is identical to the corresponding conditional probability in O_2. The result follows. *Q.E.D.*

A.5. *Proof of Proposition 1*

Because posterior expected values are ranked in the same way as the estimates, and because the DM chooses the maximal expected value, move x is chosen from some choice set if and only if its realized u-value is larger than the realized u-values of all other moves in the choice set. Thus, to prove the result, it suffices to establish the following:

Lemma 2: *Let $X_1, \ldots, X_N$ be N independently distributed normal random variables where $X_i \sim N(\mu_i, \sigma_i)$, $\mu_1 = \mu_2$, and $\sigma_1 < \sigma_2$. Then,*

$$P_2 = \mathbb{P}\left[X_2 > \{X_j\}_{j\neq 2}\right] > \mathbb{P}\left[X_1 > \{X_j\}_{j\neq 1}\right] = P_1.$$

Proof: Without loss of generality, we rescale all RVs so that each X_i is replaced by $\frac{X_i - \mu_1}{\sigma_1}$. After rescaling, we have that $X_1 \sim N(0,1)$ and $X_2 \sim N(0,\sigma)$ where $\sigma = \sigma_2/\sigma_1 > 1$.

Let $h(x)$ denote the probability that $X_3, \ldots, X_N \leq x$. Then, $h(x)$ strictly increases in x. Let F and f (G and g) denote the CDF and PDF of X_1 (X_2) respectively. Then,

$$P_2 - P_1 = \int_{-\infty}^{\infty} h(x)\Big(F(x)g(x) - G(x)f(x)\Big)dx.$$

Because $X_2 - X_1$ is distributed normal with mean 0, we have that

$$0 = \mathbb{P}(X_2 > X_1) - \mathbb{P}(X_1 > X_2) = \int_{-\infty}^{\infty} \Big(F(x)g(x) - G(x)f(x)\Big)dx.$$

Because $h(x)$ strictly increases in x, if we were to show that the function $m(x) = F(x)g(x) - G(x)f(x)$ crosses 0 exactly once from below at some $\hat{x}$ then the conclusion of the lemma would follow because in this case,

$$\begin{aligned}
\int_{-\infty}^{\infty} h(x)\Big(F(x)g(x) - G(x)f(x)\Big)dx &= \int_{-\infty}^{\hat{x}} h(x)m(x)dx + \int_{\hat{x}}^{\infty} h(x)m(x)dx \\
&> \int_{-\infty}^{\hat{x}} h(\hat{x})m(x)dx + \int_{\hat{x}}^{\infty} h(\hat{x})m(x)dx \\
&= h(\hat{x})\int_{-\infty}^{\infty} m(x)dx = 0.
\end{aligned}$$

To conclude the proof, it thus suffices to show that $m(x)$ crosses 0 exactly once from below at $\hat{x} > 0$. We do so by showing that

(i) $m(x) \leq 0$ for all $x \leq 0$ with strict inequality for $x = 0$,

(ii) $m(x) > 0$ for some $x > 0$, and

(iii) $m(x)$ cannot cross 0 from above at some $x > 0$.

To verify (i), observe first that $m(0) = 1/2(g(0) - f(0)) < 0$. To prove that $m(x) \leq 0$ for any $x < 0$, fix $x < 0$ and consider the interval $[a, x]$ with $a < x$. By Cauchy's mean value theorem, there exists $a < c < x$ such that

$$(F(x) - F(a))g(c) = (G(x) - G(a))f(c).$$

Since the ratio $g(x)/f(x)$ decreases in x for $x < 0$, we have that

$$(F(x) - F(a))g(x) < (G(x) - G(a))f(x),$$

which is true for any $a < x$. We take a to $-\infty$ to obtain $F(x)g(x) \leq G(x)f(x)$, i.e., $m(x) \leq 0$.

To verify (ii), observe that at the point of intersection x between f and g to the right of 0, $m(x)$ is proportional to $F(x) - G(x)$ and $F(x) - G(x) > 0$.

To verify (iii), assume to the contrary that $m(x)$ crosses 0 from above at some $x > 0$. Then, the following two equations should hold at x:

(1) $F(x)g(x) - G(x)f(x) = 0$, and

(2) $m'(x) \leq 0$ where m' is the derivative of m.

We develop $m'(x)$ to show (1) and (2) cannot hold simultaneously. Because the PDF z of a normal distribution with mean 0 and variance σ^2 satisfies the identity $z'(x) = -\frac{xz(x)}{\sigma^2}$, we have that

$$0 \geq m'(x) = f(x)g(x) + F(x)g'(x) - f(x)g(x) - G(x)f'(x) = xf(x)G(x) - \frac{xg(x)}{\sigma^2}F(x),$$

implying $f(x)G(x) \leq \frac{g(x)}{\sigma^2}F(x) < g(x)F(x)$ in contradiction to (1). *Q.E.D.*

Appendix B: Data Appendix

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

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

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

Appendix C: Robustness Checks

C.1. *Weighting Observations Equally*

Our findings are very similar when we do not reweight individual observations so that all players receive equal weight. To show this, we replicate the key tables and figures in the main text, weighting all observations equally. The respective results are shown in Appendix Figures AF.2–AF.5 and Tables AT.3–AT.4.

C.2. *Controlling for Time Left*

Since the timing of decisions is endogenous, we do not control for it in our main analysis. We do, however, obtain qualitatively equivalent findings when we account for it. To show this, we replicate the regression-based tables and figures in the main text. In Appendix Figures AF.6–AF.8 and Tables AT.6–AT.7, we control for the time that remains on the player's clock when it is her turn to move. In Figures AF.9–AF.11 and Tables AT.9–AT.10, we control for the time that was left per move if the player were to follow the shortest W-path. Regardless of how we account for the possibility that players face time pressure, our findings on how complexity affects decision-making are qualitatively equivalent.

C.3. *Restricting Attention to Games with Classical Time Controls*

In Appendix Figures AF.13–AF.16 and Tables AT.12–AT.13, we replicate the results in the main text restricting attention to games played under "classical" time controls, which typically run longer than 25 minutes. More specifically, Lichess classifies the time controls in a game as classical if and only if the estimated game duration exceeds 1,500 seconds, with the estimated duration defined as: (initial clock time) + 40 × (clock increment). The results with classical time controls are qualitatively equivalent to those in the main text.

C.4. *Restricting Attention to Board Positions with High Minimal DTM*

In Appendix Figures AF.18–AF.20 and Tables AT.15–AT.16, we replicate the results in the main text, restricting attention to board positions in which the minimal DTM among W-moves exceeds fifty. These are positions in which it is *a priori* highly unlikely that players can recognize either the type of a move or its DTM without careful evaluation, as assumed in our model. Reassuringly, the results from this smaller sample are qualitatively equivalent to those in the main text.

Appendix D: Replication with Independent Data from *This Week in Chess*

Appendix Figures AF.21–AF.25 and Tables AT.17–AT.19 replicate the results in the main text, using an independent dataset that we obtained from *This Week in Chess* (TWIC). TWIC is a free, weekly publication that "rounds up the most important chess" games from the previous week (see `https://theweekinchess.com`). Most of these games are played between elite players in national and international tournaments, or chess leagues.

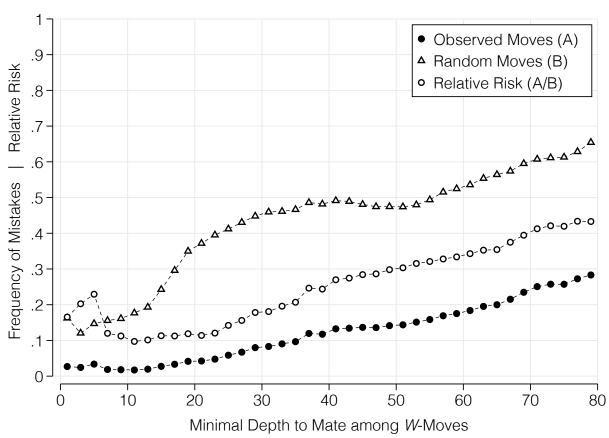
Our data include all games covered in TWIC between September 1994 and May 2020. In total, we observe 536,674 decision problems in endgame positions with six or fewer pieces, one or more legal W- and at least one D- or L-moves. The choice sets in these decision problems contain 9,067,040 legal moves.

Besides being more than two orders of magnitude smaller, the most important difference between the TWIC and Lichess data is that the former admit much less variation in players' skill. Chess players in high-profile tournaments tend to be better than the average experienced player on Lichess. This fact is reflected in a significantly lower frequency of mistakes in the TWIC data. Nonetheless, the comparative statics in Appendix Figures AF.21–AF.25 and Tables AT.17–AT.19 are qualitatively very similar to those in the main text.

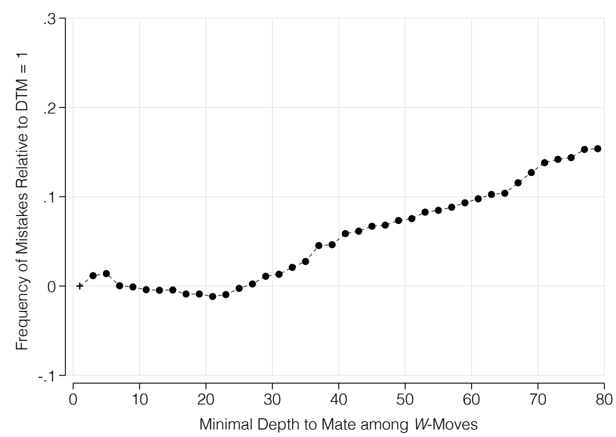
Appendix Figures

Appendix Figure AF.1: Replication of Figure 6, Weighting All Observations Equally

(a) Raw Error Rates vs. Chance



(b) Controlling for the Composition of the Choice Set

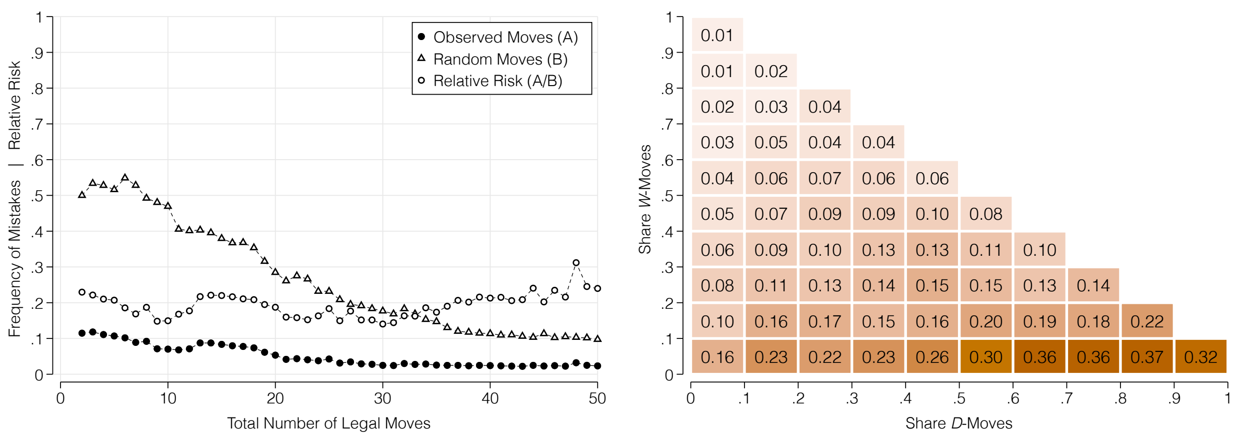


Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6 is that the results above are based on equally weighted observations.

Appendix Figure AF.2: Replication of Figure 7, Weighting All Observations Equally

(a) Mistakes vs. Total Number of Moves

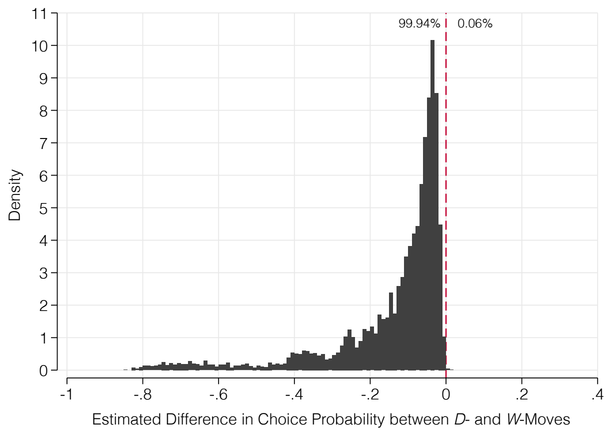
(b) Mistake Probabilities and Choice-Set Composition



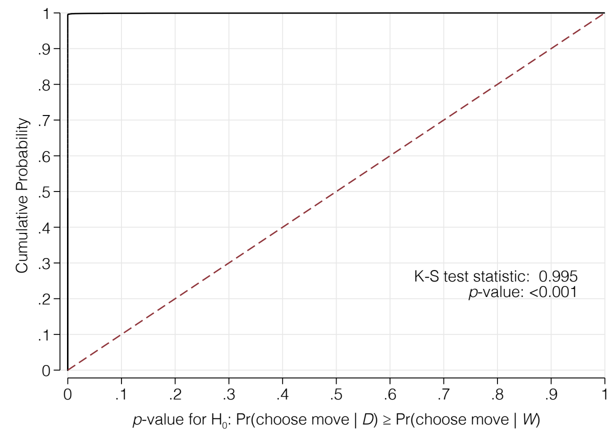
Notes: See Figure 7 in the main text. The only difference between this figure and Figure 7 is that the results above are based on equally weighted observations.

Appendix Figure AF.3: Replication of Figure 8, Weighting All Observations Equally

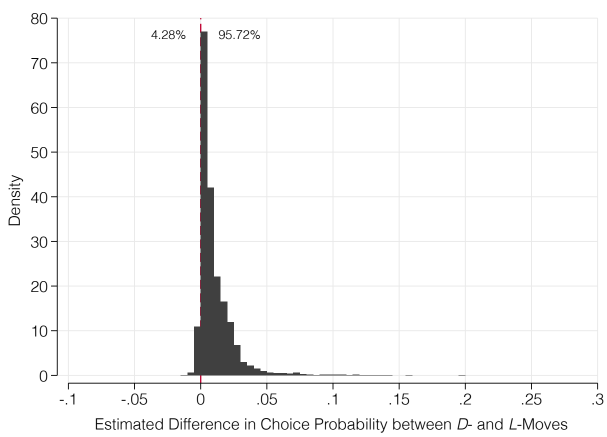
(a) Difference in Probability of Choosing D- and W-Moves



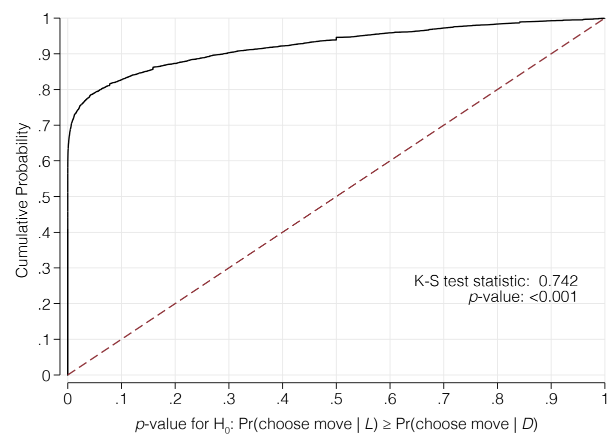
(b) CDF of p-values for Differences in Panel (a)



(c) Difference in Probability of Choosing D- and L-Moves



(d) CDF of p-values for Differences in Panel (c)

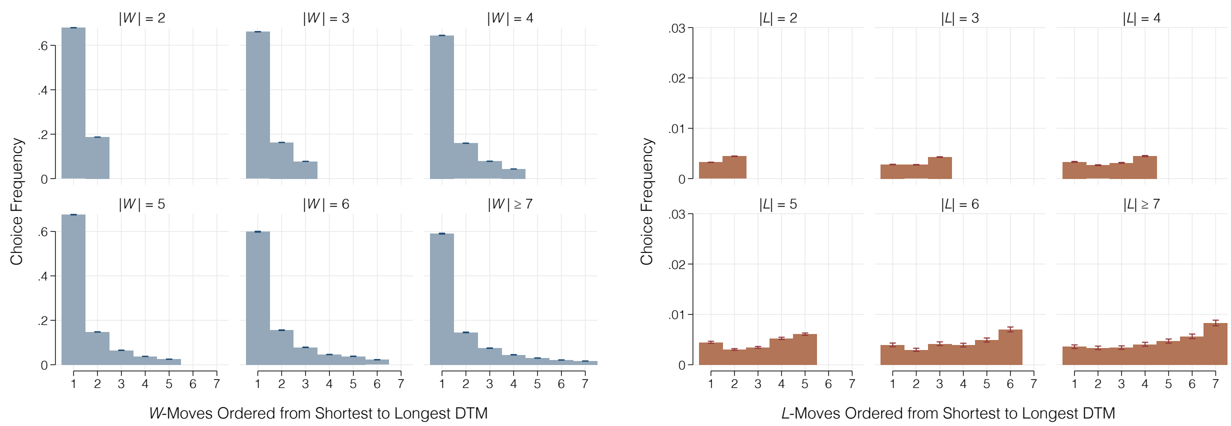


Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above are based on equally weighted observations.

Appendix Figure AF.4: Replication of Figure 9, Weighting All Observations Equally

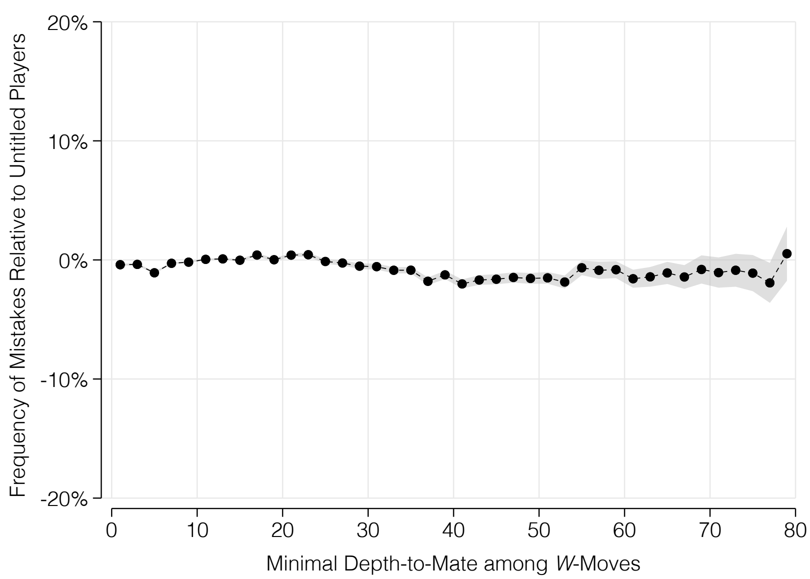
(a) W-Moves

(b) L-Moves



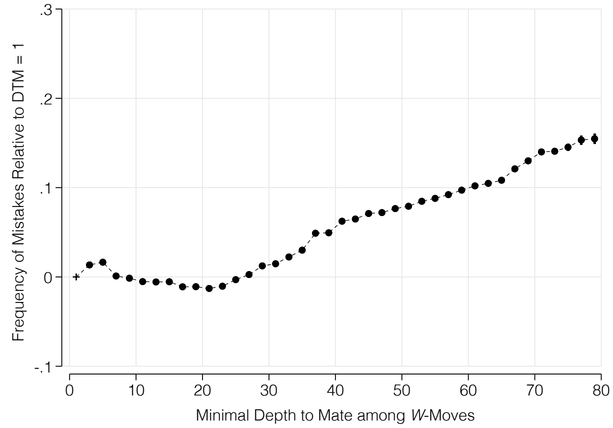
Notes: See Figure 9 in the main text. The only difference between this figure and Figure 9 is that the results above are based on equally weighted observations.

Appendix Figure AF.5: Replication of Figure 10, Weighting All Observations Equally



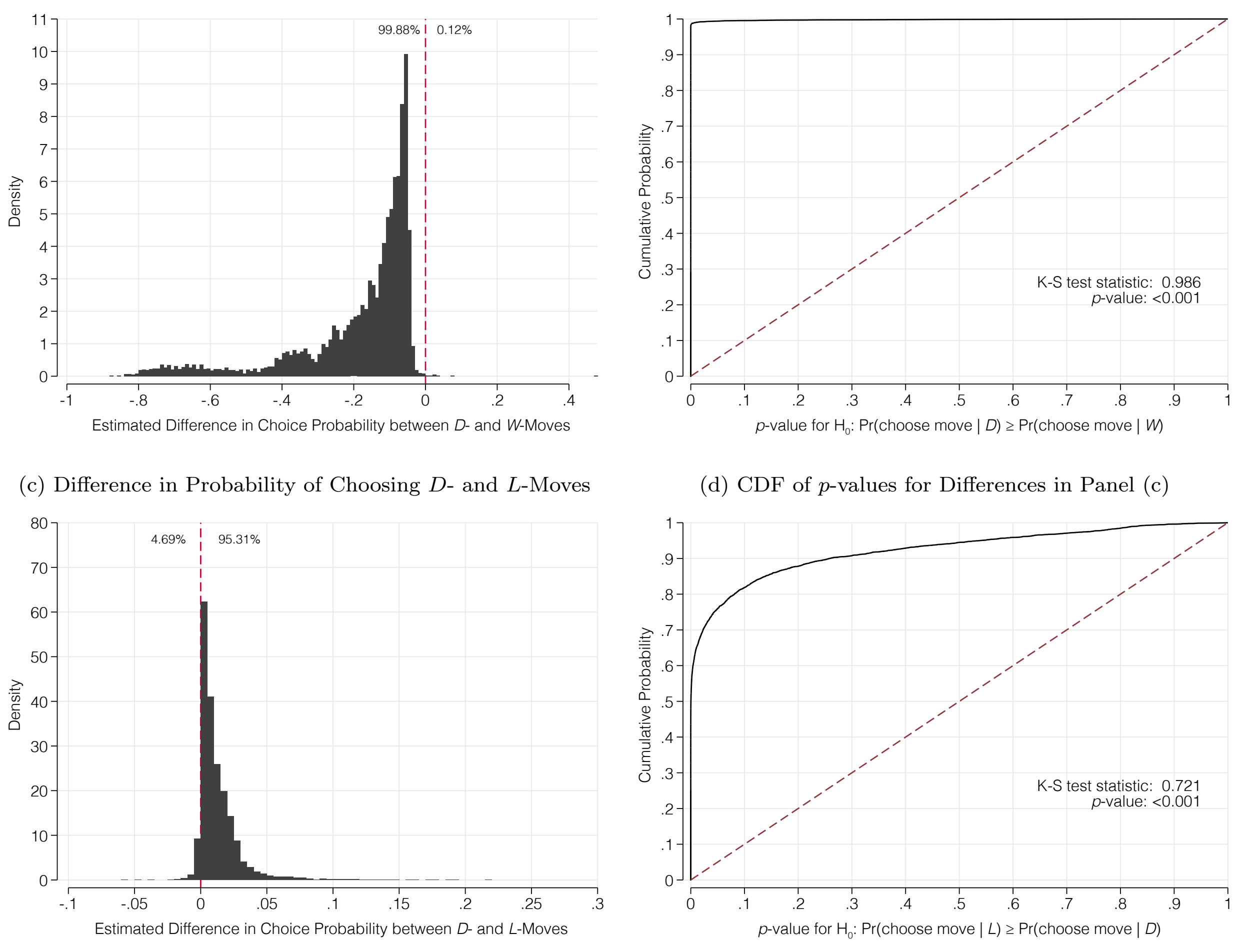
Notes: See Figure 10 in the main text. The only difference between this figure and Figure 10 is that the results above are based on equally weighted observations.

Appendix Figure AF.6: Replication of Figure 6(b), Controlling for Time Left on Clock



Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6(b) is that the results above also control for the clock reading when it is a player's turn to move.

Appendix Figure AF.7: Replication of Figure 8, Controlling for Time Left on Clock



(a) Difference in Probability of Choosing D- and W-Moves

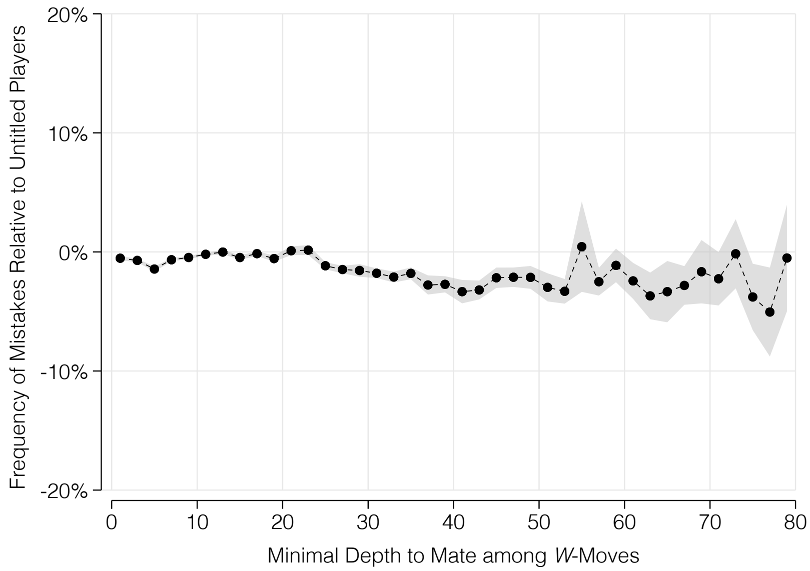
(b) CDF of p-values for Differences in Panel (a)

(c) Difference in Probability of Choosing D- and L-Moves

(d) CDF of p-values for Differences in Panel (c)

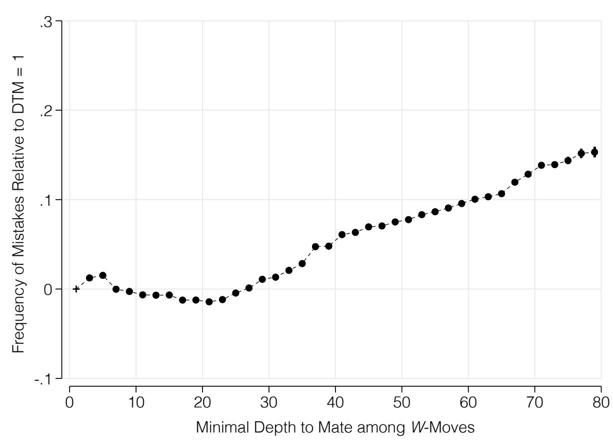
Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above also control for the clock reading when it is a player's turn to move.

Appendix Figure AF.8: Replication of Figure 10, Controlling for Time Left on Clock



Notes: See Figure 10 in the main text. The only difference between this figure and Figure 8 is that the results above also control for the clock reading when it is a player's turn to move.

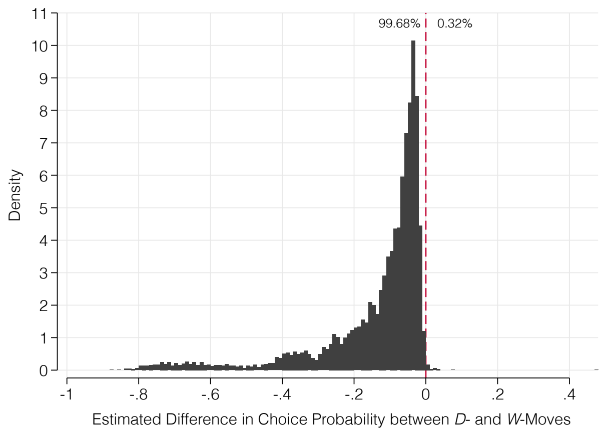
Appendix Figure AF.9: Replication of Figure 6(b), Controlling for Time Left per Move



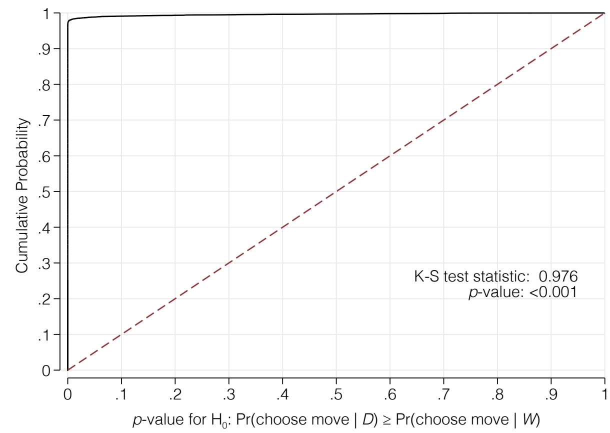
Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6(b) is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Figure AF.10: Replication of Figure 8, Controlling for Time Left per Move

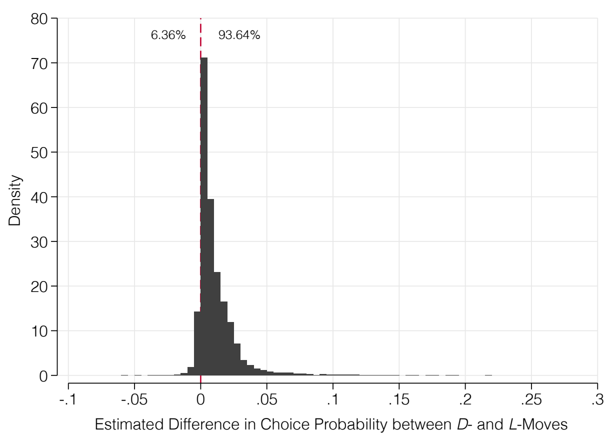
(a) Difference in Probability of Choosing D- and W-Moves



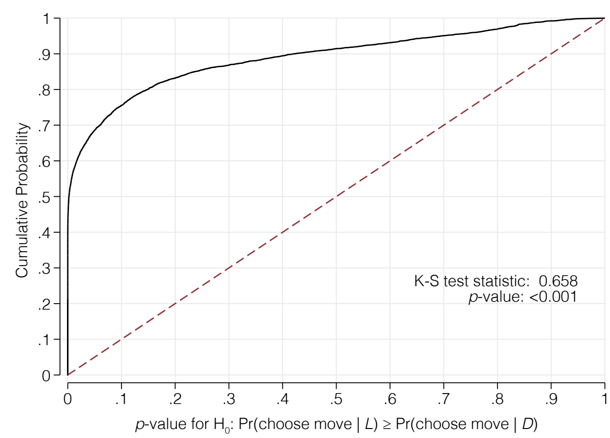
(b) CDF of p-values for Differences in Panel (a)



(c) Difference in Probability of Choosing D- and L-Moves

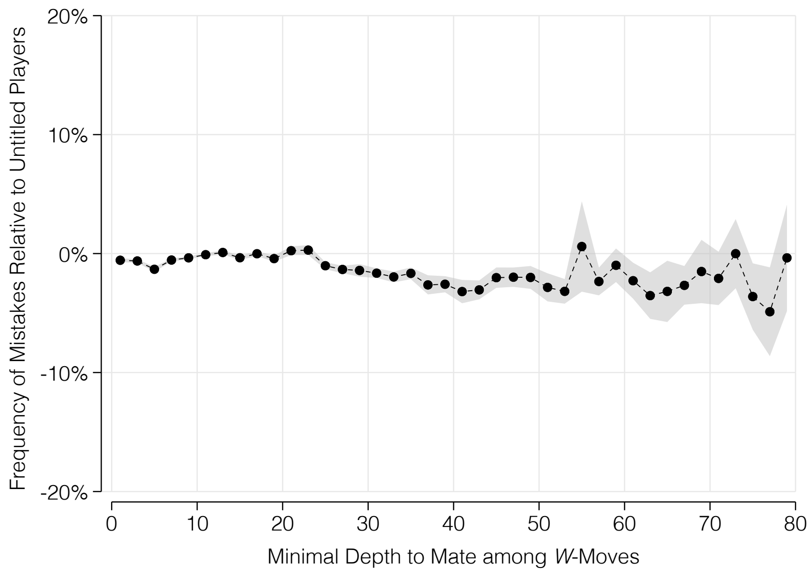


(d) CDF of p-values for Differences in Panel (c)



Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Figure AF.11: Replication of Figure 10, Controlling for Time Left per Move

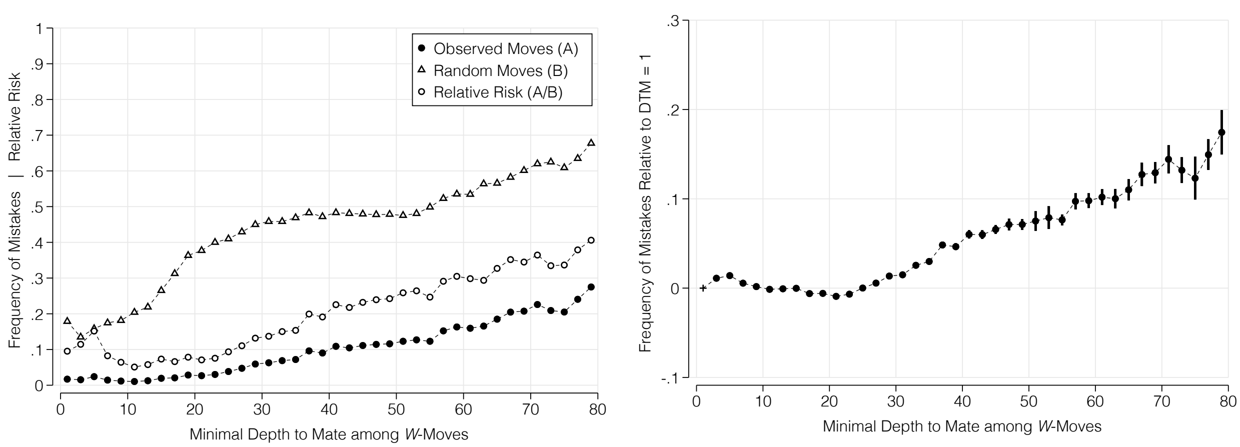


Notes: See Figure 10 in the main text. The only difference between this figure and Figure 6(b) is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Figure AF.12: Replication of Figure 6, Games with Classical Time Controls Only

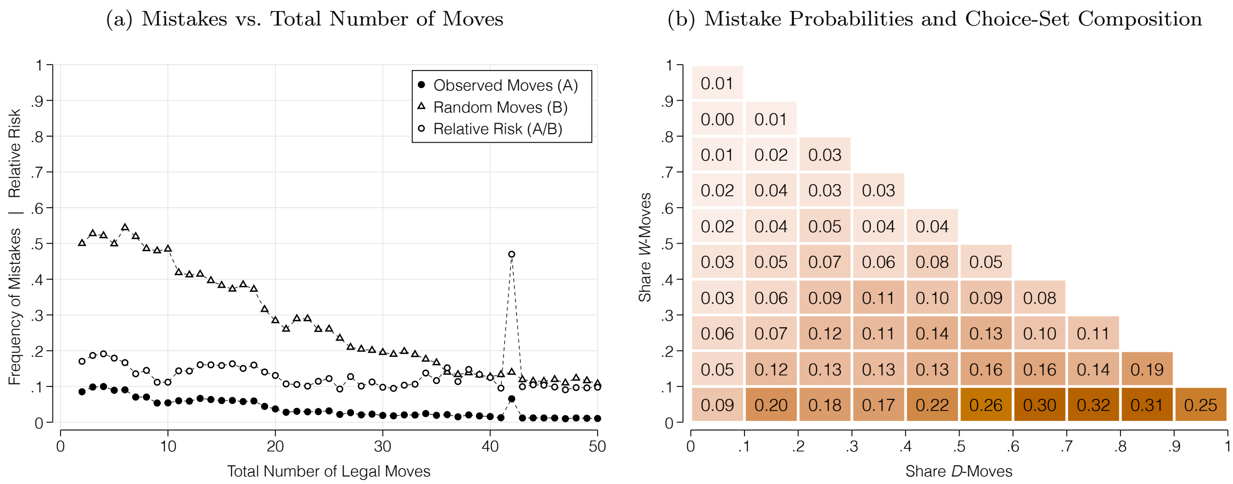
(a) Raw Error Rates vs. Chance

(b) Controlling for the Composition of the Choice Set



Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6 is that the results above restrict attention to games played under classical time controls.

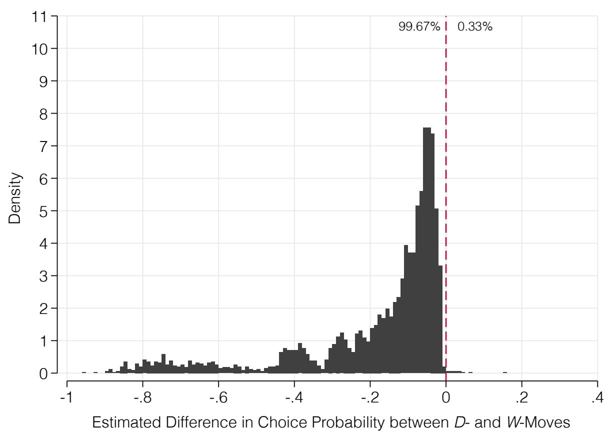
Appendix Figure AF.13: Replication of Figure 7, Games with Classical Time Controls Only



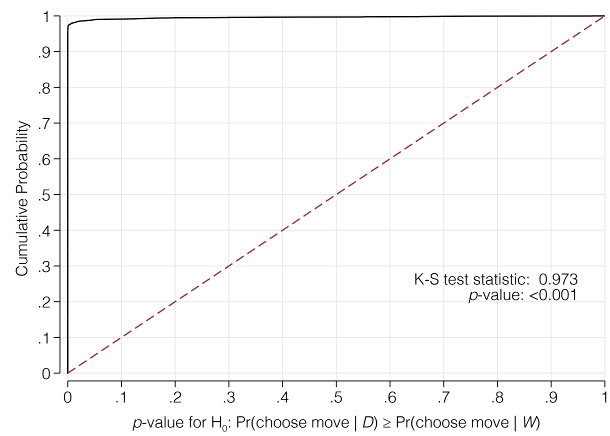
Notes: See Figure 7 in the main text. The only difference between this figure and Figure 7 is that the results above restrict attention to games played under classical time controls.

Appendix Figure AF.14: Replication of Figure 8, Games with Classical Time Controls Only

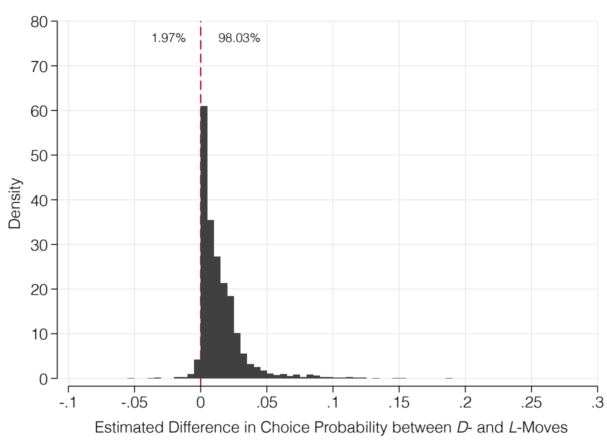
(a) Difference in Probability of Choosing D- and W-Moves



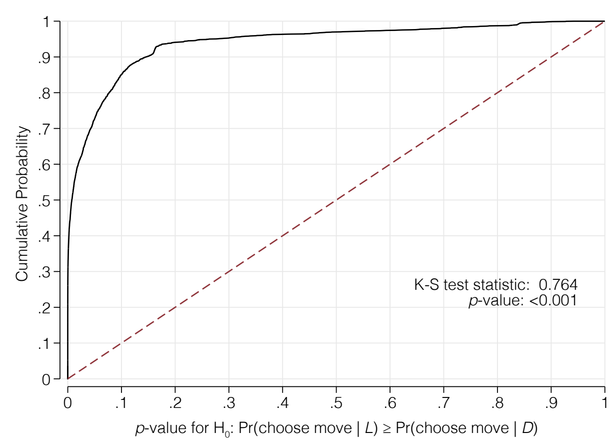
(b) CDF of p-values for Differences in Panel (a)



(c) Difference in Probability of Choosing D- and L-Moves

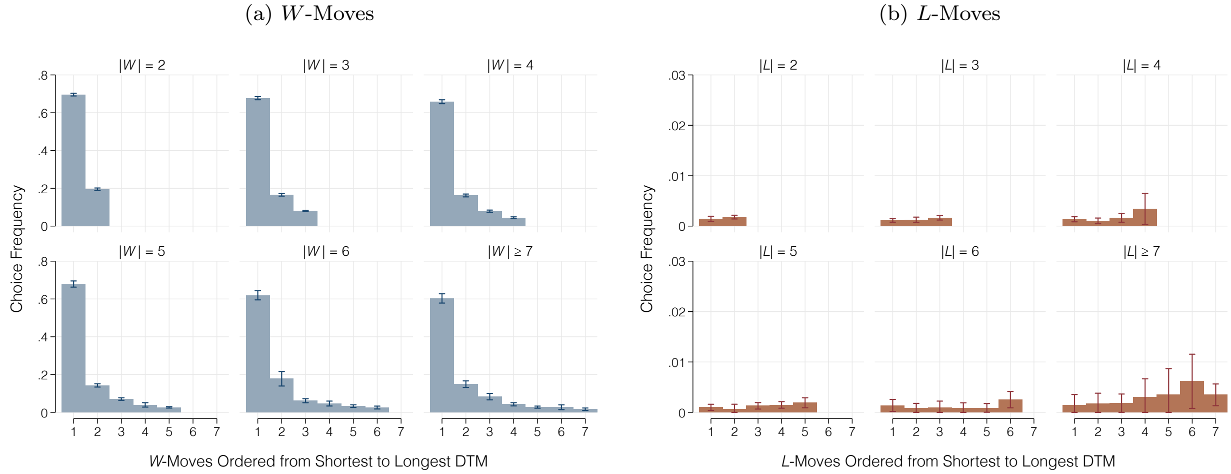


(d) CDF of p-values for Differences in Panel (c)



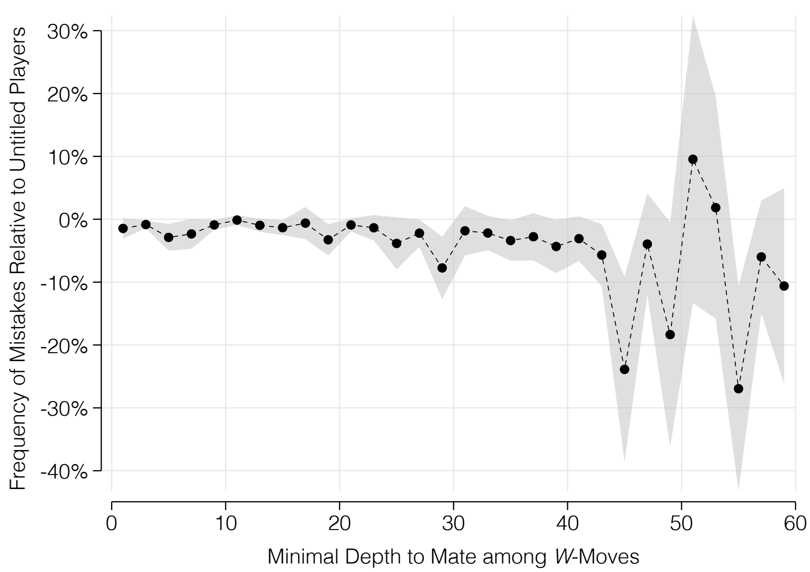
Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above restrict attention to games played under classical time controls.

Appendix Figure AF.15: Replication of Figure 9, Games with Classical Time Controls Only



Notes: See Figure 9 in the main text. The only difference between this figure and Figure 9 is that the results above restrict attention to games played under classical time controls.

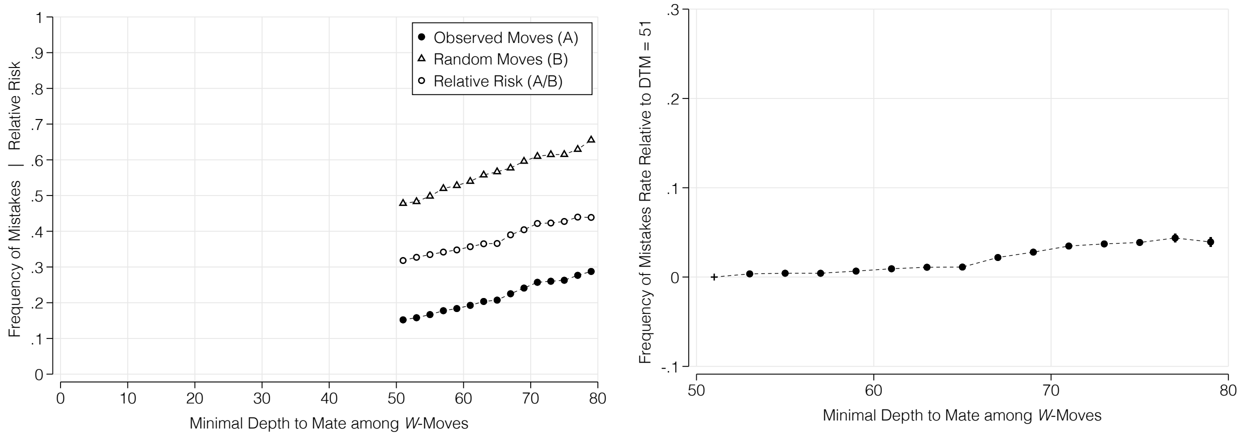
Appendix Figure AF.16: Replication of Figure 10, Games with Classical Time Controls Only



Notes: See Figure 10 in the main text. The only difference between this figure and Figure 10 is that the results above restrict attention to games played under classical time controls.

Appendix Figure AF.17: Replication of Figure 6, Board Positions with High Minimal DTM Only

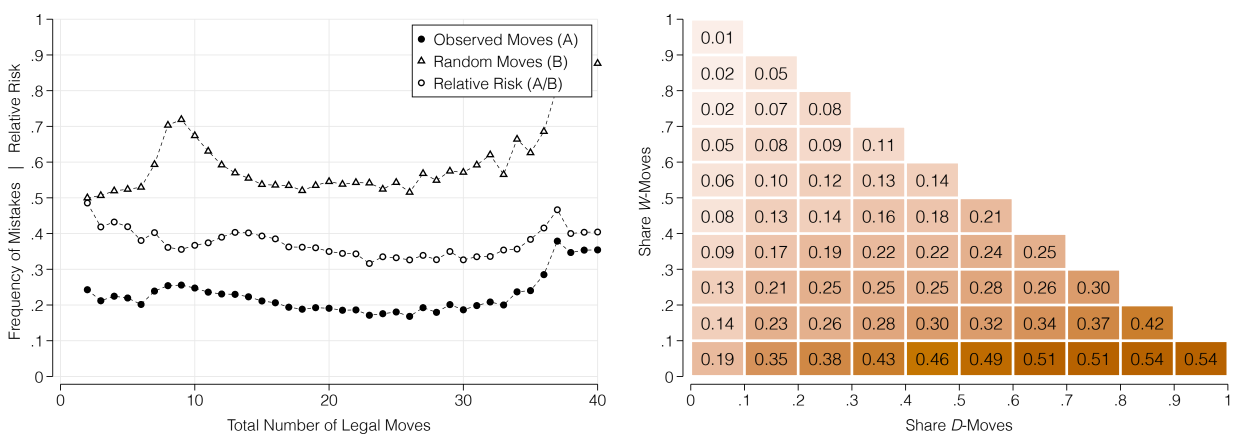
(a) Raw Error Rates vs. Chance (b) Controlling for the Composition of the Choice Set



Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

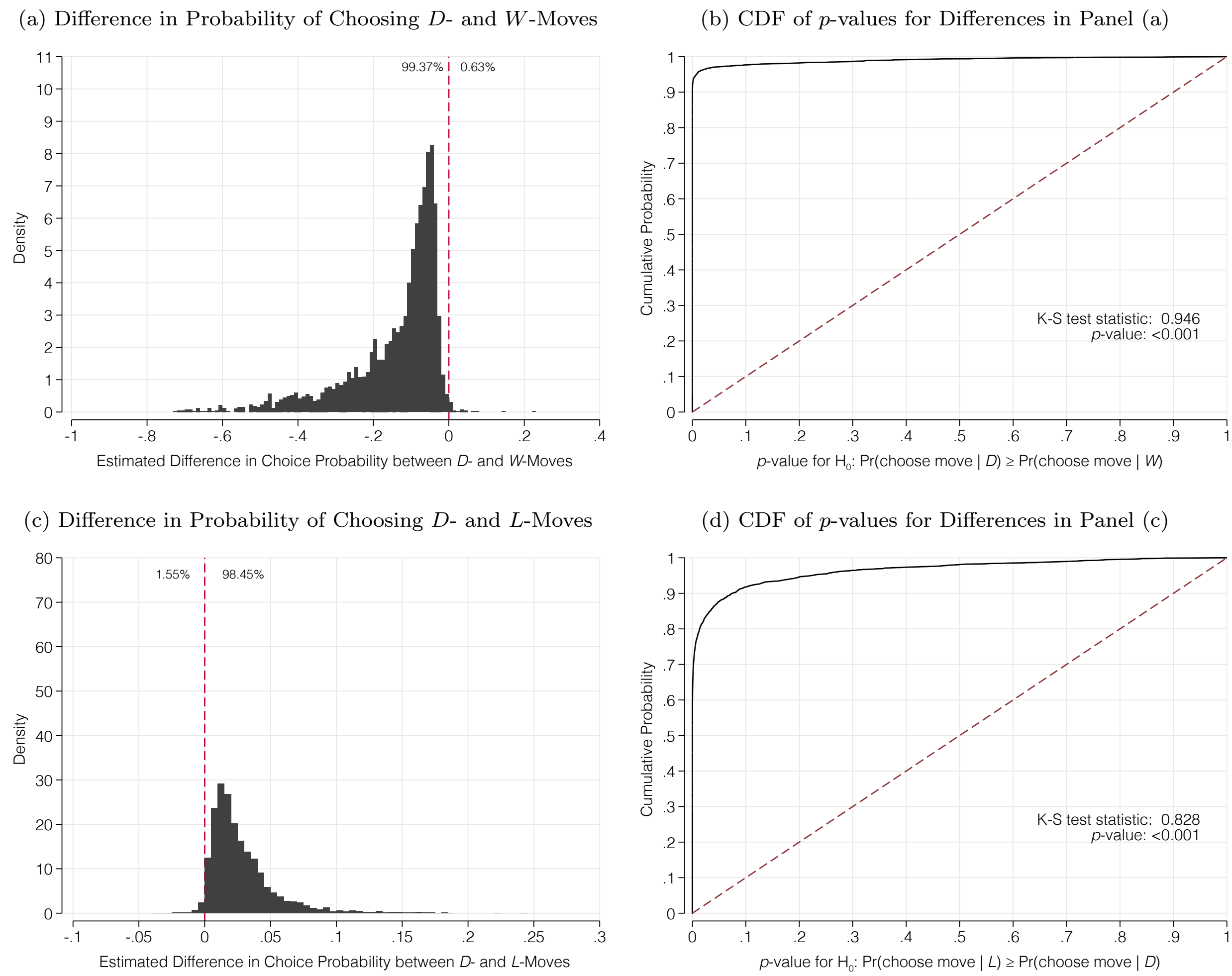
Appendix Figure AF.18: Replication of Figure 7, Board Positions with High Minimal DTM Only

(a) Mistakes vs. Total Number of Moves (b) Mistake Probabilities and Choice-Set Composition



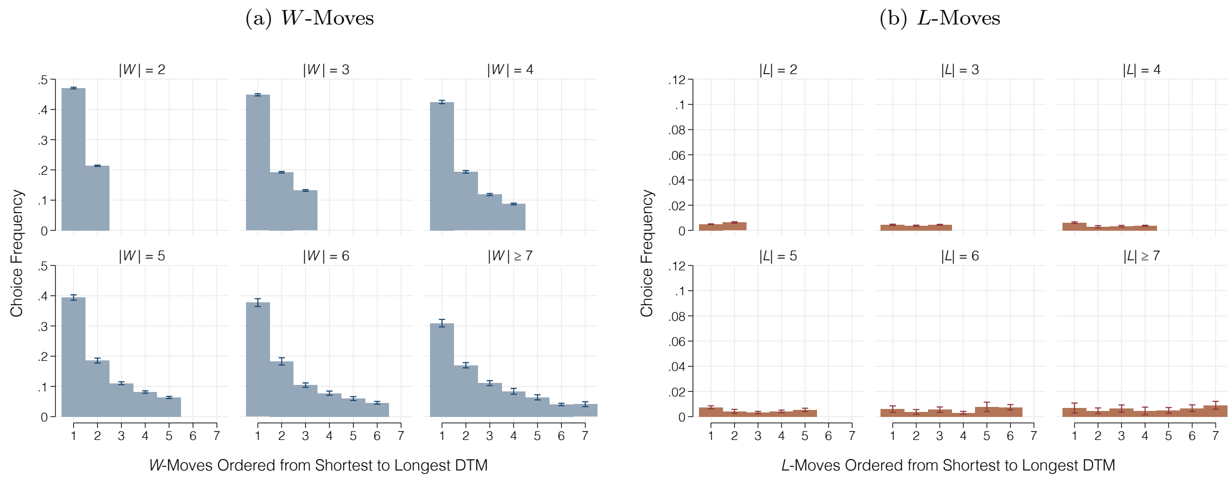
Notes: See Figure 7 in the main text. The only difference between this figure and Figure 7 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

Appendix Figure AF.19: Replication of Figure 8, Board Positions with High Minimal DTM Only



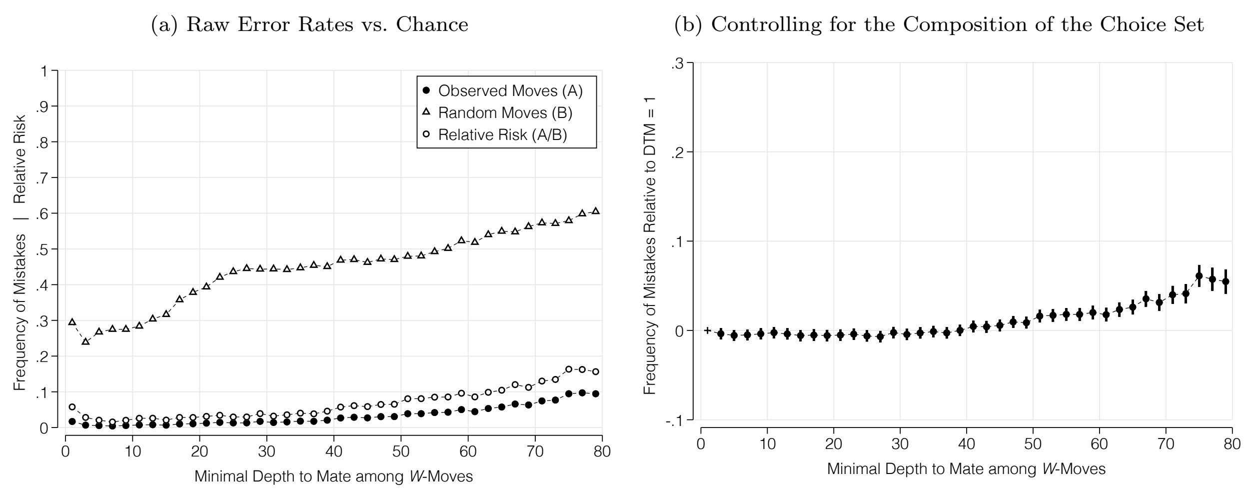
Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

Appendix Figure AF.20: Replication of Figure 9, Board Positions with High Minimal DTM Only



Notes: See Figure 9 in the main text. The only difference between this figure and Figure 9 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

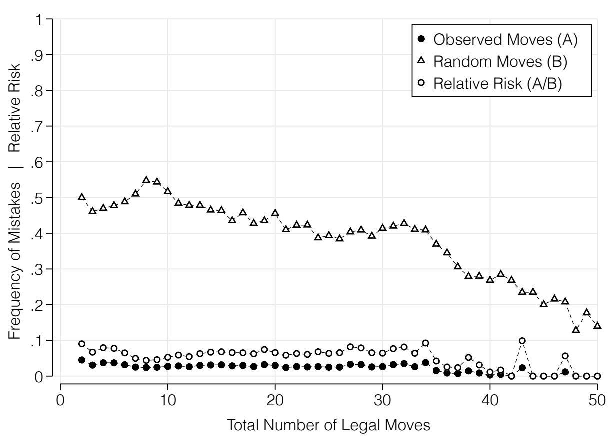
Appendix Figure AF.21: Replication of Figure 6, TWIC Data



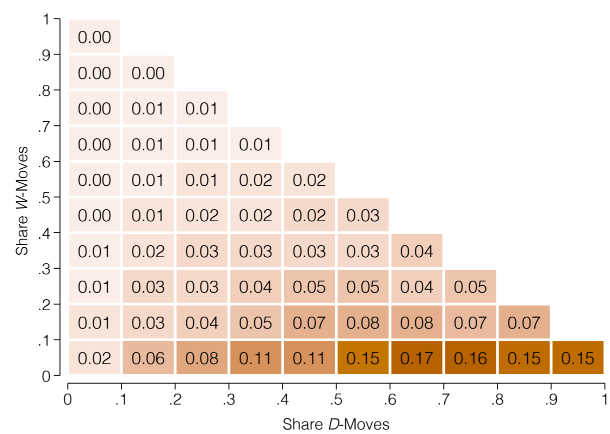
Notes: See Figure 6 in the main text. The only difference between this figure and Figure 6 is that the results above are based on data from *This Week in Chess.*

Appendix Figure AF.22: Replication of Figure 7, TWIC Data

(a) Mistakes vs. Total Number of Moves

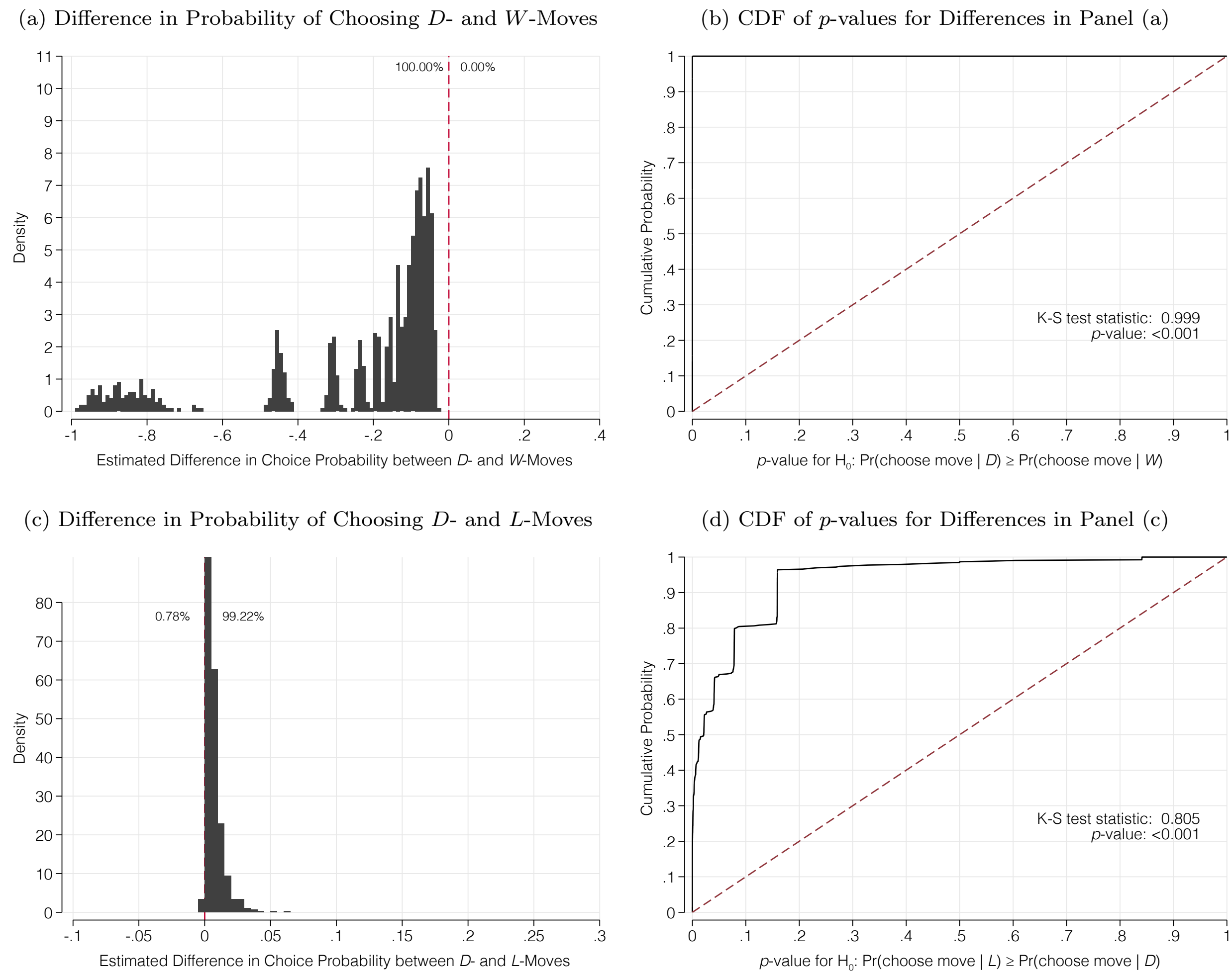


(b) Mistake Probabilities and Choice-Set Composition



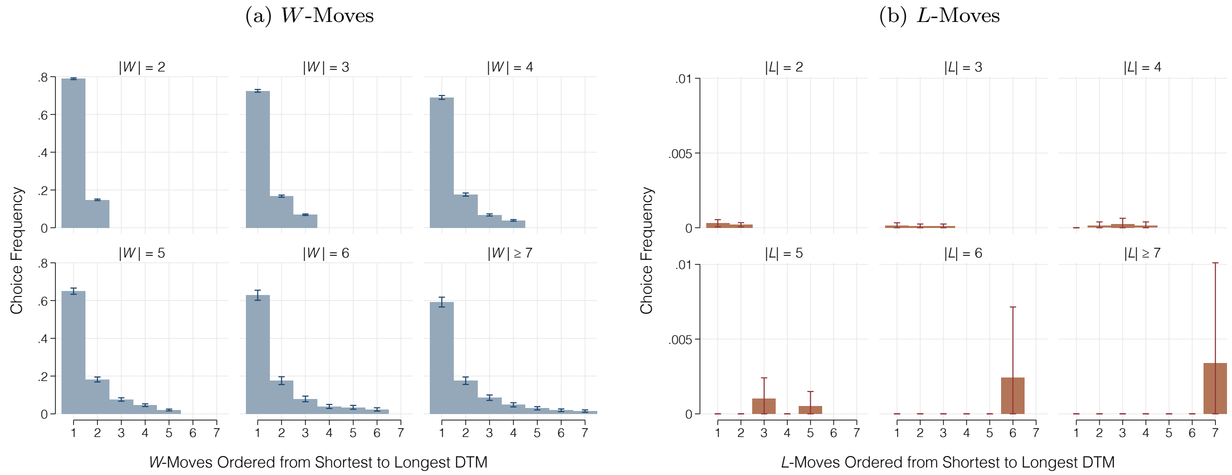
Notes: See Figure 7 in the main text. The only difference between this figure and Figure 7 is that the results above are based on data from *This Week in Chess*.

Appendix Figure AF.23: Replication of Figure 8, TWIC Data



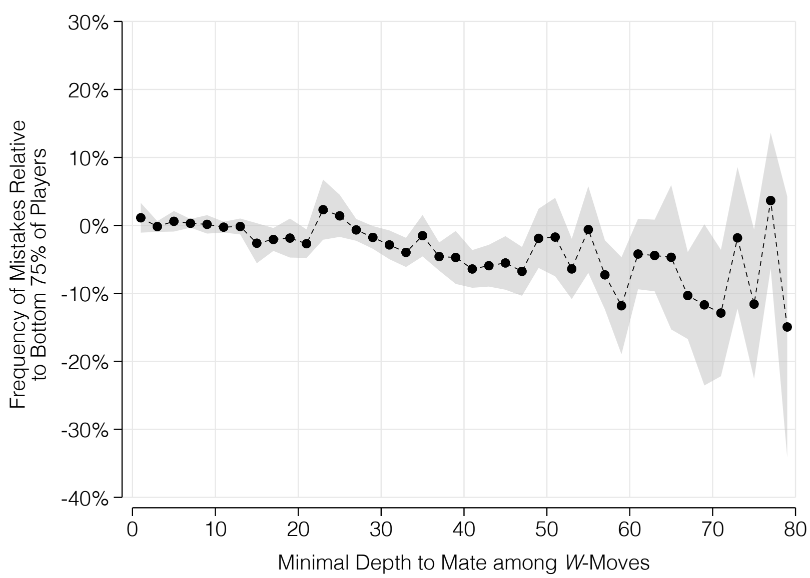
Notes: See Figure 8 in the main text. The only difference between this figure and Figure 8 is that the results above are based on data from *This Week in Chess*.

Appendix Figure AF.24: Replication of Figure 9, TWIC Data



Notes: See Figure 9 in the main text. The only difference between this figure and Figure 9 is that the results above are based on data from *This Week in Chess*.

Appendix Figure AF.25: Replication of Figure 10, TWIC Data



Notes: See Figure 10 in the main text. The main difference between this figure and Figure 10 is that the results above are based on data from *This Week in Chess*. Since the TWIC data do not consistently list players' titles, we rely on ELO ratings to differentiate players by skill. We say that a player is an "expert" if her ELO rating exceeds the 75th percentile in the TWIC data.

Appendix Tables

Appendix Table AT.1: Replication of Table 3, Without Controlling for Object Complexity

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves ($\div$ 100)	–0.307			–1.344	–1.173	
	(0.001)			(0.002)	(0.003)	
Number of D-Moves ($\div$ 100)		0.819				0.766
		(0.002)				(0.003)
Number of L-Moves ($\div$ 100)			0.361			
			(0.003)			
Fixed Effects:						
Number of D- $\times$ L-Moves	Yes	No	No	No	No	No
Number of W- $\times$ L-Moves	No	Yes	No	No	No	No
Number of W- $\times$ D-Moves	No	No	Yes	No	No	No
Total Moves $\times$ Number of L-Moves	No	No	No	Yes	No	No
Total Moves $\times$ Number of D-Moves	No	No	No	No	Yes	No
Total Moves $\times$ Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.101	0.127	0.139	0.103	0.128	0.136
N	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above do not control for the minimal DTM among W-moves.

Appendix Table AT.2: Replication of Table 2, Weighting All Observations Equally

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.171		–0.931	–0.696		–0.315
	(0.001)		(0.001)	(0.002)		(0.001)
DTM × L-Move (÷ 100)		0.024	–0.069		0.066	0.002
		(0.000)	(0.000)		(0.001)	(0.001)
W-Move (÷ 100)			51.783			28.183
			(0.035)			(0.044)
Fixed Effects:						
Composition ×						
Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$	$\mid D \mid = 0$
Mean of LHS Variable (%)	6.186	0.457	5.734	6.743	0.631	5.404
R^2	0.278	0.183	0.274	0.348	0.175	0.340
N	3,457,878,398	296,522,573	3,754,400,971	398,856,135	111,905,262	510,761,397

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above are based on equally weighted observations.

Appendix Table AT.3: Replication of Table 3, Weighting All Observations Equally

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves (÷ 100)	–0.271			–1.223	–0.985	
	(0.000)			(0.001)	(0.002)	
Number of D-Moves (÷ 100)		0.800				0.814
		(0.001)				(0.001)
Number of L-Moves (÷ 100)			0.254			
			(0.001)			
Fixed Effects:						
Number of D- × L-Moves	Yes	No	No	No	No	No
Number of W- × L-Moves	No	Yes	No	No	No	No
Number of W- × D-Moves	No	No	Yes	No	No	No
Total Moves × Number of L-Moves	No	No	No	Yes	No	No
Total Moves × Number of D-Moves	No	No	No	No	Yes	No
Total Moves × Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.105	0.130	0.140	0.106	0.129	0.138
N	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095	226,955,095

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above are based on equally weighted observations.

Appendix Table AT.4: Replication of Table 4, Weighting All Observations Equally

	(1)	(2)	(3)	(4)
	Probability of Mistake			
Titled Player (÷ 100)	–0.367		–0.005	
	(0.052)		(0.048)	
Other Title (÷ 100)		–0.342		0.006
		(0.055)		(0.050)
Grandmaster (÷ 100)		–0.589		–0.095
		(0.137)		(0.144)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	0.923	0.800
H_0: Grandmasters = Other Titled Players		0.092		0.508
Fixed Effects:				
Composition × Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	5.751	5.751	2.559	2.559
Board Configurations	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$
R^2	0.301	0.301	0.271	0.271
N	226,955,095	226,955,095	27,600,514	27,600,514

Notes: See Table 4 in the main text. The only difference between this table and Table 4 is that the results above are based on equally weighted observations.

Appendix Table AT.5: Replication of Table 2, Controlling for Time Left on Clock

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.140		–0.975	–0.518		–0.332
	(0.002)		(0.002)	(0.004)		(0.003)
DTM × L-Move (÷ 100)		0.034	0.007		0.066	0.055
		(0.001)	(0.001)		(0.002)	(0.003)
W-Move (÷ 100)			63.533			41.926
			(0.071)			(0.118)
Fixed Effects:						
Composition × Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$	$\mid D \mid = 0$
Mean of LHS Variable (%)	6.186	0.457	5.734	6.743	0.631	5.404
R^2	0.372	0.211	0.368	0.513	0.212	0.494
N	3,238,254,715	276,991,030	3,515,245,745	372,397,046	104,556,541	476,953,587

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above also control for the clock reading when it is a player's turn to move.

Appendix Table AT.6: Replication of Table 3, Controlling for Time Left on Clock

	(1)	(2)	(3)	(4)	(5)	(6)
			Probability of Mistake			
Number of W-Moves (÷ 100)	–0.281			–1.245	–0.990	
	(0.001)			(0.002)	(0.003)	
Number of D-Moves (÷ 100)		0.811				0.842
		(0.002)				(0.003)
Number of L-Moves (÷ 100)			0.247			
			(0.003)			
Fixed Effects:						
Number of D- × L-Moves	Yes	No	No	No	No	No
Number of W- × L-Moves	No	Yes	No	No	No	No
Number of W- × D-Moves	No	No	Yes	No	No	No
Total Moves × Number of L-Moves	No	No	No	Yes	No	No
Total Moves × Number of D-Moves	No	No	No	No	Yes	No
Total Moves × Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.120	0.144	0.154	0.121	0.143	0.152
N	212,295,223	212,295,223	212,295,223	212,295,223	212,295,223	212,295,223

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above also control for the clock reading when it is a player's turn to move.

Appendix Table AT.7: Replication of Table 4, Controlling for Time Left on Clock

	(1)	(2)	(3)	(4)
		Probability of Mistake		
Titled Player (÷ 100)	–0.985		–0.339	
	(0.049)		(0.063)	
Other Title (÷ 100)		–0.937		–0.312
		(0.053)		(0.065)
Grandmaster (÷ 100)		–1.318		–0.531
		(0.131)		(0.200)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	< 0.001	< 0.001
H_0: Grandmasters = Other Titled Players		0.007		0.294
Fixed Effects:				
Composition × Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	5.751	5.751	2.559	2.559
Board Configurations	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$
R^2	0.367	0.367	0.412	0.412
N	212,295,223	212,295,223	25,771,678	25,771,678

Notes: See Table 4 in the main text. The only difference between this table and Table 4 is that the results above also control for the clock reading when it is a player's turn to move.

Appendix Table AT.8: Replication of Table 2, Controlling for Time Left per Move

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.140		–0.975	–0.518		–0.332
	(0.002)		(0.002)	(0.004)		(0.003)
DTM × L-Move (÷ 100)		0.034	0.007		0.066	0.055
		(0.001)	(0.001)		(0.002)	(0.003)
W-Move (÷ 100)			63.532			41.925
			(0.071)			(0.118)
Fixed Effects:						
Composition ×						
Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$	$\mid D \mid = 0$
Mean of LHS Variable (%)	6.186	0.457	5.734	6.743	0.631	5.404
R^2	0.372	0.211	0.368	0.513	0.211	0.494
N	3,238,254,715	276,991,030	3,515,245,745	372,397,046	104,556,541	476,953,587

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Table AT.9: Replication of Table 3, Controlling for Time Left per Move

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves (÷ 100)	–0.280			–1.245	–0.989	
	(0.001)			(0.002)	(0.003)	
Number of D-Moves (÷ 100)		0.812				0.844
		(0.002)				(0.003)
Number of L-Moves (÷ 100)			0.247			
			(0.003)			
Fixed Effects:						
Number of D- × L-Moves	Yes	No	No	No	No	No
Number of W- × L-Moves	No	Yes	No	No	No	No
Number of W- × D-Moves	No	No	Yes	No	No	No
Total Moves × Number of L-Moves	No	No	No	Yes	No	No
Total Moves × Number of D-Moves	No	No	No	No	Yes	No
Total Moves × Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.119	0.144	0.154	0.121	0.143	0.151
N	212,295,223	212,295,223	212,295,223	212,295,223	212,295,223	212,295,223

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Table AT.10: Replication of Table 4, Controlling for Time Left per Move

	(1)	(2)	(3)	(4)
	Probability of Mistake			
Titled Player (÷ 100)	–0.864		–0.218	
	(0.048)		(0.063)	
Other Title (÷ 100)		–0.818		–0.193
		(0.052)		(0.065)
Grandmaster (÷ 100)		–1.188		–0.395
		(0.131)		(0.199)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	< 0.001	0.002
H_0: Grandmasters = Other Titled Players		0.008		0.332
Fixed Effects:				
Composition × Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	5.751	5.751	2.559	2.559
Board Configurations	All	All	$\mid D \mid= 0$	$\mid D \mid= 0$
R^2	0.367	0.367	0.412	0.412
N	212,295,223	212,295,223	25,771,678	25,771,678

Notes: See Table 4 in the main text. The only difference between this table and Table 4 is that the results above also control for the time that was left per move if the player were to follow the shortest W-path.

Appendix Table AT.11: Replication of Table 2, Games with Classical Time Controls Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.200		–1.085	–0.497		–0.429
	(0.014)		(0.012)	(0.018)		(0.017)
DTM × L-Move (÷ 100)		0.016	0.023		0.027	0.032
		(0.004)	(0.006)		(0.007)	(0.008)
W-Move (÷ 100)			74.413			54.774
			(0.449)			(0.601)
Fixed Effects:						
Composition ×						
Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid= 0$	$\mid D \mid= 0$	$\mid D \mid= 0$
Mean of LHS Variable (%)	10.172	0.363	9.102	11.718	0.585	8.419
R^2	0.426	0.347	0.426	0.568	0.363	0.568
N	92,278,190	8,727,277	101,005,467	10,144,036	3,015,640	13,159,676

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above restrict attention to games played under classical time controls.

Appendix Table AT.12: Replication of Table 3, Games with Classical Time Controls Only

	(1)	(2)	(3)	(4)	(5)	(6)
			Probability of Mistake			
Number of W-Moves ($\div$ 100)	–0.243			–0.922	–0.681	
	(0.003)			(0.009)	(0.015)	
Number of D-Moves ($\div$ 100)		0.558				0.702
		(0.009)				(0.016)
Number of L-Moves ($\div$ 100)			0.122			
			(0.015)			
Fixed Effects:						
Number of D- $\times$ L-Moves	Yes	No	No	No	No	No
Number of W- $\times$ L-Moves	No	Yes	No	No	No	No
Number of W- $\times$ D-Moves	No	No	Yes	No	No	No
Total Moves $\times$ Number of L-Moves	No	No	No	Yes	No	No
Total Moves $\times$ Number of D-Moves	No	No	No	No	Yes	No
Total Moves $\times$ Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.128	0.150	0.162	0.127	0.150	0.158
N	6,478,012	6,478,012	6,478,012	6,478,012	6,478,012	6,478,012

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above restrict attention to games played under classical time controls.

Appendix Table AT.13: Replication of Table 4, Games with Classical Time Controls Only

	(1)	(2)	(3)
		Probability of Mistake	
Titled Player ($\div$ 100)	–3.148		–1.405
	(0.526)		(1.009)
Other Title ($\div$ 100)		–2.844	
		(0.479)	
Grandmaster ($\div$ 100)		–11.386	
		(3.199)	
Hypothesis Tests (p-value):			
H_0: No Differences between Players	< 0.001	< 0.001	0.164
H_0: Grandmasters = Other Titled Players		0.008	
Fixed Effects:			
Composition $\times$ Complexity of Moves	Yes	Yes	Yes
Mean of LHS Variable (%)	4.547	4.547	1.221
Board Configurations	All	All	$\mid D \mid = 0$
R^2	0.418	0.418	0.678
N	6,478,012	6,478,012	730,547

Notes: See Table 4 in the main text. The main difference between this table and Table 4 is that the results above restrict attention to games played under classical time controls. In contrast to Table 4, board positions without D-moves do not allow us to separately identify coefficients for gransmasters and other titled players because there is no within-variation in the grandmaster indicator conditional on composition $\times$ complexity fixed effects. For this reason, the table above does not replicate col. (4) of its counterpart in the main text.

Appendix Table AT.14: Replication of Table 2, Board Positions with High Minimal DTM Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–0.450		–0.357	–0.474		–0.318
	(0.004)		(0.003)	(0.009)		(0.008)
DTM × L-Move (÷ 100)		0.034	0.141		0.053	0.168
		(0.003)	(0.004)		(0.007)	(0.009)
W-Move (÷ 100)			55.228			56.068
			(0.281)			(0.634)
Fixed Effects:						
Composition ×						
Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$	$\mid D \mid = 0$
Mean of LHS Variable (%)	15.365	0.795	11.064	13.552	1.295	9.381
R^2	0.379	0.330	0.394	0.435	0.387	0.443
N	92,066,019	39,282,812	131,348,831	22,051,140	9,811,399	31,862,539

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

Appendix Table AT.15: Replication of Table 3, Board Positions with High Minimal DTM Only

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves (÷ 100)	–1.465			–2.986	–1.931	
	(0.005)			(0.006)	(0.010)	
Number of D-Moves (÷ 100)		1.524				1.338
		(0.006)				(0.010)
Number of L-Moves (÷ 100)			0.235			
			(0.009)			
Fixed Effects:						
Number of D- × L-Moves	Yes	No	No	No	No	No
Number of W- × L-Moves	No	Yes	No	No	No	No
Number of W- × D-Moves	No	No	Yes	No	No	No
Total Moves × Number of L-Moves	No	No	No	Yes	No	No
Total Moves × Number of D-Moves	No	No	No	No	Yes	No
Total Moves × Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.224	0.237	0.247	0.224	0.237	0.240
N	12,927,786	12,927,786	12,927,786	12,927,786	12,927,786	12,927,786

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

Appendix Table AT.16: Replication of Table 4, Board Positions with High Minimal DTM Only

	(1)	(2)	(3)	(4)
	Probability of Mistake			
Titled Player (÷ 100)	–2.228		–0.032	
	(0.331)		(0.239)	
Other Title (÷ 100)		–2.052		–0.272
		(0.367)		(0.177)
Grandmaster (÷ 100)		–3.342		1.436
		(0.663)		(1.271)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	0.894	0.157
H_0: Grandmasters = Other Titled Players		0.089		0.182
Fixed Effects:				
Composition × Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	20.525	20.525	3.593	3.593
Board Configurations	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$
R^2	0.470	0.470	0.487	0.487
N	12,927,786	12,927,786	2,121,748	2,121,748

Notes: See Table 4 in the main text. The main difference between this table and Table 4 is that the results above restrict attention board configurations in which the DTM of the shortest W-path exceeds 50.

Appendix Table AT.17: Replication of Table 2, TWIC Data

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Choosing Move					
	Only W-Moves	Only L-Moves	W- & L-Moves	Only W-Moves	Only L-Moves	W- & L-Moves
DTM × W-Move (÷ 100)	–1.121		–0.977	–1.052		–0.805
	(0.018)		(0.014)	(0.077)		(0.051)
DTM × L-Move (÷ 100)		–0.001	0.052		–0.000	0.042
		(0.002)	(0.014)		(0.001)	(0.028)
W-Move (÷ 100)			82.040			70.528
			(0.816)			(2.669)
Fixed Effects:						
Composition ×						
Complexity of Other Alternatives	Yes	Yes	Yes	Yes	Yes	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Board Configurations	All	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$	$\mid D \mid = 0$
Mean of LHS Variable (%)	10.155	0.025	8.265	8.218	0.039	6.109
R^2	0.445	0.335	0.450	0.408	0.349	0.422
N	5,132,343	1,177,248	6,309,591	853,980	296,615	1,150,595

Notes: See Table 2 in the main text. The only difference between this table and Table 2 is that the results above are based on data from *This Week in Chess.*

Appendix Table AT.18: Replication of Table 3, TWIC Data

	(1)	(2)	(3)	(4)	(5)	(6)
	Probability of Mistake					
Number of W-Moves ($\div$ 100)	–0.168			–0.588	–0.256	
	(0.004)			(0.009)	(0.010)	
Number of D-Moves ($\div$ 100)		0.396				0.352
		(0.009)				(0.012)
Number of L-Moves ($\div$ 100)			0.044			
			(0.010)			
Fixed Effects:						
Number of D- $\times$ L-Moves	Yes	No	No	No	No	No
Number of W- $\times$ L-Moves	No	Yes	No	No	No	No
Number of W- $\times$ D-Moves	No	No	Yes	No	No	No
Total Moves $\times$ Number of L-Moves	No	No	No	Yes	No	No
Total Moves $\times$ Number of D-Moves	No	No	No	No	Yes	No
Total Moves $\times$ Number of W-Moves	No	No	No	No	No	Yes
Player	Yes	Yes	Yes	Yes	Yes	Yes
Minimal DTM among W-Moves	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.158	0.162	0.169	0.154	0.162	0.168
N	536,674	536,674	536,674	536,674	536,674	536,674

Notes: See Table 3 in the main text. The only difference between this table and Table 3 is that the results above are based on data from *This Week in Chess*.

Appendix Table AT.19: Replication of Table 4, TWIC Data

	(1)	(2)	(3)	(4)
	Probability of Mistake			
Top 25% of Players ($\div$ 100)	–3.260		–0.032	
	(0.336)		(0.032)	
75th to 99th Percentile of Players ($\div$ 100)		–3.184		–0.031
		(0.342)		(0.033)
Top 1% of Players ($\div$ 100)		–4.873		–0.051
		(0.605)		(0.046)
Hypothesis Tests (p-value):				
H_0: No Differences between Players	< 0.001	< 0.001	0.313	0.427
H_0: Grandmasters = Other Titled Players		0.005		0.683
Fixed Effects:				
Composition $\times$ Complexity of Moves	Yes	Yes	Yes	Yes
Mean of LHS Variable (%)	5.133	5.133	0.786	0.786
Board Configurations	All	All	$\mid D \mid = 0$	$\mid D \mid = 0$
R^2	0.520	0.520	0.650	0.650
N	499,331	499,331	65,113	65,113

Notes: See Table 4 in the main text. The main difference between this table and Table 2 is that the results above are based on data from *This Week in Chess*. Since the TWIC data do not consistently list players' titles, we rely on ELO ratings to differentiate players by skill.