Gutenberg School of Management and Economics

& Research Unit “Interdisciplinary Public Policy”

Discussion Paper Series

Time Pressure and Strategic Risk-Taking in Professional Chess

Johannes Carow, Niklas M. Witzig

February 22, 2024

Discussion paper number 2404

Johannes Gutenberg University Mainz
Gutenberg School of Management and Economics
Jakob-Welder-Weg 9
55128 Mainz
Germany
https://wiwi.uni-mainz.de/

Contact details

Johannes Carow
Chair of Applied Statistics and Econometrics and Ministry for Wirtschaft, Verkehr, Landwirtschaft und Weinbau, Rheinland-Pfalz
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany
johannes.carow@uni-mainz.de

Niklas M. Witzig
Chair of Public and Behavioral Economics
Johannes Gutenberg University Mainz
Jakob-Welder-Weg 4
55128 Mainz
Germany
niklas.witzig@uni-mainz.de

Time Pressure and Strategic Risk-Taking in Professional Chess*

Johannes Carow§ **Niklas M. Witzig**¶

February 22, 2024

Abstract

We study the impact of time pressure on strategic risk-taking of professional chess players. We propose a novel machine-learning-based measure for the degree of strategic risk of a single chess move and apply this measure to the 2013-2023 FIDE Chess World Cups that allow for plausibly exogenous variation in thinking time. Our results indicate that time pressure leads chess players to opt for more risk-averse moves. We additionally provide correlational evidence for strategic loss aversion, a tendency for risky moves after a mistake/ in a disadvantageous position. This suggests that high-proficiency decision-makers in high-stake situations react to time pressure and contextual factors more broadly. We discuss the origins and implication of this finding in our setting.

JEL classification C26, C45, D91

Keywords: Chess, Risk, Time Pressure, Loss Aversion, Machine Learning

*We thank Raphael Brade, Julia Engel, Katharina Hartinger, Greg Howard, Thorsten Schank, Daniel Schunk, Christian Seel and seminar participants at JGU Mainz as well as participants of the FUR 2022 conference for their helpful comments. We thank Chess24 and ChessBase for access to their extensive databases. We also thank Anna Dreber, Christer Gerdes and Patrik Gränsmark for sharing their classification of chess openings as solid, neutral or aggressive with us. The authors gratefully acknowledge the computing time granted on the supercomputer MOGON 2 at Johannes Gutenberg-University Mainz (hpc.uni-mainz.de). The views expressed in this paper are those of the authors and do not necessarily reflect those of the MWVLW RLP.

§JGU Mainz and MWVLW RLP. johannes.carow@uni-mainz.de

¶JGU Mainz. niklas.witzig@uni-mainz.de

1 Introduction

Highly strategic decisions often have to be made under considerable time pressure. Imagine financial traders quickly responding to changing market situations, auctioneers in close competition bidding against each other, or negotiators trying to reach a consensus close to a deadline. Many professional domains require strategic decision-making under tight time constraints. How time pressure affects strategic decision-making has therefore attracted the attention of economists. Prior research shows that time pressure leads to less strategic voting (Alós-Ferrer and Garagnani, 2022), less complex decision rules in normal-form games (Spiliopoulos et al., 2018), reduced market entry (Lindner, 2014), lower bids in auction games (Haji et al., 2019) and more disagreements in bargaining (Karagözoglu and Kocher, 2019). Time pressure thus appears to induce more cautious, simpler and less strategic choices.[1] The impact of time pressure on *nonstrategic* (i.e., not involving another player) decision-making, in particular risk-taking, has produced more mixed evidence: Kocher et al. (2013) document increased risk aversion under time pressure in the loss domain. Additionally, Wu et al. (2022) find that subjects in their lab experiment choose high-variance options less often when time pressure is introduced, indicating a shift towards risk-aversion under time constraints. In contrast, Olschewski and Rieskamp (2021) suggest that time pressure coincides with more frequent choices of risky gambles which is, however, driven by choice inconsistency and not risk preferences. Kirchler et al. (2017) present mixed evidence, indicating that time pressure increases risk-aversion for gains and risk-seeking for losses. Ultimately, time pressure does impact non-strategic risk-taking, albeit less consistently compared to strategic domains.

A common feature in the aforementioned studies across strategic and non-strategic domains is their reliance on *laboratory settings*. There, neither the proficiency of decision-makers nor the stakes of are typically very high. Whether the mentioned findings translate into high-stake settings with highly proficient and experienced decision-makers who regularly decide under time pressure still remains an open question. Drawing this distinction is important as an exclusive focus on laboratory data might conceal how time pressure affects proficient decision-makers – like those in many professional settings – and lead to inaccurate inferences, ultimately rendering the conclusions less relevant to potential policy implications.

Studying how expert decision-makers handle time pressure situations is, however, very difficult. Both getting high-quality data of their decisions and isolating the impact of time pressure on these decisions is an extremely challenging task. We address both aspects by turning to an ideal setting for our research question: professional chess. Chess is often credited as *the* strategic game, frequently serves as an analogy to strategic management (see, e.g. Saloner (1991)) and chess players are regularly credited for a remarkable level of skill, expertise and intelligence.[2] Chess requires highly strategic considerations by both players, selecting between

[1]Note that Lindner and Sutter (2013) find "perhaps incidentally" more closer-to-equilibrium level-k thinking in an 11-20 request game, and Kocher and Sutter (2006) find no deterioration in decision quality in a beauty contest under time pressure.

[2]There is still an ongoing scientific discussion if there is actually a relationship between chess skill and general intelligence, which produced mixed evidence so far (see e.g. Bilalić et al. (2007)).

numerous strategies, anticipating possible responses by the opponent to each move and trying to decide which move will lead to the most advantageous outcome. In addition, *time* is a key component of chess as chess is played with different time limits like classic or Blitz. Sigman et al. (2010) highlight how time pressure can prevent chess players from calculating long variations, which is, however, key to deal with the wide range of possible move sequences resulting from a current board position. Assessing how to allocate time efficiently thus represents a cornerstone for chess mastery. Time usage within chess games is recorded with very high precision thanks to digital clocks showing the remaining time for each player next to the chess boards. Professional chess is thus the ideal setting to study high-stake strategic decisions by very proficient decision makers under time pressure.

In this paper, we study the move-selection of chess players during the six FIDE World Cups 2013-2023. The World Cup assembles the strongest chess players in the world, offering generous monetary prizes. It is part of the so-called World Championship Cycle, a series of tournaments determining who challenges the current World Champion for the title. The World Cups are furthermore — this is a specialty among chess tournaments on that level — organized as *knock-out*-tournaments, where two players in each match determine the winner who proceeds to the next round, which further increases the stakes of each match. But crucially, the World Cups provide the following feature: In case of a tiebreak in a given match, players play another game against each other *in a tighter time control*, i.e. with a reduced initial time budget for each player. If reaching a tiebreak is exogenous to the outcome under consideration – risk-taking in our analysis –, this allows for exogenous variation in available thinking time. We apply a rich specification where we rule out any confounding factors that are constant per player-match-game-combination. By instrumenting remaining time with the initial time of each time control, interacted with half-move dummies[3], we attempt to unveil the causal effect of time pressure on risk-taking by professional chess players.

Building on the theoretical foundation from Calford (2020), we first develop a measure for the degree of revealed strategic risk of a single chess move. This measure is based on the *variance* of outcomes that might occur after the opponent's response to a given move. We interpret larger variance indicating higher strategic risk as the subsequent outcome depends more strongly on the opponent's action. This measure allows us to study move-by-move decision-making of chess players and importantly how the remaining time budget impacts these sequential decisions. We consistently find that time pressure leads to a systematically different selection of strategies towards safer strategies, i.e., that chess player behave more risk-averse under time pressure. This finding is robust across subsamples based on the strength of players or the initial time budget and is particularly pronounced for games in which a player has to win to stay in the competition. Importantly, the results remain stable when applying different weighting schemes of the opponent's responses in the computation of the risk measure or when considering varying numbers of considered move options from the opponent. In addition, we find correlational support for strategic loss aversion, indicated by a higher tendency to play risky moves after a mistake and in disadvantageous positions. Both are important findings as we highlight that time

[3] The concept of half-moves is explained in Sections 2.1 and 4.

pressure and loss frames can shift behavior by highly proficient decision-makers in high-stake environments. This reiterates on the previously mentioned notion that time pressure impacts strategic decision-making, in fact in similar ways for highly proficient decision-makers as for more lay samples. While we discuss the peculiarities of chess as a setting and how to actually interpret (the origins of) this effect, our results suggest that time pressure has a general impact on strategic decision making towards safer choices.

With this paper, we contribute to several literatures: First, to an economic literature that uses chess settings to study risk-taking decisions: Notably, with our measure, we can unveil dynamic risk patterns *within games*, which in turn advances existing approaches on risk in chess that are based on categorizing chess openings or entire games: Dreber et al. (2013) provide a seminal contribution with a classification system of 500 chess openings as "risky", "neutral" or "safe", based on surveying chess experts. They argue that the choice of a specific opening is a major driver for the outcome of the game. Risky openings are supposed to increase both the winning and losing probability, making a draw less likely. Consequently, Dreber et al. (2013) state that players reveal their risk preferences by choosing an opening of a specific type. Similarly, Chassy and Gobet (2015) as well as Chassy and Gobet (2020) classify opening strategies as risky where the empirical variance of the outcome of the game (win, lose or draw) is high and focus on whether the first move is 1.e4 ("risky") or 1.d4 ("conservative"). Their measure is thus based on past games rather than expert opinions. Klingen and van Ommeren (2020) interpret a high frequency of drawing games as risk-averse behavior of players. All these approaches have in common that the resulting risk measures vary per *game*, not on the *move-by-move-level*. Consequently, the scope of previous analyzes was limited to compare risk between males and females (Gerdes and Gränsmark, 2010; Dreber et al., 2013; Dilmaghani, 2021), between different time controls (Dilmaghani, 2021) or between games with varying levels of air pollution (Klingen and van Ommeren, 2020). Our risk measure is more flexible than, yet consistent with the previous classification from Dreber et al. (2013): We show that our move-specific measure strongly correlates with the existing classification of openings (assuming that risky openings beget risky play to some extent). Our measure thus allows studying dynamic developments of risky play in chess beyond the chosen openings or the game's eventual outcome. Consequently, it can offer insights on how players vary their risk-taking behavior in response to increasing time pressure. This type of analysis would not be possible when defining risk only depending on the opening of each game. Another novelty of our measure is its basis on the *difficulty of a chess position* - the measurement of which we outline below. This difficulty measure is based on machine-learning techniques applied to a large database of human chess games to assess in which positions humans are likely to make mistakes, based on past games. This puts a strong emphasis on how humans play chess and our difficulty measure is directly transportable to other studies.

Further, investigating the link between time pressure and risk-taking enriches the literature covering the effects of time pressure in chess: Sigman et al. (2010) highlight the importance of the remaining time budget: They find that it is a central determinant for the winning probability in addition to the position evaluation. Gränsmark (2012) links time allocation of chess players to both time preferences and time inconsistencies. He documents different patterns of time inconsistencies across genders, stating that males under-consume and females over-consume thinking time in the beginning of a chess game. Howard (2024) tests for rational time allocation for players of different proficiency levels. He documents a trade-off between searching for a better move and avoiding excessive time consumption which could be harmful for later stages of the game. Using a very similar IV strategy, his findings indicate that skilled players are better able to deal with the trade-off than weaker players and that an existing intervention from an online chess platform can improve time allocation. Similarly, Russek et al. (2022) show that stronger players consume more time in situations where calculating is actually beneficial. Künn et al. (2023) report that time pressure reinforces the adverse effect of air pollution on the quality of played moves.

More broadly, this paper also contributes to a growing and diverse literature that uses chess to study a plethora of economic phenomena. For instance, studies consider chess to uncover determinants for cognitive performance such as a positive emotional state (González-Díaz and Palacios-Huerta, 2016), remote work (Künn et al., 2022), air quality (Künn et al., 2023) or the position on the life cycle (Strittmatter et al., 2020). Further research discusses whether chess players apply a more rational decision-making than non-chess-players (Palacios-Huerta and Volij, 2009; Levitt et al., 2011). Additionally, Salant and Spenkuch (2022) use a chess setting to investigate satisficing in the presence of complex choices. Linnemer and Visser (2016) study the self-selection of chess players into tournament sections based on their strength.

The remainder of the paper proceeds as follows. Section 2 provides the conceptual background of our measure of revealed strategic risk in chess. We discuss the relevant empirical quantities for our measure, including the difficulty of a chess position, in detail. Afterwards, Section 3 describes the FIDE Chess World Cups, the key dataset in this study. In Section 4, we present the empirical specification and discuss the identifying assumptions for exogenous variation in available thinking time. We then present our empirical results in Section 5 as well as a discussion in Section 6. Finally, Section 7 concludes.

2 Measuring Strategic Risk in Chess

A key part of our contribution is a proposition for measuring the degree of strategic risk entailed *in a single chess move*. In this section, we first lay out the conceptual foundations for our measure before discussing the empirical quantities necessary for its construction. We conclude with a concrete example from chess.

2.1 conceptualization

For conceptualizing strategic risk-taking, we closely follow the discussions in Calford (2020). We understand strategic risk as the uncertainty which strategy an opponent will implement.

To illustrate how (revealed) risk-preferences can influence strategic decision-making, consider Table 1. This table depicts a slightly adapted version of the illustrative example in Calford (2020). Here – closely following the wording in Calford (2020) – two players (row and column) engage in a simple normal-form game. The row player can select among three strategies A, B and C, whereas the column player can select between X and Y. For the row player, strategy C is dominated by a mixture of A and B. Yet, C becomes a justifiable strategy for the row player with sufficient risk aversion: Assuming a belief of equal probability of the opponent to play X or Y, the relation $2 \times u(18) > u(25) + u(14)$ might hold because of classical *risk* preferences due to a curvature of utility functions. Note that also *ambiguity*-aversion might induce the row player to choose C. We present a more extensive explanation of that argument in Appendix Section A.3.6.[4]

Table 1: A simple game

	X	Y	σ^2_{row}
A	25,20	14,12	30.25
B	14,20	25,12	30.25
C	18,12	18,22	0

Source: Adapted from Calford (2020)

We can attach a quantitative notion to this motive: Based on the assumption of equal probability of the opponents' actions, the *variance of outcomes*, σ^2, is instructive of the revealed degree of strategic risk entailed in a given strategy. Intuitively, if a given strategy results in the same or a similar outcome no matter how the opponent responds, that action's strategic risk can be considered low. Accordingly, a risk-averse agent will seek to minimize the variance of the consequences of his available strategies, which is fulfilled by implementing C in Table 1. Conditional on the resulting variance of alternative strategies (in the above case 30.25 for X and Y), σ^2_s, therefore, allows for a measure of the revealed risk aversion/seeking of an agent if strategy s is implemented.

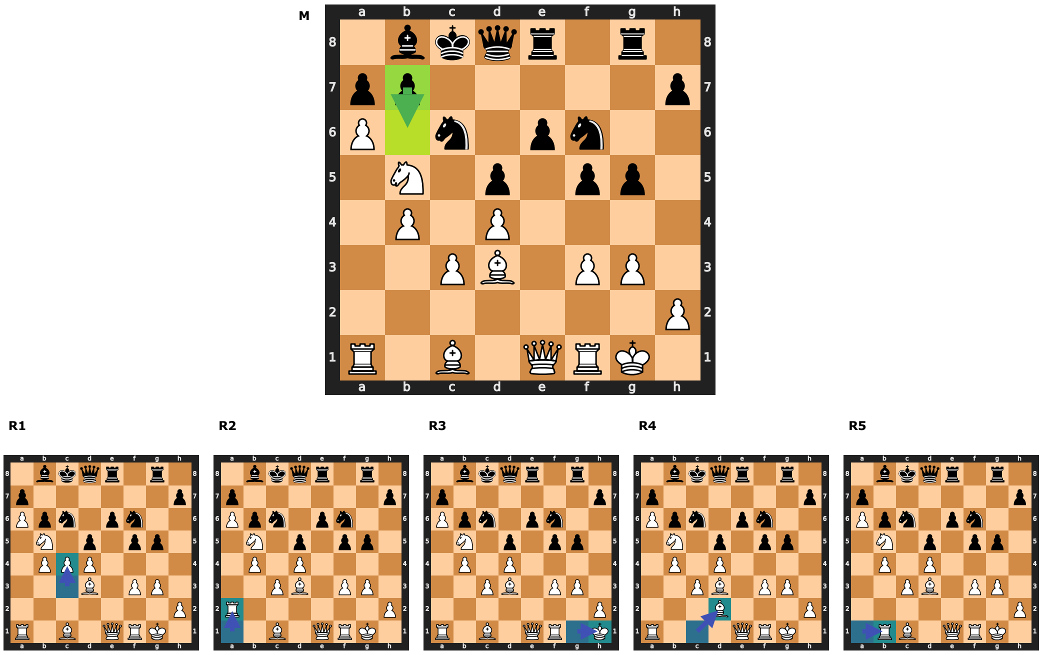
We analyze chess moves through this lens: Figure 1 depicts a situation (i.e. board position) of the game Brkic - Barrientos from the World Cup 2021 after White's 21st move 21. a5-a6. Black decided to play 21... b7-b6 highlighted in green. This is the strategy or the *half-move* h (in chess terms, one move consists of a half-move by White and a half-move by Black[5]) which Black selected out of several possible options. Relevant for the assessment of the revealed degree of strategic risk of this half-move is (i) the set of available half-moves H for Black and (ii) the set of potential responses by White R_h to each $h \in H$. Both consideration sets are not directly observable, of course, but the high proficiency of our sample allows us to make educated guesses about these sets in the empirical analysis. For the computation of the risk measure, we first focus only on the half-move b6 and assume that Black considered the responses R1-R5 by White (more details on how we identify R1-R5 follow), which are highlighted in blue. Each response by White, therefore, leads to a unique new position for Black and Black can influence what type

[4]In Appendix Section A.3.6, we also show how we can capture ambiguity attitudes in our chess setting. We repeat the empirical analyzes and largely confirm the main findings based on the risk measure.

[5]The distinction between moves and half-moves is mainly relevant in Sections 4 and thereafter.

of positions will emerge in the game by selecting a given half-move, in this case, 21... b7-b6.

Figure 1: Illustrative example for the computation of the risk measure: game Brkic vs Barrientos, FIDE World Cup 2021, position after White's move 21. a5-a6



Note: The upper panel of the figure shows a position that occurred in the game Brkic - Barrientos, FIDE World Cup 2021. In standard chess notation, White's last half-move was to advance the pawn to a6 (21. a5-a6). In the diagram position, Black reacted by playing 21... b7-b6. In the resulting position after Black's half-move, the 5 best responses (shown in the smaller diagrams below) for White, according to the chess engine Stockfish 11, are 22. c3-c4 (R1), 22. Ra1-a2 (R2), 22. Kg1-h1 (R3), 22. Bc1-d2 (R4) and 22. Ra1-b1 (R5). Our measure as defined in Section 2.3 assumes that Black anticipated R1-R5 before deciding on 21... b7-b6 versus alternatives, to which Black considered the best responses respectively.

This leads to the key intuition behind our later measure: We assume that Black, when selecting a half-move out of several possible options, *anticipates* the most likely responses by White to each half-move and subjectively assesses[6] which of these positions is most favorable to play.

This opens up the possibility for risk attitudes to play a role: Assume that Black decides between making two different half-moves. Each response by White to the first does not change Black's subjective assessment of the subsequent positions. In contrast, if Black plays the second half-move, White could either considerably worsen Black's position via a specific response or allow Black to further increase their advantage if White fails to implement this response. Thus, for the second half-move, the assessment of the corresponding position strongly depends on White's response, whereas this is not the case for the first move. Intuitively, this implies that the first half-move can be considered safe, whereas the second is more risky. Chess half-moves thus can mirror the considerations from the introductory normal-form game. Like the row player in Table 1, Black in Figure 1 must assess the consequences of his actions given responses by White before deciding on a given half-move. Given sets of H and R and an appropriate assessment metric, we can thus repeat the proposed quantitative measures of the degree of strategic risk of

[6]The assessment of corresponding positions – the analogy to payoffs in normal form games – and which aspects of positions are relevant to human chess players are discussed in detail in Section 2.2.

strategies in normal-form games for single chess moves.

Note that this neglects considerations about the risk of potential *future moves*. Chess players – especially highly proficient ones – often try to anticipate a series of half-moves into the future and decide accordingly. While our risk measure could be expanded to consider the positions and risk of the corresponding options after k half-moves, we abstract from this mostly for tractability reasons. However, as Howard (2024) discusses, considering many moves into the future will often not be beneficial for a rational decision-maker, as the marginal benefit of thinking about a continuation that may not realize is small compared to the costly time in chess.

2.2 Measuring Relevant Outcomes in Chess

Key for applying our conceptualization to real chess games is to obtain empirical quantities of relevant outcomes — the analogy of the payoffs from Table 1 — human players consider during play. As stated, we consider *position evaluations* as outcomes, e.g. considerations how favorable a given position is for a player. These outcomes then build the basis for measuring the degree of strategic risk of a given half-move. We propose that players consider two distinct outcomes: the (i) relative (dis-)advantage of a position as well as the (ii) *difficulty* of a position.

Relative (dis-)advantage of a position The first candidate for a relevant outcome is the relative (dis-)advantage in a position. Players try to find moves that shift the board to their advantage. Many factors, such as the material (dis)advantage or the king's safety, contribute to this outcome. Different chess players may weight these position characteristics differently and, hence, evaluate the (dis-)advantage in a given position differently. Nonetheless, an objective measure of the relative (dis-)advantage is available, namely via chess engines. Chess engines are computer programs that consistently outplay humans for a long time and are generally considered rational benchmarks, e.g. for selecting optimal moves in a given position (Anderson et al., 2017; Zegners et al., 2020; Sunde et al., 2022). We provide more details on how chess engines work in Section A.1, but the key takeaway is that chess engines can evaluate positions and measure the (dis)advantage of one side in terms of *centipawns (CP)*. The unit *centipawns* measures how many hundredths of a pawn (i.e. the weakest piece in chess) are equivalent to the (dis-)advantage of a given side compared to the opponent if both sides play optimally from this position onwards for a given length of the continuation. The evaluation function of a chess engine takes different aspects of board positions, such as the activity of pieces, into account, which also differs between chess engines. We use Stockfish 11, the most widely used open-source chess engine. To provide an example for a Stockfish evaluation, consider the initial position in Figure 1, where Stockfish assigns an advantage of 77 CP for White. This corresponds to a moderate, albeit not decisive advantage.

It may appear counter-intuitive to use engine evaluations, given that we are interested in human evaluations of chess positions and human players do not have access to an engine during play. However, since we only consider chess games in the FIDE World Cup, in which predominantly grandmasters (the highest level of chess titles) participate[7], we can *restrict* a chess engine and approximately match its playing strength to those of players in our sample. Specifically, we allow Stockfish only to explore continuations of *depth 15* for its evaluation of a board position, which corresponds to the playing strength of grandmasters (Ferreira, 2013) and thus implies that the responses and moves considered by the restricted engine should be in line with what a human grandmaster considers to be likely and relevant. Thus, our sample allows us to approximate the relative (dis-)advantage of a position a human grandmaster would estimate by evaluating that position with a restricted chess engine.

Difficulty of a position However, when evaluating a given position, human players will not only consider the relative (dis-)advantage of that position: A position may yield an advantage for one side, consequently being evaluated as such by a chess engine, yet a chess player would still try to avoid getting into this positions, because of its *difficulty*. While there is no general consensus among chess players what makes a position difficult, it is considered to be an important part of the game. Magnus Carlsen — World Champion between 2013 and 2023 — allegedly said, "I am trying to beat the guy sitting across from me and trying to choose the moves that are most unpleasant for him".[8] Chess players, therefore, need to keep watch of emerging situations, in which their opponent may be able to make very unpleasant moves.

Taking the difficulty of a position into account for our measure of strategic risk requires additional information about positions beyond the evaluation by a restricted chess engine. Yet, how the difficulty of a chess position can be measured is far from trivial, and no established way how to do this exists. What makes this more complicated is the large number of possible positions in chess[9], implying that the vast majority of games enter unknown territory. Thus, we need a way to measure the difficulty of a chess position that no human player has ever played.

We take a data-driven approach which allows us to propose a way to measure the difficulty of *any* given chess position for a human player based on historical playing data. We gathered data from approximately 2 million chess games from professional chess formats, which are played by chess players with a FIDE-Elo rating (the official chess rating system) of at least 2400, thus focusing on the grandmaster level. The games are from various formats, such as open tournaments, championships or Olympiads, dating back to 1975. We downloaded this data from ChessBase (2021). For each game, we use Stockfish to evaluate each move's (arguably objective) quality, which leads to a total of around 100 million moves with their respective quality rating in our database. The quality of a move is again measured in CP and equals to zero if the objectively best move (or an equivalent) was played in that position. The more this measure deviates from zero, the worse is the move. A move with a quality rating of 100 CP (i.e. deviating a full pawn from the optimal move) can be considered to be a serious blunder. For reference, the evaluation

[7] The sample mean of the Elo rating (the official chess rating system) of the players in our sample amounts to 2674 which would correspond approximately to World rank 56 as of February 2024, see `https://2700chess.com/`, last accessed on February 13, 2024.

[8] `https://www.chessable.com/blog/magnus-carlsen-quotes-chess/`.

[9] Steinerberger (2015) calculates the number of possible positions on the chess boards without any pawn being promoted to be 10^{40}.

of b7-b6 in Figure 1 is 115 CP, which indicates a sizable mistake by Black. In contrast, the optimal move suggested by Stockfish is f5-f4, attacking the White king (see Section A.1 for more details).

With our full dataset of 2 million games, we develop a difficulty measure which links the quality of a human chess move to the board position: A long-standing research documents that identifying patterns, i.e. configurations of multiple pieces on the board, constitutes an integral part of human chess expertise (de Groot, 1965). These patterns can help chess experts to navigate through the enormous variety of possible chess positions. If the player is not familiar with the type of structures on the board, deviations from the best possible move are more likely to be played. This suggests that the *expected quality of a given human move is a function of the board position*. However, the nature of this function is likely to be very complex and possibly infeasible to identify.

To overcome this issue, we employ supervised machine learning techniques. Given sufficient data, supervised machine learning, such as deep artificial neural networks, perform exceptionally well at function approximation, known as the "Universal approximation theorem" (Heinecke et al., 2020; Zhou, 2020; Goodfellow et al., 2016). We take inspiration from this notion: We train a convolutional neural network (CNN) on our dataset to predict the quality of a human move in a given chess position, as measured by Stockfish, using *only the configuration of the chess board as an input*. We provide details on how we encode a board position into numerical features, the exact model architecture, and the training of the neural network, including a quality comparisons with other predictive models (which speaks in favor of the CNN architecture), in Section A.2. Ultimately, our neural network predicts the expected deviation from the optimal move in a particular chess position (measured in CP) if a proficient human player plays that position.[10] The difficulty of the initial position in Figure 1 is predicted to be 44.93 CP for Black, which indicates a moderately-difficult position. This matches chess intuition as the White and the Black king have castled to opposite sides, motivating both players to launch an attack on the opponent's king; In addition, White's last move 21. a5-a6 intends to weaken the position of Black's king. If a human player would play that position, f5-f4, the optimal move, would seldom be played. For further examples of low, medium and high difficulty, see Figure A6.

Our difficulty measure, although reliant on chess engines in its construction, thus entails a strong *human perspective* on playing chess. This advances existing, purely engine-based approaches. Guid and Bratko (2006) propose to measure the difficulty of a chess position by the time an engine takes to find the optimal move. Similarly, Sunde et al. (2022) use the amount of nodes considered by an engine to find the best moves in a given position as measure of position difficulty. In contrast to these approaches, we ask how large the deviation from the optimal move(s) would be if a human played that position.

[10] An interactive website of our model is available at `https://nmwitzig.github.io/chess-app.html`.

2.3 From Outcomes to Strategic Risk: Example and Definitions

The difficulty of a position serves as a basis for the risk measure proposed in Section 2.1. This, however, requires assumptions on the anticipated responses by the opponent to the half-move played, which are necessary to compute the variance of difficulty as a risk measure for the half-move played.

To gather data on anticipated responses, we make use of chess engines restricted to a search depth of 15 half-moves (see Section 2.2 for the link between search depth and playing strength). For the move played in the game, we obtain the opponent's 5 best responses R1-R5 as suggested by the chess engine. While we cannot know the exact half-moves a player considered and anticipated, examining five best half-moves and responses (according to the restricted engine) appears to be a reasonable range from which grandmasters pick.[11] To rule out that implausible responses of the opponent dilute our measure, we exclude responses that are more than 300 CP worse than the opponent's best response.[12] We also repeat our analyzes varying the number of the opponent's responses under consideration to be between 3-7.

We can now demonstrate the computation of the risk measure based on the example in Figure 1 (see Section A.1 and the corresponding Figure A1 for chess-specific explanations of this position.). After White's move 21. a5-a6, Black faces a choice how to react to the advance by the pawn to a6. In the game, Black responded by playing 21... b7-b6, highlighted in green in the upper board of Figure 1. To calculate the risk of 21... b7-b6, we consider the five best responses by White to that move (shown in the lower part of Figure 1): 22. c3-c4, 22. Ra1-a2, 22. Kg1-h1, 22. Bc1-d2 and 22. Ra1-b1. We call the resulting board positions *hypothetical* since at most one of them materializes in the game. This set of hypothetical boards permits the computation of our risk measure: The variance of difficulty computed over the five boards amounts to 23.

In the position of Figure 1, Black chose the fifth best move 21... b7-b6 over the first best move 21... f5-f4. Considering the optimal move 21... f5-f4 goes along with different best responses by the White player: 22. Ra1-a2, 22. Kg1-g2, 22. Qe1-f2, 22. g3xf4 as well as 22. c3-c4 (see Figure B1). Now, the variance of difficulty becomes considerably larger and amounts to 445, indicating a higher risk of 21... f5-f4.

This highlights several aspects: By playing 21... b7-b6 instead of the objectively strongest move 21... f5-f4, Black worsens his position. The move 21... f5-f4, however, coincides with a higher variance of difficulty and the subsequent difficulty for Black depends more strongly on White's response to 21... f5-f4 compared to 21... b7-b6. Thus, playing the strongest move comes at a cost of being exposed to more variability in the position difficulty.

[11] Table 2 shows that in the Classic mode of the FIDE World Cup sample, 48 percent of half-moves in the sample are the first best half-moves while 91 percent of all moves are among the five best half-moves.

[12] In our sample, less than one percent of moves were more than 300 CP worse than the best possible move in the respective position.

Our risk measure can be linked to the introductory example by Calford (2020) in Section 2.1: In both the game theoretic setting by Calford (2020) and in the above board position, an objectively best strategy exists: In the first case, the row player optimally randomizes between strategies A and B while in the second case, the player with the Black pieces would proceed optimally with the move 21... f5-f4. Yet, in both settings, risk attitudes can rationalize the choice of the dominated strategy.

Ultimately, our risk measure bears resemblance to the one proposed by Holdaway and Vul (2021). They start from a similar intuition of what constitutes the strategic risk of a chess move and assess how strongly the evaluation of a chess engine varies depending on the available responses by the opponent. They weight the responses by the empirical frequency of their occurrence and limit their analysis to responses with more than 100 occurrences. Our measure, while only reasonably applicable for very proficient chess players, does not require this empirical frequency (if available). In addition, we provide a different outcome, the difficulty of a chess position, instead of a pure engine evaluation.[13]

When applying the risk measure in Section 4 to analyze players' choices, one has to take into account the other options which were available to the player instead of the actual continuation in the game. In this regard, we assume that the five best moves in a given position constitute a relevant comparison benchmark for the move actually played in the game.[14] Therefore, as will be explained in Section 4, the regressions will include a covariate which captures the leave-out mean of risk, i.e. the average risk of the five best moves (without considering the risk of the played move if it is among the five best moves).[15]

3 Data: The FIDE World Cup

3.1 The FIDE World Cup

This section provides details on the chess tournament setting used in the empirical analysis as well as the terminology applied throughout this paper. We chose the FIDE World Cups for two reasons: First, a very high proficiency of the chess players is paramount for constructing the risk measure introduced in Section 2.3. The World Cup is part of the World Championship Cycle. The two World Cup finalists qualify for the Candidates Tournament whose winner, in turn, obtains the right to challenge the reigning World Champion. Each World Cup features 128 very skilled players whose selection is based on strict criteria. While most players qualify via the continental championships, additional participants are added based on their rating or

[13]The results for our main risk measure based on difficulty are presented in Section 5.1.1 while Section 5.1.2 considers an alternative risk measure based on the relative (dis-)advantage.

[14]Including the leave-out mean of risk as covariate assumes the existence of other moves the player could pick from. As will be explained in Section 3.2, we rule out situations with forced or obvious moves. In those positions, there is no real choice for the player and therefore, it would not be meaningful to analyze the choice of the player in those situations.

[15]In case of the played fifth best move 21... b7-b6 in the diagram of Figure 1, the leave-out mean of risk is computed over the alternatives among the five best moves, i.e. over the first, second, third and fourth best move.

due to titles such as the World Champion or the World Junior Champion (FIDE, 2021). The World Cup is a high-stake setting as the prize money increases by each round while the winner of the tournament receives 110,000 USD (FIDE, 2021).

Second, the specific format of the FIDE World Cup allows for exogenous variation in thinking time which is key for our identification strategy. We exploit that players compete in different time controls, a feature used to instrument the remaining thinking time in chess (see e.g. Howard (2024)). Further, in the FIDE World Cup, the selection into different time controls is reasonably exogenous to the choices of the chess players, a feature we discuss extensively below.

We consider the FIDE World Cups 2013 – 2023. They take place biannually so the sample consists of six World Cups (2013, 2015, 2017, 2019, 2021, 2023). Figure B2 visualizes the structure of one World Cup while Table B1 presents the corresponding *playing modes*. In each World Cup, there are 128 participants.[16] In the first *round*, those 128 participants play 64 *matches* where only the winner of each match proceeds to the second round as the World Cup is played in a knockout-system. This implies that in the seventh and final round, only one match remains which determines the winner of the respective World Cup. Each match comprises at least two *games*. The *playing mode* of those first two games is the *Classic* time control. This means that players are endowed with an initial time budget of 90 minutes for the game while they get an additional 30 seconds increment for each move played. Furthermore, they receive 30 minutes of additional time once they have reached move 40. A win in a game is rewarded with one point and a draw with half a point. If and only if the score is 1–1 after the two classic games, a tiebreak with a faster playing mode is needed. In this case, two *Rapid* games with 25 minutes and 10 seconds increment per move are played. If the match is still tied, two games with 10 minutes and 10 seconds (*Slow Blitz*) constitute the next part of the tiebreak. In case of a further tie, two games with 5 minutes and 3 seconds increments (*Fast Blitz*) are needed. If this ends in a tie as well, a final *Armageddon* game is played. In this game, the player with the White pieces receives 5 minutes while the player leading the Black pieces obtains only 4 minutes. White needs to win this game in order to claim the win in the match. In case of a draw, the player with the Black pieces wins the match and proceeds to the next round. Given those tiebreak rules and depending on the results of the games, a match can consist of 2 (if decided in the classic games), 4, 6, 8, or 9 games.[17] As detailed in Section 4, the tiebreak system forms the basis of the identification strategy. To identify the causal effect of remaining time on risk-taking, we exploit the exogenous variation in the playing modes and, hence, in initial thinking time. We discuss the identifying assumption and potential challenges to identification in Section 4.

[16] The World Cups 2021 and 2023 constitute an exception as additional participants took part so that also one additional round was played in the tournament.

[17] As the time mode is similar between *Armageddon* and *Fast Blitz* and as only very few *Armageddon* games were played in the World Cups, we do not distinguish both categories. Further, the one remaining match in the final round of each World Cup comprises 4 rather than 2 classic games. The tiebreak system, however, remains unchanged compared to matches in the other rounds.

3.2 Sample Definition and Descriptive Statistics

As stated, we use a sample of chess games from the six FIDE World Cups 2013, 2015, 2017, 2019, 2021 and 2023, downloaded from Chess24 (2023). We carry out an analysis on the level of half-moves. Our starting point is a dataset of 215,000 observations, covering the first 80 half-moves per game. We only consider these moves due to the additional time bonus players receive at half-move 80 in classic games. Further, we argue that risk considerations are particularly important for the opening and middlegame: Decisions during these phases of the game determine strongly which types of positions arise afterwards. Endgames, in contrast, often contain a rather technical component with less space for strategic considerations. Hence, when analyzing strategic risk, we are most interested in the strategic choices in the first 80 half-moves. We apply two restrictions to the data: First, we restrict the sample to players with at least 2400 Elo, reducing the sample size by approximately 5,000 observations. Second, we exclude both forced and "obvious" moves to only capture actual decisions by chess players rather than moves that do not require further reasoning. A move is *forced* if it is the only move in a given position that is compatible with the rules of chess. Obvious moves refer to situations where one player simply recaptures a piece of the opponent to avoid being material down. To define a move as *obvious*, we use the engine restricted to a low search depth. More specifically, we consider the three best moves and their engine evaluations at a search depth of five half-moves. If the difference in evaluation between the first and second best move exceeds 100 CP (i.e. signifying a serious blunder), we consider it as *obvious*. The intuition is that already a very low-ranked player will be able to find and implement this move, which is thus not necessarily a conscious decision but more a given in a specific position. This reduces the sample size by an additional 29,000 moves. Further, we drop 19,000 observations due to missing data. Hence, the final regression sample comprises approximately 162,000 half-move observations.

Table 2 provides descriptive statistics for the regression sample, stratified by the four different playing modes as described in Section 3.1. As pointed out above, the World Cup is characterized by high proficiency of the participants. This is confirmed by the move quality measures presented in Panel A: In 51 percent of positions of Classic games, players chose the first best move, in 23 percent of cases, they opted for the second best move and in 12 percent of positions, they played the third best move. Hence, deviations from the top suggested engine moves are rare due to the high proficiency prevalent in the World Cup. This reinforces the applicability of our measures as actual behavior and engine suggestions often coincide. Note that we present the move quality measures in a separate panel as they are properties of the played move and, hence, jointly determined with the risk of the move played. For this reason, they are neither included in the baseline specification nor in Panel C.

Panel B presents means and standard deviations of the dependent variables (cf. Section 2.3 for more detailed descriptions). They are expressed in logs to allow for interpretations of effects in percentage changes of the dependent variable. The baseline dependent variable, i.e. the logarithm of the variance of difficulty computed over five moves, has a mean of approximately 2.3 for all four playing modes. The more hypothetical moves are used for the computation of

Table 2: Descriptive statistics, stratified by playing mode

	(1) Classic		(2) Rapid		(3) Slow Blitz		(4) Fast Blitz	
Panel A: Quality of played moves								
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
First Best Move Played	0.51	0.50	0.49	0.50	0.47	0.50	0.45	0.50
Second Best Move Played	0.23	0.42	0.23	0.42	0.23	0.42	0.23	0.42
Third Best Move Played	0.12	0.32	0.12	0.32	0.13	0.33	0.13	0.33
Fourth Best Move Played	0.06	0.25	0.07	0.25	0.08	0.27	0.08	0.27
Fifth Best Move Played	0.04	0.19	0.04	0.20	0.04	0.20	0.05	0.22
Sixth Best Move Played	0.03	0.16	0.03	0.17	0.03	0.17	0.04	0.19
Seventh Best Move Played	0.02	0.14	0.02	0.14	0.02	0.15	0.03	0.16
Move Quality	14.50	92.32	18.62	108.38	22.19	142.33	24.30	137.52
Panel B: Dependent Variables								
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Log. Variance Difficulty (5 boards), *baseline*	2.23	1.69	2.33	1.69	2.34	1.74	2.25	1.70
Log. Variance Difficulty (7 boards)	2.32	1.58	2.41	1.59	2.42	1.63	2.34	1.59
Log. Variance Difficulty (6 boards)	2.28	1.63	2.38	1.63	2.39	1.68	2.30	1.63
Log. Variance Difficulty (4 boards)	2.14	1.80	2.23	1.81	2.25	1.84	2.16	1.81
Log. Variance Difficulty (3 boards)	1.93	2.02	2.03	2.03	2.05	2.06	1.95	2.05
Log. Maximum Difficulty (5 boards)	3.01	0.55	3.06	0.56	3.08	0.56	3.06	0.54
Log. Minimum Difficulty (5 boards)	2.25	0.94	2.31	0.93	2.33	0.94	2.33	0.91
Log. Variance Relative Advantage (5 boards)	6.43	2.38	6.52	2.36	6.54	2.34	6.43	2.40
Panel C: Independent Variables								
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Remaining Time (in minutes)	57.24	33.06	16.79	8.85	7.31	3.62	3.33	1.79
Relative Advantage	14.61	196.54	13.14	241.37	15.99	246.36	17.74	286.60
Position Difficulty	17.79	12.48	18.97	13.48	19.21	13.68	18.46	11.80
Log. Average Variance Difficulty of Alternatives	2.72	1.29	2.82	1.30	2.86	1.30	2.80	1.27
Number of half-moves (unit of aggregation):	96,508		44,529		14,906		6,081	
Number of games:	1,820		780		268		109	
Number of matches which reached the mode:	908		390		135		52	
Number of players who reached the mode:	419		289		167		78	

Note: The sample consists of World Cup games from 2013 – 2023. As explained in Section 2.3, the computation of all dependent variables is based on the hypothetical boards that can arise after the played move. The variables *Relative Advantage*, *Position Difficulty* and *Move Quality* are measured in centipawns (cf. Sections 2.2 and A.1 for further explanations). The variables *First Best Move Played* until *Seventh Best Move Played* are dummy variables.

the variance (i.e. the higher the number of anticipated responses of the opponent), the higher the mean of the respective dependent variable. Additionally, the averages of the variance of difficulty increases from the *Classic* games to the *Slow Blitz* games. The maximum of difficulty amounts on average to more than three log points in each playing mode. Further, as outlined in Section 2.3, we use the variance in engine evaluations – to capture the relative (dis-)advantage of a position – as an alternative risk measure. On average, it is considerably larger than any variance over the difficulty measure. While difficulties are always positive, the underlying engine evaluations consist of both positive and negative numbers and take on higher values in absolute terms than the difficulties in the sample, leading to a higher variance.

Further, Panel C provides descriptive statistics of the explanatory variables. The key variable of interest *Remaining Time* is measured in minutes. In line with the shorter time controls applied in tiebreaks, the averages in this variable decrease from playing mode *Classic* to the mode *Fast Blitz*. The relative advantage is measured in centipawns, as explained in Section 2.2. It is evaluated prior to the move and expected to have a mean close to zero in the sample as an advantage for the White player coincides with a disadvantage for the Black player and vice versa (Künn et al., 2022). The averages per playing mode are all below 20 centipawns which is close to an equal position, thus in line with the expectation. The difficulty of the position is also evaluated prior to the move. Positions in the sample have an average difficulty between 18.37 and 20.00 for each of the four playing modes. This implies that, based on data from previous chess games, the positions are, on average, associated with an expected decline in position evaluation between approximately 18 and 20 centipawns if a human player plays these positions.

3.3 Relating our Risk Measure to Risk Classifications of Openings

As stated, one key contribution of our study is a novel measure of strategic risk of single chess moves. In this section, we show that our main measure of strategic risk, i.e. the variance of the difficulty of the positions subsequent to the five best responses, is consistent with prior research.

An important benchmark for our measure is the most established approach to measure risk-taking in chess in the economic literature by Dreber et al. (2013). They consider chess openings which can be classified along 500 different *ECO codes* (Encyclopaedia of Chess Openings). They let chess experts evaluate each of the 500 openings as either solid, neutral or aggressive from the perspective of both Black and White. In total, there are six combinations of these three opening types (solid vs. solid, solid vs. neutral,...), i.e. six classes of openings depending on the risk that chess experts assign to those openings. We conjecture that risky openings also lead to relatively more risk-seeking moves afterwards. Hence, we compare our main measure of strategic risk in the half-moves 1 – 80 across the six classes of openings. In Table 3, we regress our risk measure on the dummies for the opening categories and employ the least risky opening class, i.e. *Opening Solid For Both Players*, as reference.[18] From top to bottom, the dummy variables represent increasingly risky openings. We find that the regression coefficients rise accordingly.

[18] The explanatory variable *Log. Average Variance Difficulty of Alternatives* is explained in Section 4.

Table 3: OLS estimates comparing risk in different classes of openings. The dependent variable is the logarithm of the variance of difficulty.

	(1)	(2)	(3)
Opening Solid For Both Players		*reference*	
Opening Solid For One Player, Neutral For The Other Player	0.0137	0.0156	0.0143
	(0.0091)	(0.0096)	(0.0095)
Opening Neutral For Both Players	0.0972***	0.0992***	0.0986***
	(0.0115)	(0.0122)	(0.0122)
Opening Solid For One Player, Aggressive For The Other Player	0.1468***	0.1365***	0.1349***
	(0.0157)	(0.0166)	(0.0165)
Opening Aggressive For One Player, Neutral For The Other Player	0.1578***	0.1554***	0.1527***
	(0.0131)	(0.0139)	(0.0138)
Opening Aggressive For Both Players	0.2529***	0.2550***	0.2507***
	(0.0224)	(0.0237)	(0.0234)
Log. Average Variance Difficulty of Alternatives	0.6762***	0.6702***	0.6801***
	(0.0030)	(0.0030)	(0.0030)
Constant	0.3524***	0.3691***	-0.0928***
	(0.0092)	(0.0094)	(0.0220)
Player FE	No	Yes	Yes
Half-Move Dummies	No	No	Yes
Observations	162024	162024	162024
R^2	0.270	0.273	0.281

Note: Robust standard errors in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. The explanatory variables *Opening Solid For Both Players* (reference category) until *Opening Aggressive For Both Players* are six dummies indicating that the opening belongs to the respective class of openings (as defined by Dreber et al. (2013)). * p < 0.10, ** p < 0.05, *** p < 0.01

Openings that are neutral for both players coincide with risk that is ten percent higher compared to openings which are solid for both players. Additionally, in games with aggressive openings, players opt for moves associated with an increase in risk of 25 percent relative to the reference category. This applies to the OLS estimation in column (1), the player fixed-effects estimation in column (2) and when additionally controlling for half-move dummies in column (3).[19] Thus, our novel risk measure is in line with the predictions of chess experts regarding the riskiness of openings.

[19] In untabulated estimation results, we show that the ranking of the six opening in terms of risk is also maintained when excluding the opening moves by restricting to the half-moves 21-80.

4 Econometric Specification

We aim to analyze the effect of remaining thinking time on revealed risk-taking of players. As discussed in Section 2.3, our main risk measure is built on the difficulty of a board as the key outcome, while we also repeat the analysis with the relative (dis-)advantage data. We estimate the following model:

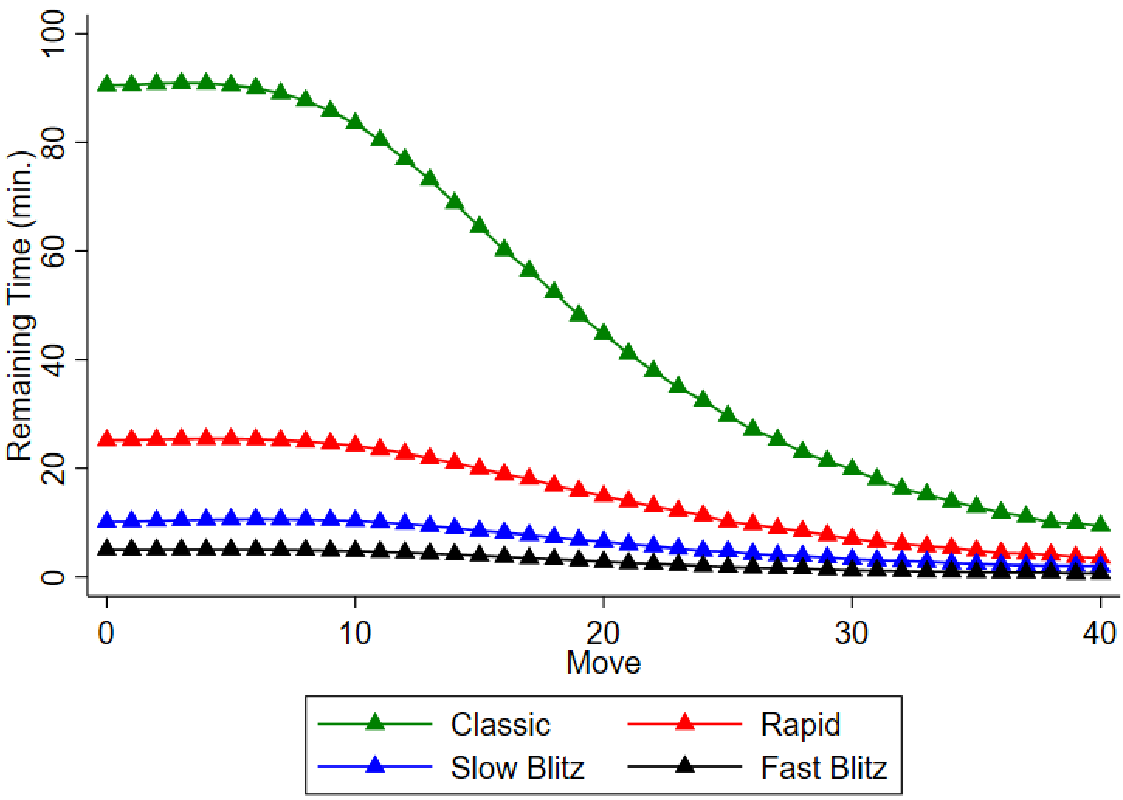
$$
\begin{aligned}
log(Risk_{pmgh}) = \alpha \cdot RemainingTime_{pmgh} + \sum_{k=1}^{80} \phi_k \cdot DHalf-Move_{kh} \\
+ X_{pmgh}\gamma + \mu_{pmg} + u_{pmgh} \qquad (1) \\
p = 1, \ldots, 354\, players \quad and \qquad m = 1, \ldots, 704\, matches, \\
g = 1, \ldots, 9\, games \qquad and \qquad h = 1, \ldots, 80\, half-moves
\end{aligned}
$$

where the indices are as follows: p indexes the player. This identifier is person-specific and does not vary across different World Cups. The subscript m refers to the identifier of the match. Due to the knockout-format of the World Cup, the number of matches reduces by the factor 0.5 in each round. The index g refers to the game number within a match. The number of games of each match depends on whether or not (multiple) tiebreaks are played. A match consists of not more than 9 games. The subscript h represents the number of the current half-move within a game. Thus, $h = 1$ refers to the first move played by the player leading the White pieces, $h = 2$ to the first move by Black, $h = 3$ to the second move by White and so on. We restrict the data to the first 80 half-moves per game due to the additional time bonus players receive at half-move 80 in classic games. We control for the current half-move to capture time trends in risk over the course of the game. Further, the vector X_{pmgh} includes control variables that vary on the move-level. We control for both the relative advantage and the difficulty of the current position on the board, both measured prior to half-move h. These covariates foster the comparability of different board positions. Additionally, we include the logarithm of the average variance of difficulty of alternative moves as control variable. This allows to proxy the set of other options which would have been available to the player instead of the move played in the game. To construct this measure, we consider the set of five best moves in a given position but exclude the played move if it is among the five best moves. Next, we compute the average risk over this set of moves. We apply the logarithm to this leave-out mean of risk and control for the resulting variable in our model. Moreover, confounding factors on the more aggregate game-level such as rating difference between the players or the choice of the opening are absorbed by the interacted player-match-game fixed-effects μ_{pmg}.

However, equation (1) suffers from an endogeneity concern: The time remaining on the clock to make a decision is driven by past time consumption in the game. Thus, when estimated via OLS, the effect of remaining time on risk might be confounded by reversed causality: Players might spend more time when making risk-seeking decisions, creating a negative link between $RemainingTime_{pmgh}$ and $Risk_{pmgh}$ for all half-moves $a < h$ of player p within game g of match m. At the same time, $Risk_{pmgh}$ might be positively serially correlated within games. This would, in turn, create a reversed channel between the contemporaneous values of $RemainingTime_{pmgh}$ and $Risk_{pmgh}$, leading to endogeneity.

Therefore, we aim to identify the effect of remaining time via an instrumental variable approach which employs identifying variation from different playing modes. Those stem from matches that reached the tiebreak, which in turn causes 2 additional games in a shorter time control. The playing-mode-specific paths of time consumption represent the identifying variation and are visualized in Figure 2. We observe large persistence in remaining thinking time within the four playing modes but significant differences in time paths across playing modes. Hence, this graph documents the relevance of the instrument.

Figure 2: Evolution of remaining time within games



Note: This graph shows the paths of average remaining time within games of the four playing modes *Classic*, *Rapid*, *Slow Blitz* and *Fast Blitz*. The sample consists of World Cup games from 2013 – 2023.

In our identification strategy, we thus instrument the remaining time with initial time, interacted with half-move dummies. The first stage is given by:

$$RemainingTime_{pmgh} = Initial\_Time_g \cdot (\sum_{k=1}^{80} \psi_k \cdot DHalf - Move_{kh})$$
$$+ \sum_{k=1}^{80} \eta_k \cdot DHalf - Move_{kh} + X_{pmgh}\kappa + \lambda_{pmg} + e_{pmgh} \tag{2}$$
$$p = 1, \ldots, 354 \, players \quad and \qquad m = 1, \ldots, 704 \, matches,$$
$$g = 1, \ldots, 9 \, games \qquad and \qquad h = 1, \ldots, 80 \, half - moves$$
$$(Classic, \; Rapid, \; Slow \; Blitz, \; Fast \; Blitz) \in Initial\_Time_g$$

The exogeneity requirement reads as follows:

$$\mathbb{E}(Initial\_Time_g * u_{pmgh}|X_{pmgh}, \mu_{pmg}, DHalf - Move_h) = 0$$

Thus, the empirical approach assumes that the initial thinking time per game g is exogenous with respect to risk measured at half-move h when controlling for interacted player-match-game fixed-effects. The exclusion restriction states that the playing mode only influences risk indirectly via the variable $RemainingTime_{pmgh}$ but not directly. Equivalently, there shall be no time-variant effect of initial time on risk-taking behavior.

To elaborate on the relationship between playing modes and risk, it should be noted that the risk measure does not use information on the playing mode. Further, Figure B3 plots the evolution of average risk over the half-moves for the four different playing modes. Overall, none of the four risk trends considerably exceeds or falls below the trends of the other three playing modes. Moreover, the notion that initial time affects risk only through remaining time is consistent with players being focused on their games and only perceiving the current time on the clock rather than reflecting on their initial time budget.

It is worth noting that the initial time of each playing mode is exogenously given by the tournament regulation. Therefore, players who reach the tiebreak play the tiebreak game with the prescribed time mode. Thus, the playing mode in tiebreaks is unaffected by time consumption or other measures from the previous game. Hence, the present setting ensures that the identifying variation is not conflated by any sort of pre-trend in risk.

Our identification strategy resembles, yet improves upon the existing approach by Howard (2024) regarding selection concerns. His study uses data from an online chess platform where players can freely choose a playing mode. He instruments the time a player spends on a move with the playing mode. Using a sample of the World Cup as in the present study yields the advantage that participation in tiebreaks is conditional on the outcome of the match and, thus, also depends on the choices of the *opponent*. Other playing modes occur only in the case of tiebreaks and are beyond the direct choice of a player. Hence, using data from a knockout-tournament remedies selection concerns into a certain playing mode.

The inclusion of the interacted player-match-game fixed-effects μ_{pmg} as used by Zegners et al. (2020) in a similar setting is crucial as it limits the set of potential confounders to variables that vary within games and players. Thus, time-invariant traits of players such as higher preference for faster time controls are no threat to identification. In a similar manner, the rating difference of two players might be a determinant for a match reaching the tiebreak. However, as the rating of both players is constant within a game, it is absorbed by the fixed-effects as well. Similarly, players might adapt their choice of strategic risk to the reputation or the strength of the opponent. Still, this would not undermine identification of the effect of remaining time as long as these characteristics of the opponent are invariant per game.

Further, the effect of interest is identified both by matches with and without tiebreaks. While the two classic games are played in each match, the emergence of tiebreak games depends on the results of the classic games. The instrumental variable estimation, however, is based on exogeneity of the initial time. It assumes that the selection of matches into tiebreaks is not driven by unobservables in the risk-equation. Players might consider to self-select into tiebreaks by choosing quiet openings that are likely to result in draws and, consequently, in tiebreaks. The chosen opening, however, is invariant per game, as each game is attributable to one specific opening. Thus, the concern of endogenous sorting is partially remedied by the fixed effects as well.

We further discuss the question of selection into tiebreaks based on Table 4. It analyzes determinants of the probability that a match is not decided in the *Classic* games but proceeds at least to the *Rapid* tiebreak. We run this analysis for the set of all 908 matches played in the six World Cups from 2013 to 2023. In all four specifications, we include the average risk in classic games (i.e. the match-level average of the dependent variable in equation (1)) as the explanatory variable of interest. The further independent variables represent the match-level averages of X_{pmgh} from equation (1). This includes the average risk of alternatives to proxy the set of available moves as well as the average position difficulty. Specification (3) additionally features the inclusion of average move quality in the classic games as explanatory variable. It captures the emergence of blunders in the classic games which might coincide with more decided games and less tiebreaks. Morever, we add the absolute rating difference in the match as a determinant in specification (4). This variable does not vary within games and is thus captured by the fixed-effects μ_{pmg} from equation (1). The present estimation, however, features a cross-section of match-level observations where no fixed-effects are included, allowing to include this covariate. Among all specifications, the coefficient of *Average Risk in Classic Games* remains insignificant. Overall, we do not find an effect of average risk-taking in the classic games on the probability that the match reaches the tiebreak. Therefore, we infer that also the unobservable component of risk-taking in the classic games is unrelated to the existence of tiebreaks. This would imply that the error term in equation (1) does not depend on the actual mode or on initial time, respectively. Hence, this finding supports the exclusion restriction. In Table B3, we repeat this analysis for reaching the Slow Blitz tiebreaks or the Fast Blitz tiebreaks. Results remain qualitatively unchanged. Further, in untabulated results we reproduce Tables 4 and B3 for the subsample of half-moves 21-80 and obtain similar results.

Table 4: Placebo Test: Linear probability model on the match-level. The dependent variable is a dummy indicating whether the match reaches the Rapid tiebreaks.

	(1)	(2)	(3)	(4)
Average Risk in Classic Games	0.0010	0.0167	0.0408	0.0921
	(0.1007)	(0.0956)	(0.0953)	(0.0929)
Average Risk of Alternatives in Classic Games	-0.2425**	0.1071	0.1023	-0.0002
	(0.1097)	(0.1176)	(0.1182)	(0.1176)
Average Position Difficulty in Classic Games		-0.0458***	-0.0418***	-0.0361***
		(0.0069)	(0.0068)	(0.0067)
Average Move Quality in Classic Games			-0.0039**	-0.0035**
			(0.0016)	(0.0016)
Absolute Rating Difference in Match				-0.0007***
				(0.0002)
Constant	1.0881***	0.9130***	0.8567***	1.0006***
	(0.1199)	(0.1209)	(0.1217)	(0.1259)
Observations	908	908	908	908
R^2	0.040	0.086	0.095	0.133

Note: Robust standard errors in parentheses. The sample consists of World Cup games from 2013 – 2023. The covariate *Average Risk of Alternatives in Classical Games* represents the match-level average of the variable *Log. Average Variance Difficulty of Alternatives* as included in the other specifications.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Endogeneity might arise only from variables that are correlated with initial time and constitute determinants of risk while they are not absorbed by the fixed effects. Thus, only residuals from the second stage equation, i.e. patterns in risk within the game that are unexplained by the general time trends in risk and by the covariates can lead to endogeneity. This implies that fatigue effects across games, i.e. players feeling exhausted after having played multiple games and thus opting for riskier moves, are accounted for by the fixed effects. Similarly, exhaustion within games is controlled for by the half-move dummies as long as there is no playing mode specific development of exhaustion over half-moves. Thus, our instrumental variable approach assumes only that there is no pattern of fatigue which leads to playing mode-specific paths of risk.

Further, it is noteworthy that the different playing modes do not differ in their importance in the match. Conditional on achieving playing mode *Rapid*, the emerging two games are as important as the initial two *classic* games, the same holds for the faster time controls. Each time control is potentially the last time control played in the match in case no further tiebreak is needed. Therefore, it seems unlikely that players adjust their risk-taking behavior when switching between two playing modes ceteris paribus.

5 Results

5.1 The Effect of Time Pressure on Strategic Risk

This section analyzes the causal effect of time pressure on strategic risk. Subsection 5.1.1 focuses on the variance of difficulty as our risk measure of main interest (see Section 2.3 for more details). In Subsection 5.1.2, we use the variance of relative (dis-)advantage as an alternative measure.

5.1.1 Main Measure: Variance of Difficulty

Table 5 shows the baseline estimates for OLS, fixed effects and instrumental variable regressions.[20] The OLS estimates in specification (1) imply that a decrease in remaining time by one minute increases the variance of difficulty by 0.36 percent. The significant negative coefficient persists when including interacted player-match-game fixed effects μ_{pmg} in specification (2). In the third specification, we additionally control for the relative advantage and the position difficulty as well as the logarithm of the average variance of difficulty of alternative moves, i.e. the leave-out mean of risk. Despite controlling for these covariates and fixed effects, estimates in columns (1) to (3) might be subject to endogeneity of remaining time as explained in Section 4. Therefore, specification (4) takes the endogeneity into account and runs an instrumental variable approach. As described above, it instruments remaining time by initial time, interacted with half-move dummies. Thus, the identifying variation stems from the four different playing modes in the World Cup.

We find that a reduction in the time budget by one minute reduces the variance of difficulty by 0.12 percent. In the classic games with a time budget of 90 minutes for 40 moves, the average change in remaining time per move amounts to 2.25 minutes. Multiplying this value by the coefficient of 0.0012 yields an average change in the dependent variable of 0.0027 log points. This represents a share of about 0.2 percent of the standard deviation of the dependent variable (1.69 for classic games). When comparing the IV specification to the first three columns, it turns out that the coefficients in the estimations without IV are downward-biased. This is related to the direct link of past time consumption on current remaining time: If players consume more time to play moves that entail more risk until half-move h, they dispose of a smaller time budget in h. If risk is additionally serially correlated within games, a downward bias occurs.

To demonstrate the robustness of the positive effect of remaining time on risk, Table 6 considers different subsamples. In specification (1), we restrict to observations with a remaining time budget of at least 30 seconds. This rules out that the previously observed effect is only due to situations in severe time pressure. We show that the effect of remaining time on risk persists, indicating that it also applies to scenarios where players can apply conscious decision-making.[21] Additionally, column (2) considers the half-moves 21-80 rather than the half-moves 1-80. This restriction targets the behavior of professional chess players to follow established opening systems. Thus, the opening moves – similar to obvious and forced moves – do not always entail practical decision-making of players. We show that the sample of observations after the opening provides a similar estimate for the effect of remaining time. The three remaining columns contain specifications for a subset of games: Specification (3) features observations of players who

[20] Table B2 presents the corresponding first stage estimates of the baseline. The coefficients of the interaction terms 'Initial Time × Half-Move h' are mostly negative. Hence, larger values of initial time coincide with smaller (i.e. 'more negative') changes in remaining time between half-moves 1 and h. This is in line with larger time consumption in slower time controls between half-moves 1 and h. Additionally, coefficients of 'Initial Time × Half-Move h' become smaller once h increases. This reflects the overall decline in remaining time during a game. Table B2 also contains the corresponding F-statistic which is larger than 500.

[21] In untabulated tests, we document that the effect remains stable when restricting to observations with at least one minute or two minutes of remaining time.

Table 5: The effect of remaining time on risk-taking, based on equation (1), for multiple estimation techniques. The dependent variable is the logarithm of the variance of difficulty.

	(1) OLS	(2) interacted player-match-game FE	(3) interacted player-match-game FE	(4) IV with interacted player-match-game FE
Remaining Time (in minutes)	-0.0036***	-0.0046***	-0.0008**	0.0012***
	(0.0003)	(0.0004)	(0.0003)	(0.0004)
Relative Advantage			-0.0000	-0.0000
			(0.0000)	(0.0000)
Position Difficulty			0.0275***	0.0276***
			(0.0006)	(0.0006)
Log. Average Variance Difficulty of Alternatives			0.5231***	0.5238***
			(0.0042)	(0.0042)
Interacted player-match-game FE	No	Yes	Yes	Yes
Half-Move Dummies	No	No	Yes	Yes
Observations	162024	162024	162024	162024
R^2	0.005	0.144	0.338	0.230

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. The corresponding first stage estimation including an F-statistic for the relevance of the instrumental variables is presented in Table B2.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

have to win the current game in order to stay in the tournament. This is the case whenever a player lost the first game in a given playing mode. The coefficient is significant and more than twice as large as the coefficient from the baseline estimation. This suggests that the effect of remaining time on risk-taking behavior is particularly pronounced for observations which are associated with tournament pressure. Specifications (4) and (5) consider drawn and decided games, respectively. They imply that the overall effect is driven by lost and won rather than by drawn games.

Additional robustness checks refer to the definition of the dependent variable. Our risk measure, i.e. the logarithm of variance of difficulties, requires an assumption about how the difficulties of the five hypothetical boards are weighted. The example of the normal-form game in Section 2.1 demonstrates the importance of beliefs on how likely the respective strategies by the opponent are. Our baseline measure is computed as the logarithm of an unweighted variance. Thus, it assumes equal beliefs about the opponent's possible responses. In Table 7, we relax this restriction and employ two different weighting schemes. Both specifications conjecture that the high proficiency of players in the sample induces them to attach a higher weight to objectively stronger moves. In Specification (1), we modify the measure presented in Section 2.1 by weighting the difficulties of the five hypothetical boards with their respective engine evaluations.[22]

[22]More precisely, we run the following procedure:
1. We compute the difference in relative advantage after the opponent's first best (e.g. advantage of 20 CP) and after the opponent's sixth best move (e.g. advantage of 220 CP): 220-20 = 200
2. We calculate for each of the five responses their difference in relative advantage to the first best response. (e.g. for an advantage of 40 CP after the second best move): 40 - 20 = 20
3. We determine which component of the difference in 1. is due to the difference computed in 2.: 200 - 20 = 180

Table 6: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check with different subsamples. The dependent variable is the logarithm of the variance of difficulty.

	(1) at least 30s remaining time	(2) Half-Moves 21-80	(3) only must win games	(4) drawn games	(5) lost and won games
Remaining Time (in minutes)	0.0011***	0.0017***	0.0030***	0.0003	0.0015***
	(0.0004)	(0.0006)	(0.0010)	(0.0006)	(0.0005)
Relative Advantage	-0.0000	-0.0000	-0.0001	-0.0001	-0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0000)
Position Difficulty	0.0278***	0.0265***	0.0266***	0.0317***	0.0244***
	(0.0006)	(0.0007)	(0.0013)	(0.0009)	(0.0007)
Log. Average Variance Difficulty of Alternatives	0.5219***	0.5402***	0.5266***	0.5135***	0.5292***
	(0.0043)	(0.0051)	(0.0118)	(0.0061)	(0.0060)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	160181	108520	19491	80925	81099
R^2	0.228	0.237	0.251	0.221	0.241

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * p $<$ 0.10, ** p $<$ 0.05, *** p $<$ 0.01

Similarly, Specification (2) takes into account that better moves might receive a higher weight. In this case, we weight positions by a geometric series with the factor 0.8.[23] For both weighting procedures, we replicate the effect found in the baseline.

We include additional robustness checks in the Appendix: Table B4 lifts the assumption that five hypothetical boards should be considered in the computation of the variance of difficulty. We show that the results persist when switching to 7, 6, 4 or 3 hypothetical boards. Two robustness checks concern the inclusion of covariates: First, Table B5 includes risk of past half-moves of player p from the respective game. This fosters exogeneity of the instrument by ensuring that past risk creates no link between initial time and current risk. We incrementally add lagged risk variables and show that the coefficient of *Remaining Time* is robust. Second, Table B6 varies the functional form of covariates. It emphasizes that results remain robust, irrespective of whether the explanatory variables are used in logarithmic form or not.[24] Additionally, Table B6 includes the move quality of the move played (specifications (3) and (4)) as well as the mean of difficulty from the hypothetical boards (specifications (5) and (6)) as additional control variables. Their inclusion does not change the effect of remaining time on risk-taking.

4. We divide each of those five differences computed in 3. by the difference between the first best and sixth best move: 180/200 = 0.90
5. We normalize those five shares so that they add up to 1.
6. We compute the variance of the five difficulties, each weighted with its respective share in 5.

[23] This factor represents the median of the geometric mean of growth rates in the engine evaluation from the first best move to the fifth best move.

[24] We do not compute the logarithm of the variable *Relative Advantage* as there are approximately 50 percent of observations with a negative value in this variable.

Table 7: IV estimates for the effect of remaining time on risk-taking, based on equation (1). The dependent variable is the logarithm of the variance of difficulty, based on different weighting schemes.

	(1) Weighting based on engine evaluations	(2) Weighting based on a geometric series with factor 0.8
Remaining Time (in minutes)	0.0021***	0.0013***
	(0.0007)	(0.0004)
Relative Advantage	0.0000	-0.0000
	(0.0000)	(0.0000)
Position Difficulty	0.0395***	0.0280***
	(0.0013)	(0.0006)
Log. Average Variance Difficulty of Alternatives	0.4252***	
	(0.0172)	
Log. Average Variance Difficulty of Alternatives		0.5072***
		(0.0043)
Interacted player-match-game FE	Yes	Yes
Half-Move Dummies	Yes	Yes
Observations	162024	162024
R^2	0.115	0.215

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. When computing the variance of difficulty, specification (1) weights each of the five hypothetical positions by its distance in engine evaluation to the first best response. In specification (2), we use a geometric row to assign weights to the difficulty predictions after the five best responses. The factor of 0.8 was calculated as follows: We compute the percentage change from the engine evaluations of the best move to the engine evaluation for the fifth best move. Next, we calculate the geometric mean of this percentage change over four growth rates. We take the median of this geometric mean of 0.8 as weighting factor. However, results remain robust when weighting either by the first or third quartile. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Further robustness checks regarding the sample composition are presented in Table B7. In contrast to the baseline, it also includes players with less than 2400 Elo and situations where there is either a clearly best ('obvious') or even a forced move. It is apparent that neither sample restriction drives the baseline result. Moreover, Table B8 applies different more restrictive Elo limits to the regression sample. We show that results hold when focusing on moves by players with a rating of at least 2500, 2600, 2700 or 2750 Elo.

Additionally, Table B9 provides four specifications where each estimation sample excludes one of the four playing modes. When dropping one of the modes (1) *Fast Blitz*, (2) *Slow Blitz* or (3) *Rapid*, the results remain unchanged. When excluding *Classic* games in specification (4), the estimated coefficient of remaining time is more than twice as large as the coefficient in the baseline but less precisely estimated. This might be due to the large share of *Classic* games in the sample (61 percent), which account for a considerable share of the identifying variation. Additionally, excluding *Classic* games mechanically compresses the distribution of $RemainingTime_{pmgh}$. A lower variance of the explanatory variable leads to lower estimation precision, thus increasing the standard errors.

Moreover, we link strategic risk to the outcomes of games in Table B10 in Section A.3.5. We show that risk is positively correlated with an indicator for winning a game, although the causal direction of this relationship is less clear.

5.1.2 Variance of Relative Advantage

An additional robustness check is presented in Table 8. Instead of the variance of difficulty, it uses the logarithm of the variance of relative (dis-)advantage of the five hypothetical boards.[25] Thus, it does not contain information based on human performance as does the baseline outcome. Still, this is a relevant comparison benchmark as the evaluation of the position is a frequently used indicator in the analysis of chess games (e.g. Künn et al. (2022)) and similar risk measures have been proposed (Holdaway and Vul, 2021). Similar to the analysis in Section 5.1.1, we control for the leave-out mean of risk of other moves that would have been possible (*Log. Average Variance Relative Advantage of Alternatives*). We document positive and significant effects of remaining time: Depending on the number of hypothetical boards under consideration (specifications (1)-(5)), a decrease in remaining time by one minute reduces risk by an amount between 0.23 and 0.29 percent. Hence, Table 8 confirms the positive effect of remaining time on risk from Table 5.

Table 8: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check using an alternative risk measure. The dependent variable is the logarithm of the variance of relative advantage.

	(1) 7 boards	(2) 6 boards	(3) 5 boards	(4) 4 boards	(5) 3 boards
Remaining Time (in minutes)	0.0023***	0.0024***	0.0024***	0.0026***	0.0029***
	(0.0007)	(0.0008)	(0.0008)	(0.0008)	(0.0008)
Relative Advantage	0.0005***	0.0005***	0.0005***	0.0005***	0.0004***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0328***	0.0337***	0.0349***	0.0355***	0.0359***
	(0.0008)	(0.0008)	(0.0009)	(0.0009)	(0.0009)
Log. Average Variance Relative Advantage of Alternatives	0.3001***				
	(0.0039)				
Log. Average Variance Relative Advantage of Alternatives		0.2967***			
		(0.0039)			
Log. Average Variance Relative Advantage of Alternatives			0.2921***		
			(0.0039)		
Log. Average Variance Relative Advantage of Alternatives				0.2922***	
				(0.0038)	
Log. Average Variance Relative Advantage of Alternatives					0.2774***
					(0.0037)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	162024	162024	162024	162024	162024
R^2	0.165	0.160	0.153	0.148	0.132

Note: Standard errors clustered on the game-level in parentheses. Sample: World Cup Games from 2013 – 2023. The dependent variable varies on the half-move-level. It is computed as the logarithm of the variance of relative advantage of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[25] We compute it as *log(1+var(relative advantage))* to avoid missing values in case of a variance of zero.

5.2 Strategic Loss Aversion

We can also provide correlational evidence for factors other than remaining time that are determinants for risk-seeking or risk-averse behavior. More specifically, we investigate whether chess players exhibit *loss aversion*, a key prediction of prospect theory. Kahneman and Tversky (1979) model how people aggregate gains and losses relative to some given reference point. Their central claim is that this aggregation is characterized by risk-aversion in the gain domain and risk-seeking in the loss domain. Loss aversion implies that losses are weighted stronger than gains of the same magnitude. We analyze two different definitions of being in the loss domain and show that the chess players in our sample act as-if they are loss averse.

First, we consider blunders in the last move as a proxy of being in the loss domain. Players might act as if they defined the status quo before their own blunder as the reference point. In terms of prospect theory, playing a blunder might shift the player to the loss domain which would coincide with risk-seeking choices. In Panel A of Table 9, we consider a definition of a blunder as 50 CP as well as similar definitions (100 centipawns, continuous measure in centipawns).[26] For all three blunder definitions, we show a positive correlation between a blunder in the previous move of player p and higher risk in the current move of player p. This implies that players are more prone to take risks in the subsequent move when their last move coincided with a loss in the relative advantage. This matches previous findings of Holdaway and Vul (2021), who find that players choose riskier strategies (see the end of Section 2.3 for a discussion of their measure) after a blunder.

Second, we focus on the relative advantage. Thus, we follow Zegners et al. (2020) and consider an equal position as the reference point.[27] Therefore, in Panel B of Table 9, we consider the relation between the relative advantage prior to the move and our risk measure. A positive (negative) relative advantage implies that the player who plays the next move has an advantageous (disadvantageous) position. We run fixed effects regressions and show that there is a negative correlation between the current evaluation and risk-taking (although partially imprecisely estimated), further supporting the notion of loss aversion. To rule out the influence of outliers in relative advantage, we consider subsamples of positions within a range of 500, 400, 300 or 200 centipawns from a fully equal position. The robust negative correlation implies that players choose a lower risk when having a better position, which we interpret as strategic loss aversion.

[26] In ten percent of observations in our sample, players choose a move that worsens the position evaluation by at least 50 CP.

[27] In the present case, a blunder in the last move leads to deviations from the reference point. Analogies can be drawn from Table 6: As pointed out, its third specification considers the subsample of must win games, i.e. situations where a player lost the previous game in the match and is thus forced to win in the next game in order to remain in the tournament. In case of must win games, deviations from the reference point were generated prior to the game. For the subsample of must win games in Table 6, we find an effect of remaining time on risk that is stronger than in the overall sample. Thus, the position of being in the loss domain leads to a higher sensitivity of risk to remaining time. Hence, we might interpret these findings as a first hint at strategic loss aversion.

Table 9: Fixed effects estimates for the effect of making a blunder in the previous move (Panel A) and for the effect of relative advantage (Panel B). The dependent variable is the logarithm of the variance of difficulty.

Panel A: effect of a blunder in the previous move			
	(1)	(2)	(3)
Blunder in Previous Move (Threshold 50 CP)	0.0955***		
	(0.0152)		
Blunder in Previous Move (Threshold 100 CP)		0.1011***	
		(0.0246)	
Move Quality of Previous Move			0.0002***
			(0.0001)
Log. Average Variance Difficulty of Alternatives	0.5280***	0.5285***	0.5289***
	(0.0043)	(0.0043)	(0.0043)
Position Difficulty	0.0268***	0.0269***	0.0270***
	(0.0006)	(0.0006)	(0.0006)
Remaining Time (in minutes)	-0.0007**	-0.0008**	-0.0008**
	(0.0003)	(0.0003)	(0.0003)
Interacted player-match-game FE	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes
Observations	156260	156260	156260
R^2	0.347	0.347	0.347

Panel B: effect of relative advantage					
	(1) All	(2) at most 500 CP advantage by either side	(3) at most 400 CP advantage by either side	(4) at most 300 CP advantage by either side	(5) at most 200 CP advantage by either side
Relative Advantage	-0.0000	-0.0001***	-0.0002***	-0.0001	-0.0001
	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Log. Average Variance Difficulty of Alternatives	0.5231***	0.5191***	0.5171***	0.5139***	0.5094***
	(0.0043)	(0.0044)	(0.0044)	(0.0045)	(0.0046)
Position Difficulty	0.0275***	0.0285***	0.0290***	0.0294***	0.0296***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0007)
Remaining Time (in minutes)	-0.0008**	-0.0009***	-0.0010***	-0.0011***	-0.0011***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	162024	157427	155016	151370	145409
R^2	0.338	0.329	0.326	0.322	0.315

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. In Panel A, we define a blunder as a move with at least 50 CP loss (specification (1)), at least 100 CP loss (specification (2)) and as a continuous variable (specification (3)). In Panel B, we consider the whole sample (specification (1)) as well as subsets of positions with a relative advantage of at most 500, 400, 300 or 200 CP by either side (specifications (2) - (5)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Discussion

The main result of our paper is a positive effect of available thinking time on revealed strategic risk-taking behavior of chess players. While the effect sizes are small-to-medium, the effect is robust across various subsamples and for different weighting schemes of our risk measure. Furthermore, we document the presence of strategic loss aversion, i.e. that the relative disadvantage of a position or blunders in the previous move increase risk-taking in subsequent moves. Loss aversion has been frequently documented since Kahneman and Tversky (1979) in human decision-making, including that of highly skilled professionals such as day traders (Larson et al., 2016), or elite athletes (Pope and Schweitzer, 2011). We thus reiterate that loss aversion is a common phenomenon of decision-making, including that of highly proficient agents.

However, we are agnostic about the specific mechanisms behind these relationships and our data does not allow us to make any claims on that end. In existing work, cognitive processes have been proposed to be responsible for the way people deal with uncertainty under time pressure. For example, Wu et al. (2022) highlight that time pressure coincides with limited cognitive capacity that can be attributed to decision-making. Choosing low-risk options thus represents a remedy to the high cognitive demands associated with decisions under time pressure. The expert chess players in our sample might cope with the external time restrictions and associated cognitive limitations by choosing moves that reduce the variability of outcomes to simplify their ongoing decision process and reduce the necessity to adapt strongly to the opponent's play. This would imply a genuine preference shift towards simpler play under time constraints.

However, in our setting, that might not necessarily be the case: In chess, not only do the players have to make complex strategic decisions, but at the same time they have to allocate their thinking time optimally throughout a game. Both Howard (2024) and Russek et al. (2022) point out that skilled chess players, as the ones in our sample, are particularly good at investing their thinking time in the most relevant positions, whereas less advanced players tend to allocate their time inefficiently. Thus, our main effect of interest may be the outcome of a difference in this resource allocation: With more time available, chess players may choose risky moves more frequently because they know that they would still have enough time to cope with a difficult or disadvantageous position if this position occurs. With less time available, they refrain from doing so.

Consequently, we only document behavioral patterns in our data that are consistent with notions of risk-aversion in case of a low time budget and strategic loss aversion. Players in our sample behave *as if* they are risk-averse under time pressure and strategically loss averse.

7 Conclusion

In this paper, we analyze the effect of time pressure on a novel measure of strategic risk of single chess moves during the FIDE Chess World Cups between 2013-2023. These tournaments permit an instrumental variable approach to obtain exogenous variation in available thinking time. We show that there is a small but robust positive effect of time pressure on revealed risk-averse behavior. We also find evidence for strategic loss aversion in our sample, a higher tendency to play risky moves after a mistake or in a disadvantageous position. This again implies that even highly proficient decision-makers show signs of established behavioral biases.

If we extrapolate from these findings, this implies that also highly-proficient decision-makers in high stake settings do not only focus on factors inherent to the decision problem when choosing between different strategies. Instead, their behavior systematically reacts to context factors such as time pressure, or being in a disadvantageous situation. This has practical implications: Managers who design timelines for bargaining processes, for instance, should take into account that decision-makers react to time pressure. Applying this knowledge in the development of bargaining strategies can provide a strategic advantage and help to predict choices of competitors who have to decide under time pressure.

Our study is not without limitations. In methodological terms, our measure has the shortcoming that it is based on (hypothetical) positions that occur only one move ahead. However, chess players often try to anticipate multiple move sequences. In principle, our measure could be adapted accordingly. However, we would require a model for predicting human play to ensure that the predicted sequences are realistic. This is an extremely ambitious endeavor beyond the scope of this paper.

Nonetheless, our paper innovates on multiple fronts: We develop a novel measure for the degree of strategic risk entailed in single chess moves. We thus advance existing approaches that focus entirely on chess openings. In addition, we introduce a way to measure the difficulty of a chess position, an interesting and transportable quantity by itself. Both our measure of strategic risk on single moves and the difficulty of given chess positions should be insightful for a plethora of possible research questions, putting the dynamic development of strategic risk and difficulty during play into focus.

References

Alós-Ferrer, C. and Garagnani, M. (2022). Voting under time pressure. *Judgment and Decision Making*, 17(5):1072–1093.

Anderson, A., Kleinberg, J., and Mullainathan, S. (2017). Assessing human error against a benchmark of perfection. *ACM Transactions on Knowledge Discovery from Data*, 11(4).

Backus, P., Cubel, M., Guid, M., Sánchez-Pagés, S., and López Mañas, E. (2023). Gender, competition, and performance: Evidence from chess players. *Quantitative Economics*, 14(1):349–380.

Bilalić, M., McLeod, P., and Gobet, F. (2007). Does chess need intelligence? - A study with young chess players. *Intelligence*, 35(5):457–470.

Calford, E. M. (2020). Uncertainty aversion in game theory: Experimental evidence. *Journal of Economic Behavior & Organization*, 176:720–734.

Chassy, P. and Gobet, F. (2015). Risk taking in adversarial situations: Civilization differences in chess experts. *Cognition*, 141:36–40.

Chassy, P. and Gobet, F. (2020). Influence of experts' domain-specific knowledge on risk taking in adversarial situations. *Journal of Expertise*, 3(2):88–100.

Chess24 (2023). Live tournaments. https://chess24.com/en/watch/live-tournaments.

ChessBase (2021). Big Database 2021. https://shop.chessbase.com/en/products/big_database_2021.

de Groot, A. (1965). *Thought and choice in chess.* Amsterdam University Press.

Dilmaghani, M. (2021). A matter of time: Gender, time constraint, and risk taking among the chess elite. *Economics Letters*, 208:110085.

Dreber, A., Gerdes, C., and Gränsmark, P. (2013). Beauty queens and battling knights: Risk taking and attractiveness in chess. *Journal of Economic Behavior and Organization*, 90:1–18.

Ferreira, D. R. (2013). The impact of search depth on chess playing strength. *ICGA Journal*, 36(2):67–80.

FIDE (2021). Regulations for the FIDE World Cup 2021. https://handbook.fide.com/files/handbook/WorldCup2021Regulations.pdf.

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

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

González-Díaz, J. and Palacios-Huerta, I. (2016). Cognitive performance in competitive environments: Evidence from a natural experiment. *Journal of Public Economics*, 139:40–52.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning.* MIT Press.

Gränsmark, P. (2012). Masters of our time: Impatience and self-control in high-level chess games. *Journal of Economic Behavior and Organization*, 82(1):179–191.

Guid, M. and Bratko, I. (2006). Computer analysis of world chess champions. *ICGA Journal*, 29(2):65–73.

Haji, A. E., Krawczyk, M., Sylwestrzak, M., and Zawojska, E. (2019). Time pressure and risk taking in auctions: A field experiment. *Journal of Behavioral and Experimental Economics*, 78:68–79.

Heinecke, A., Ho, J., and Hwang, W.-L. (2020). Refinement and universal approximation via sparsely connected ReLU Convolution Nets. *IEEE Signal Processing Letters*, 27:1175–1179.

Holdaway, C. and Vul, E. (2021). Risk-taking in adversarial games: What can 1 billion online chess games tell us. In *Proceedings of the 43rd Annual Meeting of the Cognitive Science Society*, pages 986–992.

Howard, G. (2024). A check for rational inattention. *Journal of Political Economy Microeconomics.*

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263.

Karagözoglu, E. and Kocher, M. G. (2019). Bargaining under time pressure from deadlines. *Experimental Economics*, 22(2):419–440.

Kirchler, M., Andersson, D., Bonn, C., Johannesson, M., Sørensen, E. O., Stefan, M., Tinghög, G., and Västfjäll, D. (2017). The effect of fast and slow decisions on risk taking. *Journal of Risk and Uncertainty*, 54(1):37–59.

Klingen, J. and van Ommeren, J. N. (2020). Risk attitude and air pollution: Evidence from chess. *Tinbergen Institute Discussion Paper*, 2020-027/V.

Kocher, M. G., Pahlke, J., and Trautmann, S. T. (2013). Tempus fugit: Time pressure in risky decisions. *Management Science*, 59(10):2380–2391.

Kocher, M. G. and Sutter, M. (2006). Time is money: Time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, 61(3):375–392.

Künn, S., Palacios, J., and Pestel, N. (2023). Indoor air quality and strategic decision making. *Management Science*, 69.

Künn, S., Seel, C., and Zegners, D. (2022). Cognitive performance in remote work: Evidence from professional chess. *The Economic Journal*, 132(643):1218–1232.

Larson, F., List, J. A., and Metcalfe, R. D. (2016). Can myopic loss aversion explain the equity premium puzzle? Evidence from a natural field experiment with professional traders. *NBER Working Paper Series.*

Levitt, S. D., List, J. A., and Sadoff, S. E. (2011). Checkmate: Exploring backward induction among chess players. *American Economic Review*, 101(2):975–990.

Lindner, F. (2014). Decision time and steps of reasoning in a competitive market entry game. *Economics Letters*, 122(1):7–11.

Lindner, F. and Sutter, M. (2013). Level-k reasoning and time pressure in the 11–20 money request game. *Economics Letters*, 120(3):542–545.

Linnemer, L. and Visser, M. (2016). Self-selection in tournaments: The case of chess players. *Journal of Economic Behavior and Organization*, 126:213–234.

Maharaj, S., Polson, N., and Turk, A. (2022). Chess AI: competing paradigms for machine intelligence. *Entropy*, 24(4):550.

McIlroy-Young, R., Sen, S., Kleinberg, J., and Anderson, A. (2020). Aligning superhuman AI with human behavior. pages 1677–1687.

Olschewski, S. and Rieskamp, J. (2021). Distinguishing three effects of time pressure on risk taking: Choice consistency, risk preference, and strategy selection. *Journal of Behavioral Decision Making*, 34(4):541–554.

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

Pope, D. G. and Schweitzer, M. E. (2011). Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. *American Economic Review*, 101(1):129–157.

Russek, E. M., Acosta-Kane, D., Van Opheusden, B., Mattar, M. G., and Griffiths, T. L. (2022). Time spent thinking in online chess reflects the value of computation. *PsyArxiv*.

Salant, Y. and Spenkuch, J. L. (2022). Complexity and choice. *NBER Working Papers*.

Saloner, G. (1991). Modeling, game theory, and strategic management. *Strategic Management Journal*, 12(S2):119–136.

Sigman, M., Etchemendy, P., Slezak, D. F., and Cecchi, G. A. (2010). Response time distributions in rapid chess: A large-scale decision making experiment. *Frontiers in Neuroscience*, 4(OCT):1–12.

Spiliopoulos, L., Ortmann, A., and Zhang, L. (2018). Complexity, attention, and choice in games under time constraints: A process analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(10):1609–1640.

Steinerberger, S. (2015). On the number of positions in chess without promotion. *International Journal of Game Theory*, 44(3):761–767.

Strittmatter, A., Sunde, U., and Zegners, D. (2020). Life cycle patterns of cognitive performance over the long run. *Proceedings of the National Academy of Sciences*, 117(44):27255–27261.

Sunde, U., Zegners, D., and Strittmatter, A. (2022). Speed, quality, and the optimal timing of complex decisions: Field evidence. (280092119).

Wu, C. M., Schulz, E., Pleskac, T. J., and Speekenbrink, M. (2022). Time pressure changes how people explore and respond to uncertainty. *Scientific Reports*, 12(1).

Zegners, D., Sunde, U., and Strittmatter, A. (2020). Decisions and performance under bounded rationality: A computational benchmarking approach. *Collaborative Research Center Transregio 190, Discussion Paper No. 263.*

Zhou, D.-X. (2020). Universality of deep convolutional neural networks. *Applied and Computational Harmonic Analysis*, 48(2):787–794.

A Appendix

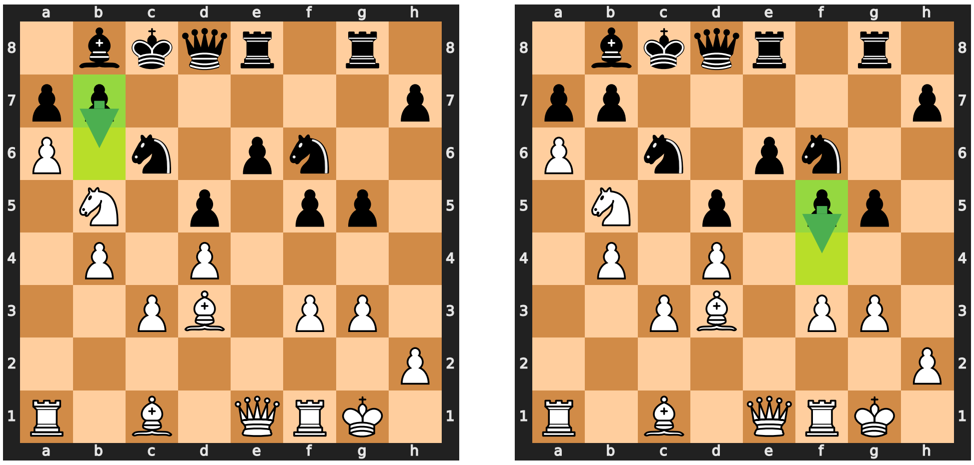
A.1 About Chess Engines

Chess engines are programs to provide evaluations of both chess positions and possible moves. Chess is considered to be a solvable game (Backus et al., 2023), .i.e. in each position exists (at least) one objectively optimal move. Identifying this move is a computational problem where computers excel and consequently, are able to back chess engines of playing strength, which are several orders of magnitude more capable than the best human chess player. Detailed explanations how chess engines such as *Stockfish* work are provided by Zegners et al. (2020), Künn et al. (2022) or Maharaj et al. (2022) and shortly summarized in this section.

Chess engines evaluate positions based on a variety of characteristics of the position, such as the available material of both players or the king safety. The evaluation functions are complex, engine-specific and beyond the scope of this paper. The unit of evaluations of both positions and moves are so-called pawn units where a pawn is the weakest chess piece. To determine the best move, the engine considers a decision tree with a specific search depth. For instance, a search depth of 20 implies that the decision tree reaches positions which can arise after 20 half-moves, starting from the current position on the board. Many positions can entail 25 or more legal moves, making it computationally intensive to consider all nodes of the decision tree. Hence, modern chess engines use algorithms to reduce the complexity of decision trees by excluding irrelevant decision nodes (Zegners et al., 2020). Moreover, the position evaluations by the engine are based on the assumption of mutual best responses. When the engine displays an advantage of one pawn for Black for a given position, or equivalently a disadvantage of one pawn for White, this means that the engine analyzed a variation of length 15 of mutual best responses that ends in a position where Black has an advantage equivalent to one pawn according to the evaluation function of the corresponding board.

Figure A1 depicts the initial position of the example outlined in Figure 1 with both the human move 21... b7-b6 (in standard chess notation) as well as the superior move suggested by the chess engine Stockfish, 21... f5-f4. Stockfish evaluates 21... b7-b6 to be 115 CP, indicating a mistake by Black. In general, the position is characterized by opposite-castled kings, implying that White's king is on the kingside (files e-h) whereas Black's king is on the queenside (files a-d). This motivates both players to attack the opponent's king. White's last move 21. a5-a6 pursues the goal of creating weaknesses in Black's camp. Consequently, White is ahead in the race of attacking the opponent's king. Therefore, the engine assigns an advantage of 77 centipawns to White. In the game, Black replied 21... b7-b6 which is considered to be only the fifth best response and which worsens Black's position by an additional 115 centipawns. The first best response would have been 21... f5-f4 to create counter-chances against White's king.

Appendix Figure A1: Illustrative example for the evaluations of chess engines: game Brkic vs Barrientos, FIDE World Cup 2021, position after 21. a5-a6



Barrientos move: 21... b7-b6 | Better alternative: 21... f5-f4

Note: The graph depicts the position after White's move 21. a5-a6 (in standard chess notation). In the game, Black replied 21... b7-b6 whereas the engine suggests the move 21... f5-f4. According to the engine, the game continuation 21... b7-b6 is 115 CP worse than the move 21... f5-f4.

A.2 Details on the Neural Network

This part of the appendix provides details on the convolutional neural network, more specifically on its training data, its model architecture as well as on typical measures logged during training. We also provide a small comparison with other types of models with respect to the predictive power in a hold-out test set.

A.2.1 Basics and Feature Representation

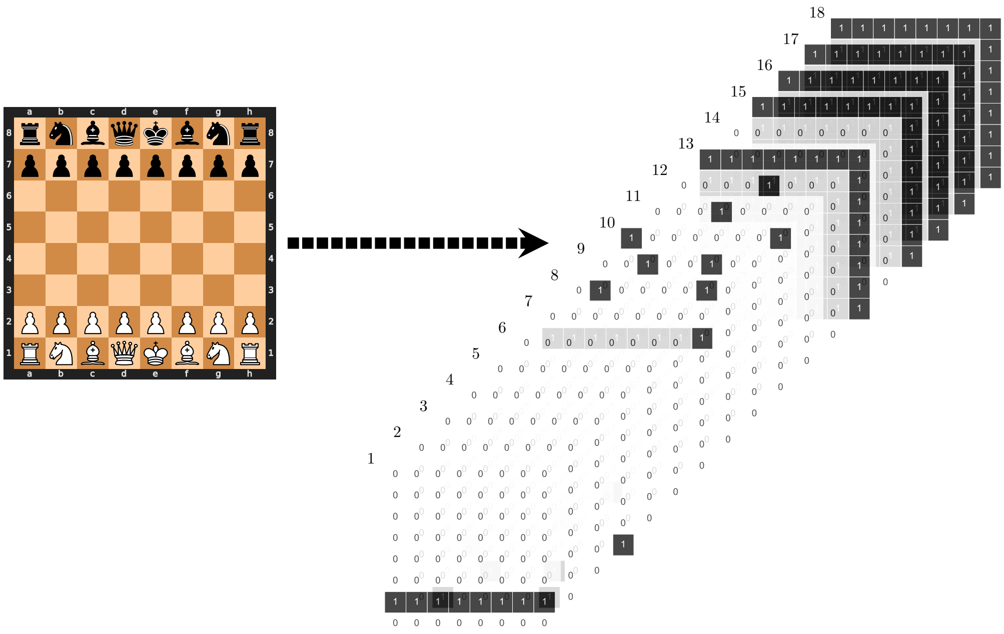
We chose a neural network architecture over other supervised machine learning algorithms mainly because neural networks have been heavily employed in chess and similar domains. For example, the reinforcement-learning agent of AlphaGo, which beat Ke Jie, the then-number-one ranked Go player in 2017, was based on a convolutional neural network to represent the Go board as input for the model. MaiaChess (McIlroy-Young et al., 2020), an engine designed to mimic human chess play, is also based on a residual convolutional neural network, from which our model takes heavy inspiration. We show in Table A2 that in fact our neural network achieves higher prediction accuracy in our setup compared to other often-used models.

Input Features The input features of the neural network are based on the so-called FEN-strings. A FEN-string describes a given chess position in a particular format and contains a sequence of characters that indicate on which of the 64 squares of a chess board each of the 12 different pieces (6 different pieces from two colors) is located. In addition, the FEN-string contains information on the color to move, remaining king- or queenside castling rights of both players, and whether en-passant[28] is possible in the position. For example, the FEN-string of the initial position in Figure 1 is:

```
1bkqr1r1/pp5p/P1n1pn2/1N1p1pp1/1P1P4/2PB1PP1/7P/R1B1QRK1 b - -
```

To translate this string into numerical information, we build on the work by McIlroy-Young et al. (2020) and convert a given FEN-string into a 8x8x18 tensor. This tensor has 18 so-called channels, each of which are of dimension 8x8. Each channel conveys one part of the information of the chess board e.g. indicating on which of the 64 squares a Black pawn is currently present.

Appendix Figure A2: Chess Position to 18x8x8 Tensor



Note: This graph showcases how the base position of chess can be translated into an 18x8x8 tensor. The FEN-string of the base position is `rnbqkbnr/pppppppp/8/8/8/8/PPPPPPPP/RNBQKBNR w KQkq - 0 1`.

Figure A2 provides a graphical example of this approach, converting the base position of chess into 18 8x8 matrices. The 18 channels contain the positional information on the 12 pieces and information on who is on the move (one channel), if castling is possible (4 channels for the two colors and long/short) and if en-passant is currently possible (one channel). These non-square-specific features of the FEN-String are represented as 64 1s or 0s in their respective channel. We thus mirror the input representation by McIlroy-Young et al. (2020) and extend it with a channel for the en-passant possibility. We discuss the advantages of this representation for the convolutional neural network in Section A.2.2.

Target For the target, we use the quality of a human move in a given position as measured by the chess engine Stockfish 11. The target is equal to 0 if the human player implemented

[28] *en-passant* refers to a special way of capturing a pawn of the opponent.

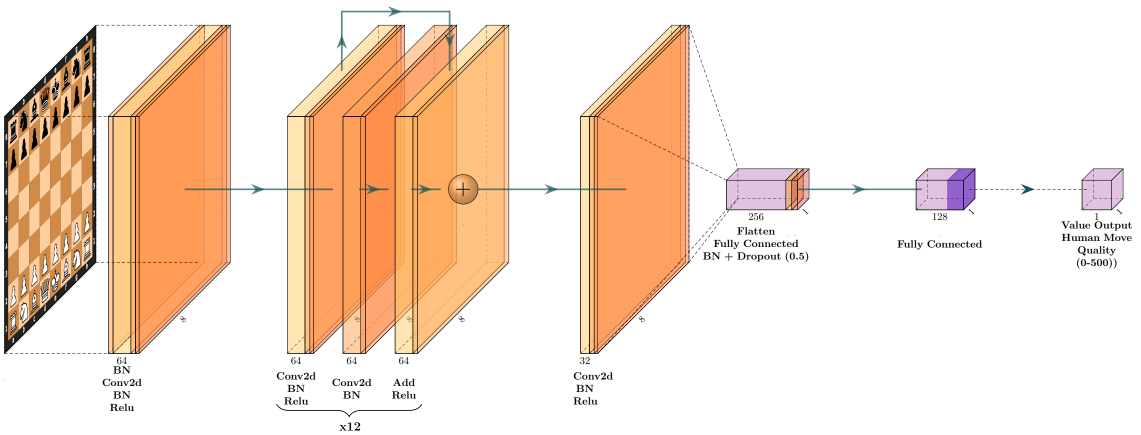
the optimal (or an equivalent) move in that position and deviates more strongly from 0 the worse the move was. Thus, we have a *regression problem*, where our targets are measured on a continuous scale. The objective of our model is to predict this expected quality deviation on a given chess board. The higher the predicted quality deviation, the more difficult a given chess board is for a human player.

A.2.2 The Model

For model architecture, we again build on McIlroy-Young et al. (2020) and similarly train a residual convolutional neural network (RCNN). We chose an input representation of chess boards similar to images (see above). Without explaining how convolutional neural networks work in detail, what is important is that they are able to learn *local features*, i.e. that parts of an input are relevant for a given prediction. In the image analogy, this means that convolutional layers are able to learn over time that a handwritten two is usually curved at its top, but not at its bottom. In our case, this implies that the model could in principle learn the importance of the centrality of the opponents' queen for the move quality prediction. In addition, convolutional neural nets have fewer parameters than e.g. classical multi-layer perceptrons, which allows them to be deeper (learning more complex relationships) and be trained more efficiently on large datasets.

Figure A3 depicts a graphical representation of our network architecture and Table A1 provides more details on the chosen hyper-parameters.

Appendix Figure A3: Residual Convolutional Neural Network Architecture



Note: This graph depicts a graphical representation of the chosen neural network architecture.

Appendix Table A1: Hyperparameters Residual Convolutional Neural Network

Hyperparameter	Value
Residual Blocks	12
Filters	64
Padding	valid (stride 1)
Filter Size	3x3
Activation	ReLU
Dropout (Fully Connected Layer)	50%
Optimiser	Adam
Learning Rate	0.0002
Batch Size	4096
Epochs	30 (only best weights kept)
Loss Function	Mean-Squared Error
Framework	Tensorflow 2.1.0

Note: This table depicts the chosen hyperparameters of the residual convolutional neural network.

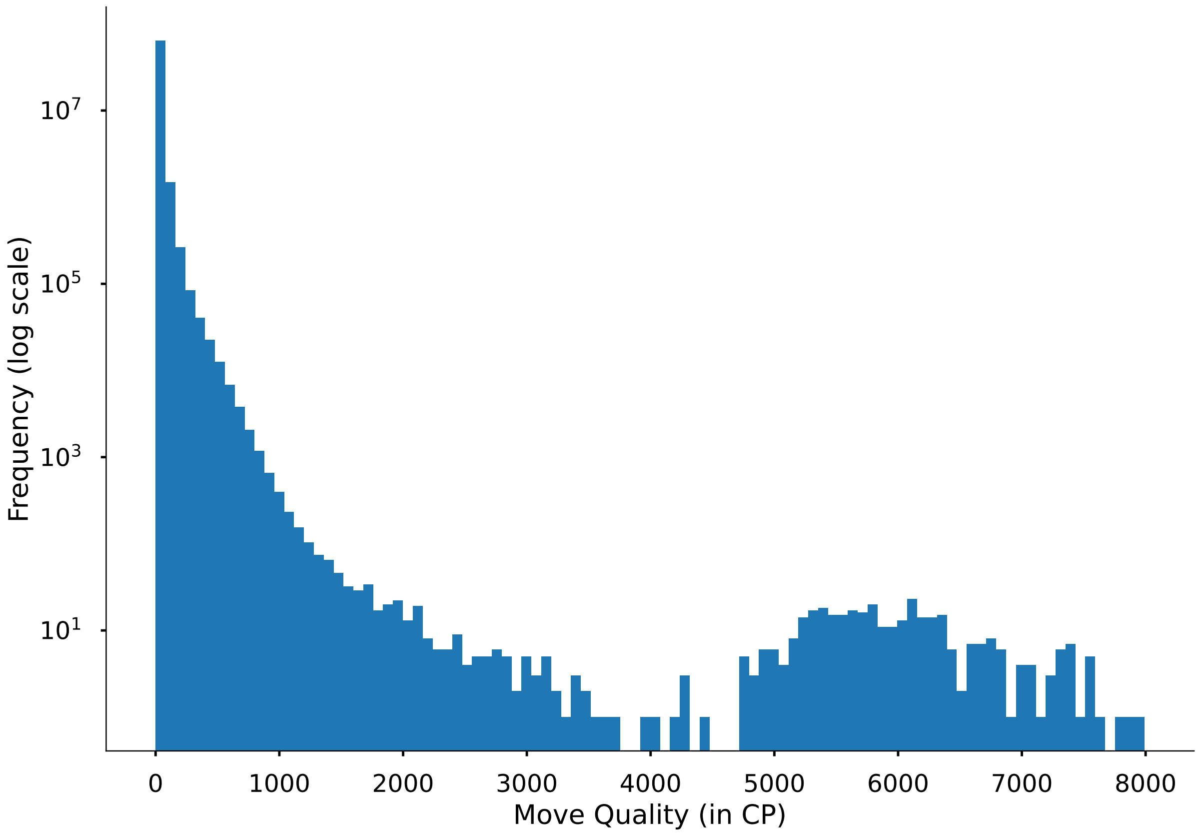
A.2.3 Training Data

The training data consists of approximately 1.9 million chess games, played by strong professionals (i.e. Elo $\geq$ 2400), between 1975 and 2021. The games stem from the *Big Database 2021* from ChessBase (2021) and include a broad range of settings such as open tournaments, championships or Olympiads. We excluded all games from the FIDE World Cups. Our training sample thus contains games from a wide variety of playing modes and varying time budgets.

For training our neural network, we only consider a subset of moves: First, we only keep moves by players with Elo $\geq$ 2400. This ensures to only have decisions by very proficient decision-makers in the dataset. Next, we only include half-moves from 1 to 80, excluding endgames. As stated in the main text, endgames are often characterized by less space for strategic considerations. Furthermore, we exclude already decided positions, i.e. which Stockfish evaluates at an (dis)advantage of $>$ 200 CP for one of the players. The main reason for this is that in such positions, again there often is only a single move retaining the disadvantage for the player behind while every other move worsens the position considerably. Keeping such positions would inflate our dataset with moves labeled as serious mistakes, which, however, are not the result of a difficult position, but more an artifact of how Stockfish evaluates decided positions.

This finally results in 70,819,615 moves in our training dataset. Figure A4 depicts a histogram of their move quality evaluation (frequency on a log scale). In the majority of cases (57.82 %), the players in our final training sample implemented the optimal (or an equivalent) move, where the quality evaluation is 0 CP. Afterwards, the quality is distributed similar to a half-normal distribution. The increased frequency starting from around 5000 CP is mainly due to moves

Appendix Figure A4: Distribution of Move Quality in Training Sample



Note: This histogram depicts the distribution of move quality evaluations in CP of 70,819,615 moves in the training dataset (on a log scale). In 57.82 % of moves, the best or an equivalent move (i.e. a move quality of 0 CP) was chosen.

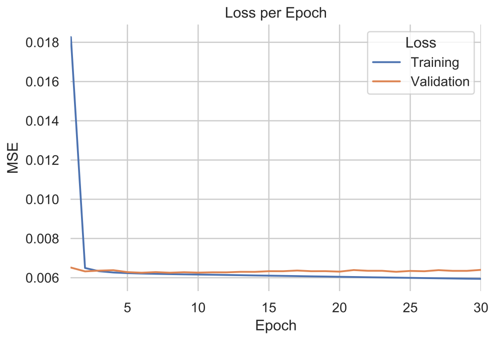
which lead to a check-mate in a given number of moves (which is then translated into the CP quality rating).

A.2.4 Training Process

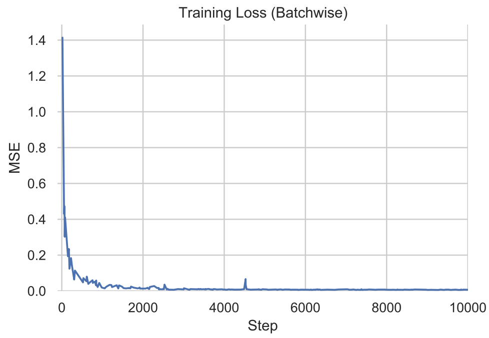
Prior to training, we clip the quality ratings to 500, to reduce the impact of outliers, while retaining their informativeness (instead of dropping them). This also helps to keep the eventual predictions within reasonable bounds. We then transform the clipped data using a min-max scaler, i.e. scaling them on a 0 (= 0 CP) to 1 (=500 CP) scale to ease updating of the weights during training. We split the dataset in 80% training and 20% validation sample.

Appendix Figure A5: Epoch and Batchwise Loss during Training

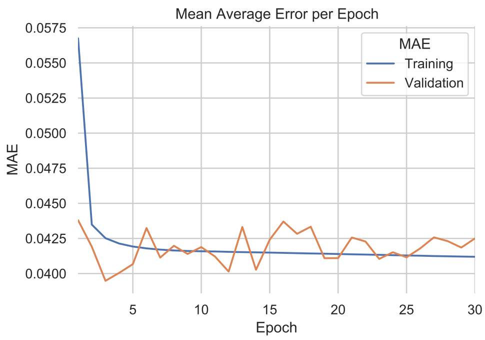
(a) Epoch Loss



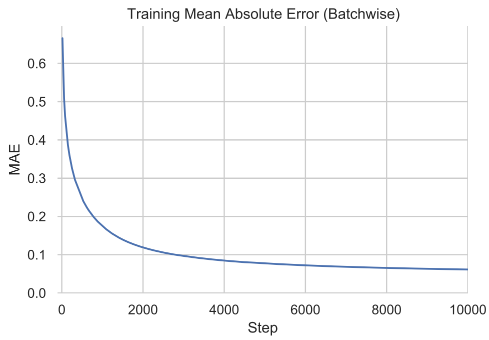
(b) Batchwise Loss



(c) Epoch MAE



(d) Batchwise MAE



Notes: This graph shows epoch- (panel a + c) and batchwise (panel b + d) development of the loss (mean-squared error (MSE)) as well as the mean-average error (MAE). In total, we rain training for 30 epochs, but saved only the model with the lowest test loss (epoch 6, MSE = 0.00626). Training with batch size 4096 ran in total for 414,929 (30 epochs) steps, but only 10,000 steps are shown in panel b.

Figure A5 depicts the epoch and batchwise mean-squared and mean-average error in both the training and validation set. For reference, an MAE of 0.042 implies a mean average deviation of 21 CP.

A.2.5 Comparison with Other Models

We compare our RCNN to a range of other possible models: A simple, two layer (1024 and 512 neurons) multi-layer-perceptron with BatchNorm and Dropout (50 percent) trained for 10 epochs, featuring the same learning rate of 0.0002. This MLP features a flat input representation, i.e. ignoring the spatial structure of the chess board. The MLP performs comparably, but has a higher MSE and MAE compared to the RCNN. We also test a Random Forest Regressor and a simple ridge regression. As these methods are unable to be computed batchwise, we need to train them on a subsample of 10% of the initial training data (which is similar to the amount of steps until which the RCNN does not improve significantly anymore (see Figure A5)). Both alternatives perform worse compared to the RCNN. We also tried a support-vector machine (SVM) regressor, but failed due to timeout or memory issues (even when using only 10 percent of training data).

Appendix Table A2: Comparison of Model Performance on hold-out Test Set

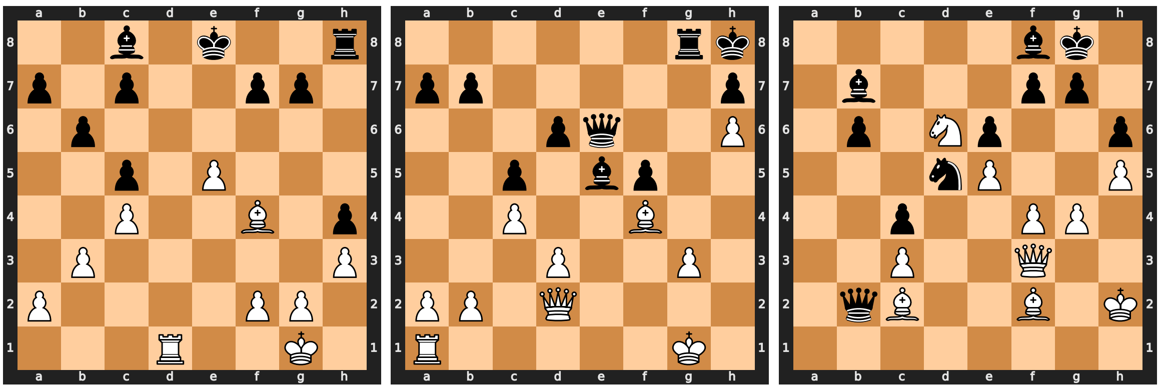
Model	MSE	MAE
RCNN	0.00626	0.0416
MLP	0.00631	0.0417
Random Forest Regressor	0.00658	0.0442
Ridge Regression	0.00644	0.0437

Notes: This table displays both the Mean-squared error (MSE) and the Mean Average Error (MAE) on the test-set for the Residual Convolutional Neural Network (RCNN); a two-layer Multi-layer Perceptron (MLP) architecture, a Random Forest Regressor as well as a Ridge Regression. Note that the latter two are only trained on 10% of training data due to memory limitations.

A.2.6 Example Positions for Low, Medium and High Difficulty

Appendix Figure A6: Examples Difficulty

(a) Low Difficulty (for Black) **(b)** Medium Difficulty (for White) **(c)** High Difficulty (for White)

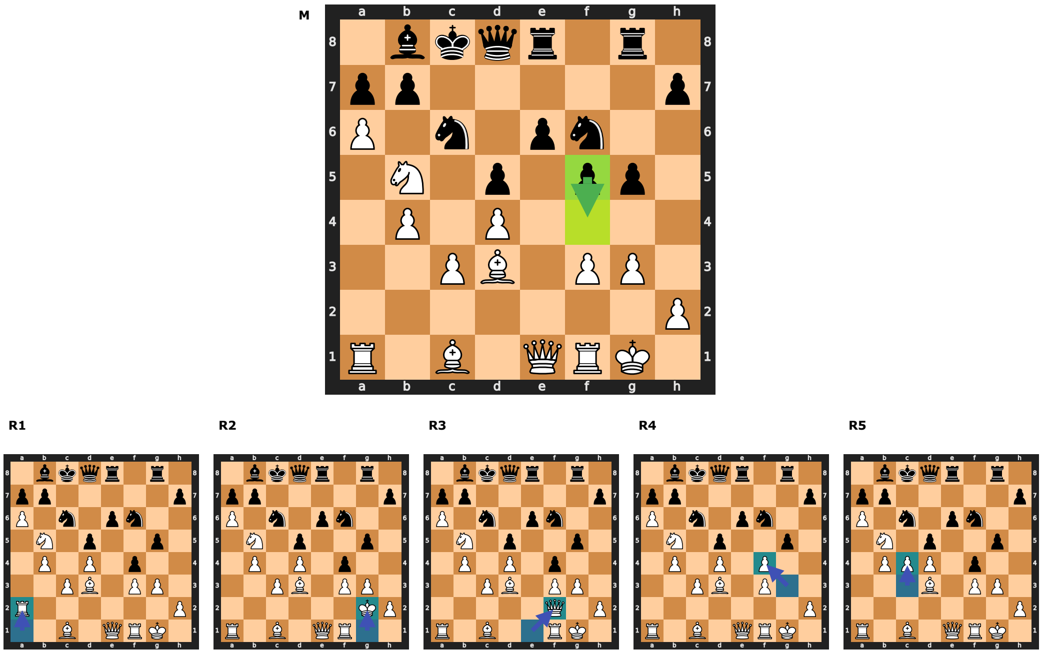


Notes: Low, medium and high difficulty examples. **(a)** Istratescu - Lysyj, World Cup 2013, Black to move. Endings with opposite-colored bishops offer a large drawing margin. Further, Black can comfortably protect the position against the invasion of the White pieces. White cannot pose specific threats to the opponent. Predicted difficulty: 10.20 CP. **(b)** Anton Guijarro - Wei, World Cup 2019, White to move: Both kings are slightly weakened: The pawn on h6 creates a permanent threat for Black's king as it supports later checkmate attacks of the White queen. White, however, has to deal with the attack of the Black rook via the half-open g-file. In the next moves, Black can increase the pressure by transferring the queen to the king side to create more threats. Predicted difficulty: 19.75 CP. **(c)** Caruana - Lenic, World Cup 2017, White to move: Several pieces are attacked at the same time (the White knight on d6 attacks the bishop on b7 while the Black queen on b2 threatens to take the bishop on c2 or the pawn on c3. Further, king safety is rather low for both players: White weakened his king by advancing the pawns on the king side so that they do not provide shelter for the White king. White, in contrast, can move the queen to e4 to threaten checkmate by playing Qh7. Predicted difficulty: 93.47 CP

A.3 Additional Figures and Tables

A.3.1 Engine Move f4

Appendix Figure B1: Illustrative example for an engine recommendation: game Brkic vs Barrientos, FIDE World Cup 2021, position after White's move 21. a5-a6. Response for the Black player suggested by the engine: 21... f5-f4.



Note: The upper panel of the figure shows a position that occurred in the game Brkic - Barrientos, FIDE World Cup 2021. In standard chess notation, White's last move was to advance the pawn to a6 (21. a5-a6). In the diagram position, Black reacted by playing 21... b7-b6. However, the engine suggests the move 21... f5-f4 (highlighted in green). In the resulting position after the optimal move 21... f5-f4, the 5 best responses (shown in the smaller diagrams below) for White according to the chess engine Stockfish 11 are 22. Ra1-a2 (R1), 22. Kg1-g2 (R2), 22. Qe1-f2 (R3), 22. g3xf4 (R4) and 22. c3-c4 (R5).

A.3.2 World Cup Setup and Time Limits

Appendix Figure B2: The World Cup tournament setting

Round 1

Match 1	Classic		Rapid	
Game	1	2	3	4
Player 1	0	1	0	0
Player 2	1	0	1	1

Match 2	Classic	
Guten Game	1	2
Player 3	0.5	1
Player 4	0.5	0

...

Match 63	Classic		Rapid	
Game	1	2	3	4
Player 125	0.5	0.5	0.5	0
Player 126	0.5	0.5	0.5	1

Match 64	Classic		Rapid		Slow Blitz	
Game	1	2	3	4	5	6
Player 127	1	0	0	1	0.5	1
Player 128	0	1	1	0	0.5	0

Round 2

Match 65	Classic	
Guten Game	1	2
Player 2	0.5	1
Player 3	0.5	0

...

Match 96	Classic	
Guten Game	1	2
Player 126	0	0.5
Player 127	1	0.5

...

Note: The graph shows the tournament setting of a World Cup including the playing modes *Classic, Rapid, Slow Blitz* and *Fast Blitz*: In Round 1, there are 128 players who compete in 64 matches. The winner of each match qualifies for Round 2. The match consists of two classic games if one player leads (as in matches 2, 65 and 96). In case of a tie after the two classic games, two rapid games are played (matches 1, 63, 64). If the match is still tied, even faster playing modes are applied (Slow Blitz in match 64, further modes would be Fast Blitz and Armageddon). One World Cup consists of seven Rounds (where Rounds 3 – 7 are omitted for reasons of space).

Appendix Table B1: Overview of the playing modes

Playing mode	Time budget
Classic	90 minutes + 30 seconds per move, additional bonus of 30 minutes after move 40
Rapid	25 minutes + 10 seconds per move
Slow Blitz	10 minutes + 10 seconds per move
Fast Blitz	5 minutes + 3 seconds per move

Note: This table depicts the time budgets of the playing modes applied in the World Cup, disentangled by initial thinking time and increment per move played.

A.3.3 Evolution of Risk and First Stage

Appendix Figure B3: Evolution of risk within games

Note: This graph shows how the risk measure employed in this study evolves within games of the four playing modes *Classic*, *Rapid*, *Slow Blitz* and *Fast Blitz*. It is computed as the logarithm of the variance of difficulty of the hypothetical boards that can arise after the played move. The sample consists of World Cup games from 2013 – 2023.

Appendix Table B2: IV first stage estimates of (2). The dependent variable is the remaining time before the move.

Initial Time × Half-Move 3	0.0034***	Initial Time × Half-Move 55	-0.7579***
	(0.0005)		(0.0067)
Initial Time × Half-Move 5	0.0041***	Initial Time × Half-Move 57	-0.7781***
	(0.0008)		(0.0064)
Initial Time × Half-Move 7	0.0045***	Initial Time × Half-Move 59	-0.7890***
	(0.0010)		(0.0066)
Initial Time × Half-Move 9	0.0020*	Initial Time × Half-Move 61	-0.8133***
	(0.0012)		(0.0064)
Initial Time × Half-Move 11	-0.0014	Initial Time × Half-Move 63	-0.8316***
	(0.0016)		(0.0064)
Initial Time × Half-Move 13	-0.0085***	Initial Time × Half-Move 65	-0.8421***
	(0.0019)		(0.0064)
Initial Time × Half-Move 15	-0.0226***	Initial Time × Half-Move 67	-0.8632***
	(0.0023)		(0.0063)
Initial Time × Half-Move 17	-0.0411***	Initial Time × Half-Move 69	-0.8656***
	(0.0028)		(0.0064)
Initial Time × Half-Move 19	-0.0671***	Initial Time × Half-Move 71	-0.8764***
	(0.0034)		(0.0062)
Initial Time × Half-Move 21	-0.0957***	Initial Time × Half-Move 73	-0.8869***
	(0.0040)		(0.0064)
Initial Time × Half-Move 23	-0.1301***	Initial Time × Half-Move 75	-0.8999***
	(0.0046)		(0.0064)
Initial Time × Half-Move 25	-0.1728***	Initial Time × Half-Move 77	-0.9033***
	(0.0053)		(0.0066)
Initial Time × Half-Move 27	-0.2258***	Initial Time × Half-Move 79	-0.9142***
	(0.0061)		(0.0067)
Initial Time × Half-Move 29	-0.2744***	Initial Time × Half-Move 4	-0.0000
	(0.0064)		(0.0007)
Initial Time × Half-Move 31	-0.3233***	Initial Time × Half-Move 6	0.0017*
	(0.0069)		(0.0009)
Initial Time × Half-Move 33	-0.3711***	Initial Time × Half-Move 8	0.0016
	(0.0073)		(0.0013)
Initial Time × Half-Move 35	-0.4126***	Initial Time × Half-Move 10	-0.0000
	(0.0075)		(0.0015)
Initial Time × Half-Move 37	-0.4628***	Initial Time × Half-Move 12	-0.0046**
	(0.0076)		(0.0018)
Initial Time × Half-Move 39	-0.5032***	Initial Time × Half-Move 14	-0.0170***
	(0.0077)		(0.0021)
Initial Time × Half-Move 41	-0.5469***	Initial Time × Half-Move 16	-0.0312***
	(0.0075)		(0.0026)
Initial Time × Half-Move 43	-0.5833***	Initial Time × Half-Move 18	-0.0526***
	(0.0074)		(0.0031)
Initial Time × Half-Move 45	-0.6174***	Initial Time × Half-Move 20	-0.0814***
	(0.0074)		(0.0039)
Initial Time × Half-Move 47	-0.6441***	Initial Time × Half-Move 22	-0.1143***
	(0.0075)		(0.0044)
Initial Time × Half-Move 49	-0.6781***	Initial Time × Half-Move 24	-0.1528***
	(0.0073)		(0.0051)
Initial Time × Half-Move 51	-0.7097***	Initial Time × Half-Move 26	-0.2013***
	(0.0068)		(0.0058)
Initial Time × Half-Move 53	-0.7334***	Initial Time × Half-Move 28	-0.2455***
	(0.0067)		(0.0064)

Continued on next page

Appendix Table B2 – continued from previous page

Initial Time × Half-Move 30	-0.2963***	Half-Move 3	0.0555*
	(0.0068)		(0.0312)
Initial Time × Half-Move 32	-0.3442***	Half-Move 5	-0.1003***
	(0.0072)		(0.0355)
Initial Time × Half-Move 34	-0.3949***	Half-Move 7	-0.2557***
	(0.0073)		(0.0436)
Initial Time × Half-Move 36	-0.4364***	Half-Move 9	-0.1030**
	(0.0075)		(0.0492)
Initial Time × Half-Move 38	-0.4833***	Half-Move 11	0.0013
	(0.0076)		(0.0539)
Initial Time × Half-Move 40	-0.5278***	Half-Move 13	0.2177***
	(0.0076)		(0.0625)
Initial Time × Half-Move 42	-0.5647***	Half-Move 15	0.4319***
	(0.0074)		(0.0734)
Initial Time × Half-Move 44	-0.6013***	Half-Move 17	0.5600***
	(0.0073)		(0.0858)
Initial Time × Half-Move 46	-0.6337***	Half-Move 19	1.0002***
	(0.0073)		(0.1018)
Initial Time × Half-Move 48	-0.6646***	Half-Move 21	1.1815***
	(0.0073)		(0.1229)
Initial Time × Half-Move 50	-0.6945***	Half-Move 23	1.3137***
	(0.0070)		(0.1407)
Initial Time × Half-Move 52	-0.7165***	Half-Move 25	1.6410***
	(0.0071)		(0.1614)
Initial Time × Half-Move 54	-0.7404***	Half-Move 27	2.0515***
	(0.0069)		(0.1817)
Initial Time × Half-Move 56	-0.7603***	Half-Move 29	2.2456***
	(0.0069)		(0.1937)
Initial Time × Half-Move 58	-0.7867***	Half-Move 31	2.4418***
	(0.0066)		(0.2057)
Initial Time × Half-Move 60	-0.7985***	Half-Move 33	2.6750***
	(0.0067)		(0.2213)
Initial Time × Half-Move 62	-0.8171***	Half-Move 35	2.6741***
	(0.0066)		(0.2280)
Initial Time × Half-Move 64	-0.8334***	Half-Move 37	2.9291***
	(0.0065)		(0.2303)
Initial Time × Half-Move 66	-0.8526***	Half-Move 39	2.9404***
	(0.0063)		(0.2385)
Initial Time × Half-Move 68	-0.8625***	Half-Move 41	3.1351***
	(0.0062)		(0.2428)
Initial Time × Half-Move 70	-0.8736***	Half-Move 43	3.1650***
	(0.0062)		(0.2438)
Initial Time × Half-Move 72	-0.8759***	Half-Move 45	3.1454***
	(0.0063)		(0.2459)
Initial Time × Half-Move 74	-0.8880***	Half-Move 47	2.9208***
	(0.0064)		(0.2467)
Initial Time × Half-Move 76	-0.8960***	Half-Move 49	2.8759***
	(0.0066)		(0.2447)
Initial Time × Half-Move 78	-0.9078***	Half-Move 51	3.0088***
	(0.0060)		(0.2370)
Initial Time × Half-Move 80	-0.9070***	Half-Move 53	2.7912***
	(0.0066)		(0.2359)

Continued on next page

Appendix Table B2 – continued from previous page

Half-Move 55	2.8677***	Half-Move 30	1.7546***
	(0.2384)		(0.2016)
Half-Move 57	2.6842***	Half-Move 32	1.9398***
	(0.2322)		(0.2136)
Half-Move 59	2.4031***	Half-Move 34	2.1257***
	(0.2348)		(0.2170)
Half-Move 61	2.6013***	Half-Move 36	2.2445***
	(0.2348)		(0.2227)
Half-Move 63	2.5212***	Half-Move 38	2.3320***
	(0.2318)		(0.2288)
Half-Move 65	2.4426***	Half-Move 40	2.5122***
	(0.2359)		(0.2351)
Half-Move 67	2.4074***	Half-Move 42	2.4982***
	(0.2269)		(0.2322)
Half-Move 69	2.0953***	Half-Move 44	2.4811***
	(0.2261)		(0.2385)
Half-Move 71	1.9521***	Half-Move 46	2.4446***
	(0.2173)		(0.2425)
Half-Move 73	1.8451***	Half-Move 48	2.4045***
	(0.2203)		(0.2379)
Half-Move 75	1.9095***	Half-Move 50	2.4213***
	(0.2186)		(0.2330)
Half-Move 77	1.6549***	Half-Move 52	2.2664***
	(0.2175)		(0.2285)
Half-Move 79	1.6805***	Half-Move 54	2.2895***
	(0.2212)		(0.2354)
Half-Move 4	-0.4858***	Half-Move 56	2.1619***
	(0.0467)		(0.2311)
Half-Move 6	-0.5327***	Half-Move 58	2.3983***
	(0.0503)		(0.2307)
Half-Move 8	-0.6583***	Half-Move 60	2.1151***
	(0.0675)		(0.2289)
Half-Move 10	-0.7574***	Half-Move 62	1.9688***
	(0.0733)		(0.2273)
Half-Move 12	-0.5572***	Half-Move 64	1.9407***
	(0.0790)		(0.2243)
Half-Move 14	-0.3067***	Half-Move 66	2.1744***
	(0.0821)		(0.2272)
Half-Move 16	-0.1874**	Half-Move 68	1.8924***
	(0.0909)		(0.2260)
Half-Move 18	-0.0039	Half-Move 70	1.8518***
	(0.1049)		(0.2271)
Half-Move 20	0.2771**	Half-Move 72	1.5422***
	(0.1193)		(0.2265)
Half-Move 22	0.6165***	Half-Move 74	1.4672***
	(0.1354)		(0.2233)
Half-Move 24	0.8742***	Half-Move 76	1.3880***
	(0.1563)		(0.2254)
Half-Move 26	1.2689***	Half-Move 78	1.5152***
	(0.1748)		(0.2015)
Half-Move 28	1.4698***	Half-Move 80	1.1218***
	(0.1881)		(0.2160)

Continued on next page

Appendix Table B2 – continued from previous page

Relative Advantage	0.0020***
	(0.0002)
Position Difficulty	-0.0380***
	(0.0050)
Log. Average Variance Difficulty of Alternatives	-0.6126***
	(0.0352)
Constant	64.3334***
	(0.1763)
Interacted player-match-game FE	Yes
Half-Move Dummies	Yes
Kleibergen-Paap F-statistic	519.100
Observations	162024
R^2	0.948

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable *Remaining Time* varies on the half-move-level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B3: Placebo Test: Linear probability model on the match-level. The dependent variable is a dummy indicating whether the match reaches the Slow Blitz tiebreaks (Panel A) or the Fast Blitz tiebreaks (Panel B).

Panel A: Reaching Slow Blitz

	(1)	(2)	(3)	(4)
Average Risk in Rapid Games	0.0913	0.0910	0.1068	0.1008
	(0.1623)	(0.1616)	(0.1589)	(0.1597)
Average Risk of Alternatives in Rapid Games	-0.2936*	-0.1198	-0.0907	-0.0917
	(0.1684)	(0.1941)	(0.1922)	(0.1929)
Average Position Difficulty in Rapid Games		-0.0199**	-0.0181*	-0.0171*
		(0.0098)	(0.0097)	(0.0097)
Average Move Quality in Rapid Games			-0.0040***	-0.0041***
			(0.0014)	(0.0014)
Absolute Rating Difference in Match				-0.0008**
				(0.0004)
Constant	0.9645***	0.8502***	0.7725***	0.8407***
	(0.1901)	(0.2009)	(0.2012)	(0.2055)
Observations	390	390	390	390
R^2	0.029	0.038	0.054	0.067

Panel B: Reaching Fast Blitz

	(1)	(2)	(3)	(4)
Average Risk in Slow Blitz Games	-0.1723	-0.0799	-0.0418	-0.0292
	(0.2780)	(0.2829)	(0.2851)	(0.2878)
Average Risk of Alternatives in Slow Blitz Games	0.0867	0.2891	0.2826	0.2577
	(0.3097)	(0.3218)	(0.3190)	(0.3245)
Average Position Difficulty in Slow Blitz Games		-0.0349**	-0.0326*	-0.0318*
		(0.0173)	(0.0171)	(0.0172)
Average Move Quality in Slow Blitz Games			-0.0029	-0.0028
			(0.0019)	(0.0019)
Absolute Rating Difference in Match				0.0004
				(0.0007)
Constant	0.5412	0.4108	0.3628	0.3584
	(0.3456)	(0.3483)	(0.3431)	(0.3447)
Observations	135	135	135	135
R^2	0.009	0.038	0.049	0.051

Note: Robust standard errors in parentheses. The sample consists of World Cup games from 2013 – 2023. Panel A comprises all matches that reached the Rapid tiebreak while Panel B is restricted to matches that reached the Slow Blitz tiebreak. The covariate *Average Risk of Alternatives in Rapid (Slow Blitz) games* represents the match-level average of the variable *Log. Average Variance Difficulty of Alternatives* as included in the other specifications. * p < 0.10, ** p < 0.05, *** p < 0.01

A.3.4 Robustness analyzes

Appendix Table B4: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check with varying numbers of hypothetical boards over which the variance is computed. The dependent variable is the logarithm of the variance of difficulty.

	(1) 7 boards	(2) 6 boards	(3) 4 boards	(4) 3 boards
Remaining Time (in minutes)	0.0011***	0.0012***	0.0014***	0.0014***
	(0.0003)	(0.0004)	(0.0004)	(0.0005)
Relative Advantage	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0255***	0.0263***	0.0296***	0.0322***
	(0.0005)	(0.0005)	(0.0006)	(0.0007)
Log. Average Variance Difficulty of Alternatives	0.5577***			
	(0.0042)			
Log. Average Variance Difficulty of Alternatives		0.5439***		
		(0.0042)		
Log. Average Variance Difficulty of Alternatives			0.4943***	
			(0.0043)	
Log. Average Variance Difficulty of Alternatives				0.4430***
				(0.0044)
Interacted player-match-game FE	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes
Observations	162024	162024	162024	162024
R^2	0.266	0.251	0.202	0.160

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B5: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check with the inclusion of past risk. The dependent variable is the logarithm of the variance of difficulty.

	(1)	(2)	(3)	(4)	(5)
Remaining Time (in minutes)	0.0013***	0.0014***	0.0014***	0.0015***	0.0016***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Relative Advantage	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0261***	0.0259***	0.0255***	0.0253***	0.0257***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Log. Average Variance Difficulty of Alternatives	0.4989***	0.5019***	0.5153***	0.5244***	0.5238***
	(0.0042)	(0.0043)	(0.0044)	(0.0045)	(0.0046)
Log. Variance Difficulty in t-1	0.0722***	0.0703***	0.0719***	0.0667***	0.0665***
	(0.0027)	(0.0028)	(0.0028)	(0.0029)	(0.0029)
Log. Variance Difficulty in t-2		0.0155***	0.0120***	0.0130***	0.0109***
		(0.0026)	(0.0026)	(0.0027)	(0.0028)
Log. Variance Difficulty in t-3			-0.0118***	-0.0097***	-0.0098***
			(0.0025)	(0.0026)	(0.0026)
Log. Variance Difficulty in t-4				-0.0115***	-0.0093***
				(0.0026)	(0.0027)
Log. Variance Difficulty in t-5					-0.0157***
					(0.0027)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	156213	150402	144591	138780	132967
R^2	0.238	0.241	0.248	0.251	0.250

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. The additional covariates *Log. Variance Difficulty in t-1* until *Log. Variance Difficulty in t-5* refer to risk choices of previous half-moves of player *p*. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B6: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check of varying the functional form of covariates. The dependent variable is the logarithm of the variance of difficulty.

	(1)	(2)	(3)	(4)	(5)	(6)
Log. Average Variance	0.5273***	0.5819***	0.5237***	0.5239***	0.5056***	0.4920***
Difficulty of Alternatives	(0.0045)	(0.0040)	(0.0042)	(0.0042)	(0.0043)	(0.0042)
Relative Advantage	-0.0001*	-0.0000	-0.0000	-0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Remaining Time (in		0.0016***	0.0012***	0.0012***	0.0010***	0.0013***
minutes)		(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Log. Remaining Time	0.1331**					
	(0.0518)					
Position Difficulty	0.0280***		0.0276***	0.0272***	0.0091***	0.0074***
	(0.0006)		(0.0006)	(0.0006)	(0.0007)	(0.0006)
Log. Position Difficulty		0.2617***				
		(0.0088)				
Move Quality			0.0000			
			(0.0001)			
Log. Move Quality				0.0149***		
				(0.0024)		
Mean Difficulty					0.0296***	
					(0.0009)	
Log. Mean Difficulty						0.9495***
						(0.0130)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	162024	162024	162024	162024	162024	162024
R^2	0.228	0.218	0.230	0.231	0.240	0.263

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Log. Mean Difficulty* is the logarithm of the mean of difficulties of the hypothetical boards. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B7: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check with relaxed sample restrictions. The dependent variable is the logarithm of the variance of difficulty.

	(1) with players below 2400 Elo	(2) with obvious and forced moves	(3) with players below 2400 Elo and with obvious and forced moves
Remaining Time (in minutes)	0.0013***	0.0016***	0.0014***
	(0.0004)	(0.0004)	(0.0004)
Relative Advantage	-0.0000	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0283***	0.0275***	0.0275***
	(0.0005)	(0.0005)	(0.0005)
Log. Average Variance Difficulty of Alternatives	0.5231***	0.4852***	0.4858***
	(0.0040)	(0.0039)	(0.0038)
Interacted player-match-game FE	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes
Observations	177021	196617	201559
R^2	0.235	0.218	0.219

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B8: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check with further sample restrictions based on the Elo rating of players. The dependent variable is the logarithm of the variance of difficulty.

	(1) at least 2500 Elo	(2) at least 2600 Elo	(3) at least 2700 Elo	(4) at least 2750 Elo
Remaining Time (in minutes)	0.0012***	0.0016***	0.0019***	0.0022**
	(0.0004)	(0.0004)	(0.0006)	(0.0011)
Relative Advantage	-0.0000	-0.0000	0.0000	0.0001
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
Position Difficulty	0.0277***	0.0278***	0.0293***	0.0287***
	(0.0006)	(0.0006)	(0.0009)	(0.0013)
Log. Average Variance Difficulty of Alternatives	0.5247***	0.5230***	0.5152***	0.5092***
	(0.0043)	(0.0047)	(0.0067)	(0.0103)
Interacted player-match-game FE	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes
Observations	155504	129622	62766	25839
R^2	0.230	0.230	0.226	0.220

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix Table B9: IV estimates for the effect of remaining time on risk-taking, based on equation (1). Robustness Check of excluding specific playing modes. The dependent variable is the logarithm of the variance of difficulty.

	(1) without Fast Blitz games	(2) without Slow Blitz games	(3) without Rapid games	(4) without Classic games
Remaining Time (in minutes)	0.0012***	0.0011***	0.0013**	0.0020
	(0.0004)	(0.0004)	(0.0005)	(0.0031)
Relative Advantage	-0.0000	-0.0000	0.0000	-0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0276***	0.0277***	0.0280***	0.0267***
	(0.0006)	(0.0006)	(0.0007)	(0.0009)
Log. Average Variance Difficulty of Alternatives	0.5239***	0.5238***	0.5274***	0.5170***
	(0.0043)	(0.0044)	(0.0049)	(0.0068)
Interacted player-match-game FE	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes
Observations	155943	147118	117495	65516
R^2	0.231	0.231	0.229	0.231

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023 where one of the four playing modes is excluded in each specification. The dependent variable varies on the half-move-level. As explained in Section 2.3, it is computed as the logarithm of the variance over the difficulties of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3.5 Consequences of Risk-Taking

The findings presented in Section 5 raise the question if higher risk-taking, e.g. induced by time pressure or loss domains, also translates into consequences for the outcomes of chess games. We, therefore, investigate whether higher levels of average risk in a game correlate with the results of the games. In the first specification of Table B10, we consider a dummy indicating whether the game is decided (i.e. not drawn) as the dependent variable. This specification suggests that an increase in the average risk during a game is unrelated to the likelihood of a decided game. Further, the coefficient remains insignificant once we add further average game-level covariates to this specification which might influence the game's outcome. Specifications 3 and 4 are based on the subset of decided games in our dataset. We regress a dummy for winning the game (rather than losing) on the same game-level covariates as in the first two specifications. We find a positive link between average risk and the dummy for winning, implying that higher risk correlates with more wins than losses. Overall, Table B10 tentatively suggests that average risk-taking can also impact the outcomes of chess games, although this correlation is neither very robust nor is the causal direction clear. Nevertheless, we read this as provisional evidence that risk-taking can influence the outcome of a game which in turn implies that managing risk-taking in our setting is a meaningful strategic dimension.

Appendix Table B10: Correlational analysis on the player-game-level on the relation between risk and the result of the game. The dependent variable is a dummy indicating that the game is decided (win or loss) rather than drawn (Columns 1 and 2) or that the game is won, conditional on the game being decided (Columns 3 and 4).

	Sample: All games		Sample: Decided games	
	(1)	(2)	(3)	(4)
Average Risk	0.0290	-0.0035	0.1100***	0.1142***
	(0.0202)	(0.0192)	(0.0310)	(0.0308)
Average Risk of Alternatives	0.1942***	-0.0018	-0.3320***	-0.3098***
	(0.0222)	(0.0243)	(0.0374)	(0.0372)
Average Position Difficulty		0.0252***	0.0097***	0.0098***
		(0.0017)	(0.0025)	(0.0025)
Average Move Quality		0.0029***		-0.0016***
		(0.0004)		(0.0004)
Absolute Rating Difference		0.0005***		0.0004***
		(0.0001)		(0.0001)
Constant	-0.1173***	-0.0549*	1.0296***	0.9472***
	(0.0310)	(0.0319)	(0.0482)	(0.0501)
Observations	5811	5811	2834	2834
R^2	0.067	0.142	0.041	0.060

Note: Robust standard errors in parentheses. The sample consists of player-game combinations from World Cup games from 2013 – 2023. Columns 1 and 2 are based on the full sample whereas Columns 3 and 4 are restricted to player-game combinations of decided games, i.e. games that did not end in a draw.
* p < 0.10, ** p < 0.05, *** p < 0.01

A.3.6 Strategic Ambiguity

In this section, we discuss how strategic ambiguity — in addition to risk — attitudes can drive behavior of chess players.

conceptualization First, recall the example from Table 1. Strategy C is dominated by mixing between A and B for the row player: If the other player (column) plays X, strategy A obtains 25, B 14 and C 18 for the row player. If column plays Y, A results in 14, B in 25 and C 18 for row. In addition to classical risk preferences, ambiguity attitudes can drive behavior here, too: Consider the case of complete uncertainty about the opponent's strategy selection. With Gilboa and Schmeidler (1989) max-min preferences, the row player will consider the worst-possible outcomes (Calford, 2020). As $\min_C > \min_A = \min_B$, strategic ambiguity aversion can lead the row player to opt for C. An ambiguity-averse player will thus seek to maximize the minimum of the consequences of his available strategies.

Measuring ambiguity attitudes in chess Again, recall the example in Figure 1: Black played 21... b7-b6 while 21... f5-f4 would have been the optimal move. We discussed in the main text that the variance of the difficulty is larger for 21... f5-f4 compared to 21... b7-b6 corresponding to a higher degree of strategic risk. For difficulty — the analogy to payoffs — instead of maximizing the minimum, a player will seek to *minimize* the maximum, as player generally will try to avoid high-difficulty positions. Considering the maximum of difficulty (instead of its variance) across the hypothetical boards leads to a similar conclusion as in the main text: After 21... b7-b6, the maximum difficulty is 49 CP, yet 62 CP after 21... f5-f4. Thus, in addition to minimizing the variability of the position evaluation and difficulty, 21... b7-b6 also ameliorates the worst-off position for Black compared to 21... f5-f4. Note that we abstain from computing the minimum or maximum of the relative (dis-)advantage. Consistent with Sections 2.2 and A.1, the minimum of the relative (dis-)advantage refers to the first best response to a given move. In contrast, its maximum is based on the worst move under consideration.

The effect of thinking time on ambiguity attitudes To investigate the effect of remaining thinking time, we repeat the main analysis with the maximum difficulty data in Table B11. We again use logs to facilitate interpretation. Panel A of Table B11 considers the main specification with five different outcome variables, varying the number of hypothetical boards between 3 and 7. We do not find an effect of remaining time on either of these dependent variables, suggesting that the logarithm of the maximum of difficulty is insensitive to time pressure. In Panel B, we employ the logarithm of the minimum of difficulty as outcome variable. Complementing the analysis in Panel A, this allows to disentangle channels of the effect of remaining time on the variance of difficulty. Specifications 1 to 5 show uniformly that remaining time has a negative and significant coefficient. This implies that players increase the minimum of difficulty once remaining thinking time decreases.

Taken together, the effects of remaining time on minimum and maximum are largely consistent with the positive effect on the variance in difficulty. Put differently, a reduction in remaining time leads to an increase in the minimum of difficulty distribution. This further condenses the distribution so that the variance reduces accordingly.

Appendix Table B11: IV estimates for the effect of remaining time on strategic ambiguity. The dependent variables are the logarithm of the maximum of difficulty (Panel A) and the logarithm of the minimum of difficulty (Panel B).

Panel A: logarithm of the maximum of difficulty					
	(1) 7 boards	(2) 6 boards	(3) 5 boards	(4) 4 boards	(5) 3 boards
Remaining Time (in minutes)	0.0002	0.0001	0.0001	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Relative Advantage	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0112***	0.0113***	0.0113***	0.0115***	0.0119***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Log. Average Maximum Difficulty of Alternatives (7 boards)	0.4592*** (0.0056)				
Log. Average Maximum Difficulty of Alternatives (6 boards)		0.4550*** (0.0055)			
Log. Average Maximum Difficulty of Alternatives (5 boards)			0.4547*** (0.0055)		
Log. Average Maximum Difficulty of Alternatives (4 boards)				0.4505*** (0.0056)	
Log. Average Maximum Difficulty of Alternatives (3 boards)					0.4294*** (0.0058)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	162024	162024	162024	162024	162024
R^2	0.636	0.647	0.657	0.668	0.674

Panel B: logarithm of the minimum of difficulty					
	(1) 7 boards	(2) 6 boards	(3) 5 boards	(4) 4 boards	(5) 3 boards
Remaining Time (in minutes)	-0.0006***	-0.0006***	-0.0005***	-0.0005***	-0.0004**
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Relative Advantage	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Position Difficulty	0.0061***	0.0065***	0.0069***	0.0074***	0.0079***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Log. Average Minimum Difficulty of Alternatives (7 boards)	0.5473*** (0.0046)				
Log. Average Minimum Difficulty of Alternatives (6 boards)		0.5339*** (0.0047)			
Log. Average Minimum Difficulty of Alternatives (5 boards)			0.5216*** (0.0048)		
Log. Average Minimum Difficulty of Alternatives (4 boards)				0.5082*** (0.0048)	
Log. Average Minimum Difficulty of Alternatives (3 boards)					0.4901*** (0.0050)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	162024	162024	162024	162024	162024
R^2	0.648	0.644	0.641	0.634	0.620

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level. It is computed as the logarithm of the maximum (Panel A) or the minimum (Panel B) of difficulty of the hypothetical boards that can arise after the played move. *Remaining Time* is instrumented with the interaction of half-move dummies and initial thinking time of the specific time control. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Strategic loss aversion for ambiguity attitudes Adding to the analysis from Section 5.2 where we consider strategic loss aversion, we reiterate this analysis for the maximum of difficulty. Hence, Table B12 applies the same specification as Table 9 but uses the logarithm of the maximum of difficulty as new outcome variable. It reports the effect of a blunder in the previous move (Panel A) and of relative advantage (Panel B) on the dependent variable. Panel A documents positive and significant coefficients for three different definitions of blunders, indicating that players are more willing to be confronted with a difficult position in the next move when their last move was a blunder. In Panel B, we analyze whether the relative advantage correlates with the maximum of difficulty that players are willing to accept in the subsequent move. We find negative correlations, suggesting that players are more prone to increase the maximum of difficulty if their disadvantage increases. Equivalent to Section 5.2, we find that negative deviations from the reference point increase the maximum of difficulty a player is willing to accept, yielding further supportive evidence of strategic loss aversion.

Appendix Table B12: Fixed effects estimates for the effect of making a blunder in the previous move (Panel A) and for the effect of relative advantage (Panel B). The dependent variable is the logarithm of the maximum of difficulty.

Panel A: effect of a blunder in the previous move			
	(1)	(2)	(3)
Blunder in Previous Move (Threshold 50 CP)	0.0423***		
	(0.0025)		
Blunder in Previous Move (Threshold 100 CP)		0.0502***	
		(0.0043)	
Move Quality of Previous Move			0.0001***
			(0.0000)
Log. Average Maximum Difficulty of Alternatives (5 boards)	0.4659***	0.4670***	0.4674***
	(0.0055)	(0.0055)	(0.0055)
Position Difficulty	0.0109***	0.0109***	0.0110***
	(0.0002)	(0.0002)	(0.0002)
Remaining Time (in minutes)	-0.0002***	-0.0002***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)
Interacted player-match-game FE	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes
Observations	156260	156260	156260
R^2	0.735	0.735	0.735

Panel B: effect of relative advantage					
	(1) All	(2) at most 500 CP advantage by either side	(3) at most 400 CP advantage by either side	(4) at most 300 CP advantage by either side	(5) at most 200 CP advantage by either side
Relative Advantage	-0.0000***	-0.0000***	-0.0000***	-0.0000***	-0.0000*
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Log. Average Maximum Difficulty of Alternatives (5 boards)	0.4543***	0.4401***	0.4347***	0.4279***	0.4194***
	(0.0056)	(0.0057)	(0.0058)	(0.0059)	(0.0060)
Position Difficulty	0.0113***	0.0120***	0.0122***	0.0124***	0.0127***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Remaining Time (in minutes)	-0.0002***	-0.0002***	-0.0003***	-0.0003***	-0.0003***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	162024	157427	155016	151370	145409
R^2	0.721	0.711	0.706	0.699	0.688

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013 – 2023. The dependent variable varies on the half-move-level, it is computed as the logarithm of the maximum of the difficulties of the hypothetical boards that can arise after the played move. In Panel A, we define a blunder as a move with at least 50 CP loss (specification (1)), at least 100 CP loss (specification (2)) and as a continuous variable (specification (3)). In Panel B, we consider the whole sample (specification (1)) as well as subsets of positions with a relative advantage of at most 500, 400, 300 or 200 CP by either side (specifications (2) - (5)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$