Journal of Economic Behavior and Organization 238 (2025) 107218

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Research paper

Time pressure and strategic risk-taking in professional chess[☆]

Johannes Carow[a,b], Niklas M. Witzig[a,*]

[a] JGU Mainz, Germany
[b] MWVLW RLP, Germany

ARTICLE INFO

JEL classification:
C26
C45
D91

Keywords:
Chess
Risk
Time pressure
Loss aversion
Machine learning

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 variation in thinking time. Our results indicate that having less thinking time consistently leads chess players to opt for more risk-averse moves. This effect is particularly pronounced in disadvantageous positions, that is, in which a player is trailing behind. We additionally provide correlational evidence for strategic loss aversion, a tendency for more risky moves after a mistake and in a disadvantageous position. Our results suggest that even high-proficiency decision-makers in high-stake situations react to time pressure and contextual factors more broadly.

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.

A common feature of this work is their reliance on samples from *experimental laboratories*. In typical lab studies, neither the proficiency of decision-makers nor the stakes of the decisions are very high. Whether the above findings translate into high-stake settings with highly proficient and experienced decision-makers who regularly decide under time pressure thus remains largely 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

☆ We thank Raphael Brade, Julia Engel, Katharina Hartinger, Greg Howard, Thorsten Schank, Daniel Schunk, Christian Seel and two anonymous referees 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 (https://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.

* Corresponding author.
E-mail addresses: johannes.carow@uni-mainz.de (J. Carow), niklas.witzig@uni-mainz.de (N.M. Witzig).

https://doi.org/10.1016/j.jebo.2025.107218
Received 2 September 2024; Received in revised form 16 August 2025; Accepted 24 August 2025
Available online 17 September 2025
0167-2681/

by turning to an ideal setting for our research question: professional chess. Chess is often credited as *the* strategic game and chess players are regularly acclaimed for a remarkable level of strategic skill and expertise. Chess requires highly strategic considerations by both players, selecting between 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 part of chess: Chess is played with different time limits, e.g., 90 min for each player (classic) or 5 min (blitz) and managing efficient time-usage represents a key component of chess mastery (Sigman et al., 2010). In turn, 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.[1] The World Cup assembles the strongest chess players in the world and, next to offering generous monetary prizes, 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 *knockout*-tournaments, where two players in each match determine the winner processing to the next round, which further increases the stakes of each match.

Crucial for our analysis, the World Cup format introduces systematic variation in thinking time through its different time controls: In case of a tie, players play additional games against each other *in a tighter time control*, i.e., with a reduced initial time budget. This structure induces variation in the *time paths* of available thinking time across games: players (need to) play faster in tighter time controls, resulting in consistently lower remaining time at each stage of the game. To illustrate, the average remaining time at half-move $h = 7$ differs considerably between Classic and Blitz games, and so does the decline in remaining time over the course of the game for example, from $h = 7$ to $h = 23$. These playing-mode-specific differences in both the level and the slope of remaining time serve as *identifying variation* in our IV analysis. Therefore, we instrument remaining time with the initial time of each time control (similar to Howard 2024), interacted with half-move dummies,[2] to use the different paths of time consumption from each playing mode as identifying variation.

Concerns about the exogeneity of this IV approach naturally arise, as the occurrence of a faster time control depends on whether a previous match reached a tiebreak. If players adjust their risk-taking in ways that affect the likelihood of reaching a tiebreak – or if inherently more risk-seeking or risk-averse players are systematically more likely to reach tiebreaks – then playing modes become endogenous to risk-taking.

We address these concerns in several ways. First, we include interacted player-match-game fixed effects, following Zegners et al. (2020), which absorb all time-invariant player characteristics and game-level confounders. This ensures that we identify effects within player-match-game-units, thereby mitigating concerns that players' general preferences for faster time controls or strategic considerations bias our estimates. Second, we show that observable pre-tiebreak characteristics do not predict the likelihood of entering a tiebreak (see Table 2). Third, we replicate our analysis in the Biel Grandmaster Tournaments, an independent tournament with a different format, where time control changes are unrelated to prior game outcomes. The results from this replication are highly consistent with our main findings, supporting the validity of our identification strategy (Section A.6).

From our main analysis, our central finding is that time pressure leads to a systematically different selection towards safer strategies, i.e., that chess players behave more risk-averse under time pressure. Subsample analyses show that this effect is particularly pronounced in disadvantageous positions, robust to different subsamples based on the strength of players or the initial time budget and also similar for different time-control specific differences in the rating of players.

Furthermore, the increase in risk-taking due to more available thinking time is particularly pronounced in games in which a player has to win to stay in the competition. In addition to our main effect, we find correlational evidence for strategic loss aversion, indicated by a higher tendency to play risky moves after a mistake and in disadvantageous positions.

This leads to the main contribution of this paper: Our results highlight that time pressure and loss frames can shift behavior among highly proficient decision-makers in high-stakes environments, with loss frames potentially moderating the effect of time pressure. These findings reinforce the previously mentioned notion that time pressure affects strategic decision-making, in fact in similar ways for highly proficient decision-makers as for more lay samples. While we discuss the particularities of chess as a setting and how to interpret the origins of this effect, our results suggest that time pressure generally pushes strategic decision-making toward safer choices. To the best of our knowledge, this paper is the first to document such effects in a sample characterized by both high stakes and high expertise.

With this paper, we connect to several strands of literature: First, to an economic literature that uses chess settings to study risk-taking: 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 a classification by a survey of 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"). 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

[1] For the World Cups prior to 2013, no clock time data is available, rendering these tournaments unsuitable for our analysis.

[2] The concept of half-moves is explained in Sections 2 and 4.

risk measures vary per *game*, not on the *move-by-move-level*. Consequently, the scope of previous analyses 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).

In contrast, our move-by-move measure 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. Importantly, our measure is nonetheless consistent with the previous classification from Dreber et al. (2013) and 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. 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. This puts a strong emphasis on how humans play chess instead of solely relying on chess engines or expert opinions. 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 as we do, 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, we also relate to a larger literature interested in the impact of time pressure on *nonstrategic* risk-taking, i.e. where no other player is involved. This literature produced mixed evidence so far: 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 thus can have an impact on (non-strategic) risk-taking, albeit less consistently compared to the strategic domains mentioned in the beginning. We add to this literature in a broader sense by providing evidence on the impact of time pressure for risk-taking of professional and highly proficient decision makers, something this literature has not investigated yet.

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. Avoyan et al. (2020) consider stopping behavior in online chess and, accordingly, derive different types of players. They also derive a theoretical framework where welfare can be enhanced by improved matching procedures. Finally, Linnemer and Visser (2016) study the self-selection of chess players into tournament sections based on their strength. We show how chess can offer insights into how highly strategic decisions are impacted by having less time to think before making them.

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. Strategic risk in chess as the variance of outcomes

As described, we require a measure of the degree of strategic risk entailed *in a single chess move*. We first lay out the intuition behind our measure before turning to more details on the empirical quantities necessary for its construction.

For illustration, consider the upper panel Fig. 1. Here, we depict a situation (i.e. board position) of the game Brkic - Barrientos from the World Cup 2021. Black just decided to play 21... b7-b6 highlighted in green (after White's 21st move 21. a5-a6). R1-R5 constitute the five best responses (more details on their identification follow) by White to Black's move and each response in turn leads to a novel position for Black, as shown in the lower panel of the Figure (R1-R5 highlighted in blue). To assess the risk of Black's move, we use data on the *variance* of outcomes – e.g., the position evaluation – of each candidate position after White's responses. If White's responses (do not) strongly weaken or strengthen Black's position depending on which move White will actually carry out, 21... b7-b6 can be considered a (safe) risky move. We thus interpret the selection of moves with a lower variance of subsequent

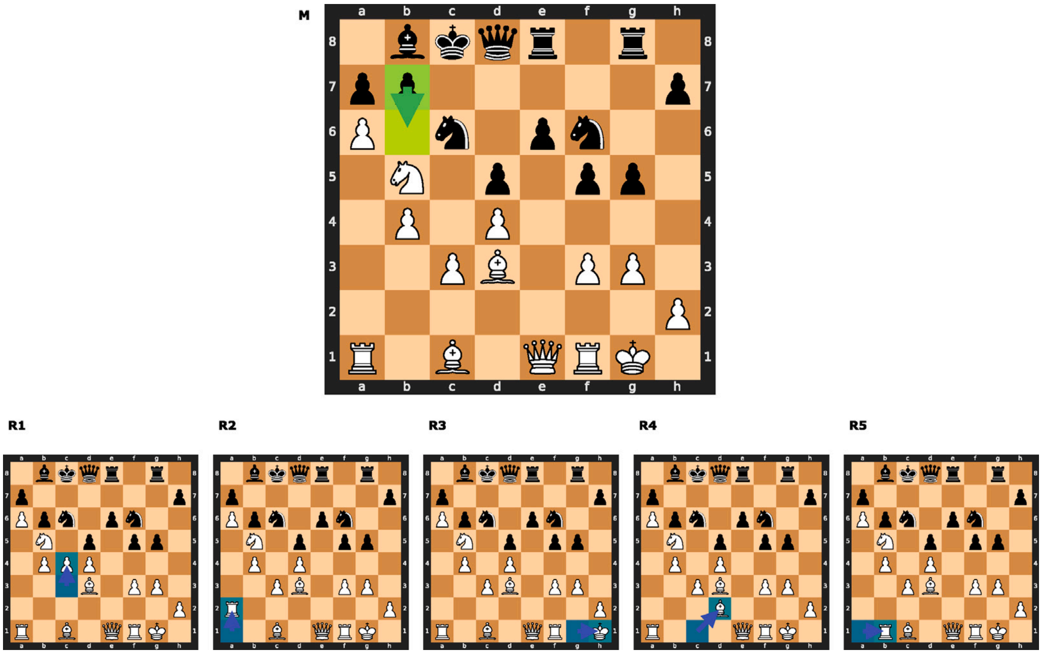

Fig. 1. Trading-off Risky and Safe Moves: Illustrative Example for the Computation of the Risk Measure. Brkic vs Barrientos, FIDE World Cup 2021, position after White's move 21. a5-a6.

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. Please note that the time control - key source of identifying variation later on - is not yet relevant in this example.

outcomes as (revealed) risk-aversion.[3] This matches intuitions linking risk aversion to the variance of wealth in expected utility calculations and has also been applied to chess moves before by Holdaway and Vul (2021). This further assumes that when selecting a move out of several possible options, a player *anticipates* the most likely responses by their opponent to *each* candidate move and afterwards assesses which continuation is most favorable to play.[4]

This explicitly 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.

2.1. *Consideration sets and relevant outcomes*

In turn, empirically measuring the risk of a chess move requires (i) consideration sets both of alternatives to a given move and, for each move, likely responses by the opponent and (ii) empirical quantities of relevant outcomes players consider with which the variance can be computed. Given the high proficiency of our sample, we are able to make educated guesses about (i) and approximate (ii) with two candidate quantities. We outline the necessary details here and provide a more in-depth example how to compute our risk measure for 21... b7-b6 from Fig. 1 which relies more strongly on chess notation in Section A.2.

[3] A theoretically more involved foundation for this argument can be found in Calford (2020). There, he further discusses how *ambiguity*-aversion might drive choices, too. We present a more extensive explanation of that argument in Appendix Section A.5.6. There, we show how to capture ambiguity attitudes in our setting and repeat the empirical analyses and largely confirm the main findings based on our risk measure.

[4] Of course, other motivations beyond risk preferences could determine the move selection of players. We discuss alternative interpretations of our main finding in detail in Section 6.

2.1.1. Consideration sets

To gather data on both considered moves and anticipated responses in a given position, we make use of *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 and use Stockfish 11, the most widely used open-source chess engine. It may appear counter-intuitive to use an engine to identify consideration sets, given that we are deal with human players which 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,[5] 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. We thus use a restricted Stockfish to both gather data on both candidate moves in a position (such as 21...b7-b6) and the opponent's 5 best responses (as R1-R5 in Fig. 1).

While we cannot know the exact half-moves a player considered and anticipated, examining the five best moves and responses (according to the restricted engine) appears to be a reasonable range from which grandmasters pick.[6] 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. 1 CP corresponds to one *centipawn (CP)*, which measures how many hundredths of a pawn (i.e. the weakest piece in chess) once side will be at a (dis)advantage if both sides play optimally from this position onwards for a given length of the continuation.[7] We also repeat our analyses varying the number of the opponent's responses under consideration to be between 3–7.

2.1.2. Relevant outcomes in chess

For relevant outcomes, we consider *position evaluations*, 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 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 again available via ()restricted chess engines. Similar to specific moves, chess engines can evaluate positions and measure the (dis)advantage of one side in terms of CP as well. 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. Consider the initial position in Fig. 1, to which Stockfish 11 assigns an advantage of 77 CP for White. This corresponds to a moderate, albeit not decisive advantage. Overall, we are thus able 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

[5] 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.

[6] Table 1 shows that in the Classic mode of the FIDE World Cup sample, 51 percent of half-moves in the sample are the first best half-moves while 96 percent of all moves are among the five best half-moves.

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

[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}.

Table 1
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)	6.47	2.35	6.57	2.33	6.58	2.32	6.47	2.38
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 A.2, 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 2.1.2 and A.1 for further explanations). The variables *First Best Move Played* until *Seventh Best Move Played* are dummy variables.

objective) quality,[10] 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 of b7-b6 in Fig. 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.3. Ultimately, our neural network predicts the

[10] Importantly, to evaluate the objective move quality in the training dataset, we do not restrict stockfish to depth 15 as we did for identifying the considerations sets of human grandmasters.

Table 2
Selection 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.

quality of the human move (measured in CP) if a proficient human player plays that position. As the optimal move – by construction – is evaluated at 0 CP, the predicted human move quality is equivalent to the deviation of this move to the optimal move and thus a measure of the position difficulty in a particular chess position.[11] The difficulty of the initial position in Fig. 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, chess intuition might advice to maintain king safety by defending with 21… b7-b6. In contrast, f5-f4, the optimal move, might appear less natural as it allows a further weakening of Black's king and hence, might seldom be played. For further examples of low, medium and high difficulty, see Figure A6.

Our difficulty measure, though constructed with the aid of chess engines, fundamentally reflects a *human perspective* on playing chess. This sets it apart from existing, purely engine-based approaches. For instance, Guid and Bratko (2006) assess the difficulty of a chess position based on the time an engine requires to find the optimal move, while Sunde et al. (2022) use the number of nodes an engine considers to find the best move as a 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. Not only does this explicitly introduce a human dimension in the measurement of the difficulty, but it also comes with methodological benefits: Both the measures by Guid and Bratko (2006) and Sunde et al. (2022) require a considerable amount of computational effort, as each position must be evaluated individually. By contrast, once our supervised learning model is trained, it can predict the difficulty of a vast number of chess positions almost instantly.

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 A.2. 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 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 variation in thinking time which is key for our identification strategy. Specifically, we leverage the fact that players compete in different time controls, a feature used to instrument the remaining thinking time in chess (see e.g. Howard (2024)).

[11] An interactive website of our model is available at https://nmwitzig.github.io/chess-app.html. First, it includes several example positions for low, medium and high difficulty. Second, it allows to check the difficulty for given chess positions by pasting a FEN string to the interactive website. To obtain a FEN string, one can set up a position on a website such as https://lichess.org/analysis. The FEN string from lichess can then be copied to our interactive website.

We consider the FIDE World Cups 2013–2023.[12] They take place biannually so the sample consists of six World Cups (2013, 2015, 2017, 2019, 2021, 2023). Figure A10 visualizes the structure of one World Cup while Table A5 presents the corresponding *playing modes*. In each World Cup, there are 128 participants.[13] 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 min for the game while they get an additional 30 s increment for each move played. Furthermore, they receive 30 min 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 min and 10 s increment per move are played. If the match is still tied, two games with 10 min and 10 s (*Slow Blitz*) constitute the next part of the tiebreak. In case of a further tie, two games with 5 min and 3 s 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 min while the player leading the Black pieces obtains only 4 min. 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.[14] 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.

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).[15] 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.[16] 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 1 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 A.2 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 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 A.2, 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.

[12] For the World Cups prior to 2013, data on the played moves is available but time consumption was not tracked.

[13] 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.

[14] 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.

[15] In January 2024, the website *chess24.com* was closed down. The downloaded game files are available upon request.

[16] We focus on the first 80 half-moves due to the additional time bonus players receive at half-move 80 in classic games.

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 2.1.2. 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 17.79 and 19.21 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 17 and 20 centipawns if a human player plays these positions. Further, we consider 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. This measure is lowest in classic games, albeit similar between playing modes.

4. Econometric specification

We aim to analyze the effect of remaining thinking time on revealed risk-taking of players. As discussed in Section A.2, 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:

$$
\begin{gathered}
log(Risk_{pmgh}) = \alpha \cdot RemainingTime_{pmgh} + \sum_{k=1}^{80} \phi_k \cdot DHalf-Move_{kh} + X_{pmgh}\gamma + \mu_{pmg} + u_{pmgh} \\
p = 1, \dots, 419\ players,\ m = 1, \dots, 908\ matches,\ g = 1, \dots, 9\ games\ and\ h = 1, \dots, 80\ half-moves
\end{gathered} \tag{1}
$$

where $Risk_{pmgh}$ the risk of half-move h played by player p in game g and match m, $RemainingTime_{pmgh}$ refers to available thinking time prior to half move h, $DHalf-Move_{kh}$ denotes dummies for half-moves, the vector X_{pmgh} includes control variables (relative advantage and difficulty prior to the half-move, logarithm of average risk of alternative half-moves) and μ_{pmg} captures the interacted player-match-game fixed-effects. The fixed-effects account for confounding factors on the more aggregate game-level such as rating difference between the players or the choice of the opening as well as confounders across games such as the results from previous games.

However, Eq. (1) suffers from an endogeneity concern: The time remaining on the clock to make a decision, $RemainingTime_{pmgh}$, is a function of past time consumption in the game. Past time consumption, however, might have depended on past positions on the board. When facing simple decisions with low strategic considerations and a very limited choice set, players might play faster than in complicated positions with tough decision problems which move to play. The dependent variable in Eq. (1) $Risk_{pmgh}$ depends on players choices, specifically the risk associated with the played move as explained in Section 2. Hence, $RemainingTime_{pmgh}$ might be correlated with past values of $Risk_{pmgh}$. Further, $Risk_{pmgh}$ can be positively serially correlated within games (e.g. due to a risky opening choice and resulting risky positions in the middlegame), creating reversed causality between $Risk_{pmgh}$ and $RemainingTime_{pmgh}$ in Eq. (1). Hence, OLS estimation of Eq. (1) might be biased.

Therefore, we aim to identify the effect of remaining time using an instrumental variable approach that exploits variation across different *playing modes* with exogenously determined initial time budgets. More specifically, the playing-mode-specific patterns of time consumption provide the identifying variation, as illustrated in Fig. 2. We observe substantial differences in time paths across playing modes. For example, the reduction in remaining time between moves 7 and 23 differs markedly between classic and blitz games. Precisely these differences in time consumption form the core of our identifying variation, which we use to explain differences in risk-taking behavior between moves 7 and 23 in classic versus fast blitz games. Fig. 2 shows a decline in remaining time from 85 min at move 7 to 30 min at move 23 in classical games, but only a much smaller decline in fast blitz games (from 3 to 2 min). Our identification strategy compares how player p's risk choice in move 7 vs. 23 changes in a fast blitz games as compared to their risk choice in move 7 vs. 23 in a classic game.

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

$$
\begin{gathered}
RemainingTime_{pmgh} = Initial\_Time_g \cdot (\sum_{k=1}^{80} \psi_k \cdot DHalf-Move_{kh}) + \sum_{k=1}^{80} \eta_k \cdot DHalf-Move_{kh} + X_{pmgh}\kappa + \lambda_{pmg} + e_{pmgh} \\
p = 1, \dots, 419\ players,\ m = 1, \dots, 908\ matches,\ g = 1, \dots, 9\ games\ and\ h = 1, \dots, 80\ half-moves \\
(Classic,\ Rapid,\ Slow\ Blitz,\ Fast\ Blitz) \in Initial\_Time_g
\end{gathered} \tag{2}
$$

This first stage estimation allows to extract the variation in $RemainingTime_{pmgh}$ which we use for identification. As the player's time budget might have evolved depending on the past board situation, there might be variation in remaining time correlated with the error term of Eq. (1). Therefore, the first stage strives for isolating the exogenous component of $RemainingTime_{pmgh}$. We do so by instrumenting $RemainingTime_{pmgh}$ with the average paths of time consumption, as given by the interactions of Initial time and the half move dummies. To interpret the coefficients ψ_k of these interactions, one has to keep in mind that the coefficients of half move dummies are interpreted relative to the reference category. Hence, we allow time consumption between the reference (i.e. the

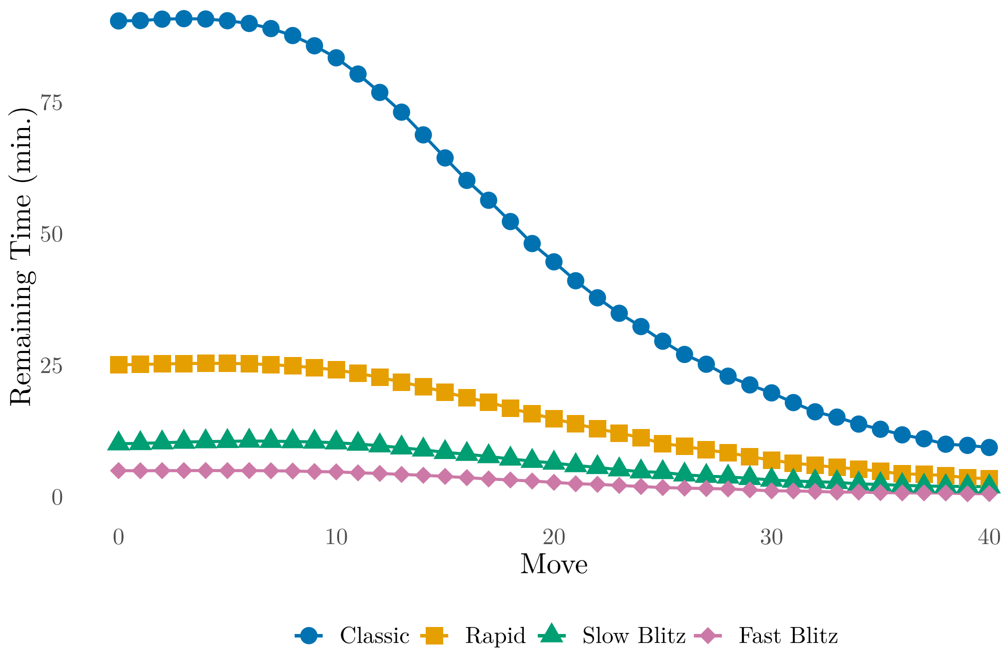

Fig. 2. Evolution of Remaining Time within Games for Different Playing Modes.
Note: This graph shows the paths of average remaining time within games of the four playing modes *Classic*, *Rapid*, *Slow Blitz* and *Fast Blitz*. The sample consists of World Cup games from 2013–2023.

first half-move of each player, i.e. initial time) and the current move to vary by $initial_time$ (thus interacting with initial time). More information how the estimates are calculated and interpreted is provided by means of a simplified example in Section A.7.

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 A9 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. Additionally, the fixed effects ensure that match dynamics do not impair our estimates. As the fixed effects capture heterogeneity across player-match-game units, identification is based on the variation within a game of a specific match for one of the players. Hence, fixed effects are crucial for identification as they rule out several potential confounders.

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, endogenous sorting (e.g., a fixed player preference for playing in a particular playing mode) is partially remedied by the fixed effects as well.

While the choice of opening is accounted for by the fixed effect, players may still adopt a more cautious strategy throughout a game to increase the likelihood of a tiebreak. If this occurs, the exclusion restriction would be violated. We therefore further examine potential selection into tiebreaks using Table 2, which analyzes the determinants of the probability that a match is not decided in the *Classic* games but extends at least to the $Rapid$ tiebreak. If the occurrence of tiebreaks is linked to risk-taking behavior, we should observe corresponding effects in this analysis. We run the 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 Eq. (1)) as the explanatory variable of interest. The further independent variables represent the match-level averages of X_{pmgh} from Eq. (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. Moreover, 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 Eq. (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 Eq. (1) does not depend on the actual mode or on initial time, respectively. Hence, this finding supports the exclusion restriction. In Table A7, 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 Table 2 and A7 for the subsample of half-moves 21-80 and obtain similar results.

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.[17]

5. Results

5.1. The effect of time pressure on strategic risk

This section analyzes the causal effect of time pressure on strategic risk. Section 5.1.1 focuses on the variance of difficulty as our risk measure of main interest (see Section A.2 for more details). In Section 5.1.2, we use the variance of relative (dis)advantage as an alternative measure.

5.1.1. Main measure: Variance of difficulty

Table 3 shows the baseline estimates for OLS, fixed effects and instrumental variable regressions.[18] 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

[17] Table A4 in Section A.4.2 further shows that the risk-level of the opening strategies, measured by the classification of Dreber et al. (2013), is very similar across playing modes.

[18] Table A6 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 A6 also contains the corresponding F-statistic which is larger than 500.

Table 3
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	162 024	162 024	162 024	162 024
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 A.2, 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 A6. Additional specifications are shown in Table A8.
* p < 0.10, ** p < 0.05, *** p < 0.01.

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 min for 40 moves, the average change in remaining time per move amounts to 2.25 min. 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 generally spend more time when making risk-seeking decisions, a negative link between $RemainingTime_{pmgh}$ and $Risk_{pmgh}$ for all half-moves $a < h$ of player p within game g of match m emerges. If, in addition, $Risk_{pmgh}$ is positively serially correlated within games, the OLS coefficient of $RemainingTime_{pmgh}$ on $Risk_{pmgh}$ will be downward biased if the endogeneity of $RemainingTime_{pmgh}$ is not accounted for. The IV approach addresses the confounding of risk and remaining time in the current move by using only the exogenous variation in remaining time – that is, the variation given by the average paths of time consumption. It thereby diminishes the contemporaneous confounding and, thus, mitigates the negative correlation, i.e., regression coefficient.[19]

Building on the previous analysis, Table 4 stratifies the estimation by position evaluation. Hence, it distinguishes scenarios where the player to move has (1) an advantageous position (advantage of at least 1 CP),[20] (2) a disadvantageous position or (3) a balanced position. We find that only the subsample of disadvantageous positions features a positive significant coefficient. Hence, the overall positive coefficient from column (4) in Table 3 is mainly driven by choices of players in disadvantageous situations. In these positions, players try to reduce their disadvantage to strive for a balanced position. Table 4 illustrates that time pressure induces players to opt for less risky options once they are in an inferior position. This is in line with the high proficiency of players - if in disadvantageous positions, they tend to further reduce the variance of future outcomes once remaining time becomes scarce. As stronger players tend to consume time when calculation is beneficial (Russek et al., 2022), they take into account when they get into time trouble. Hence, they reduce the variance of difficulty when they anticipate time pressure. Behaving this way ensures that future difficulty does not depend strongly on the opponent's move. In contrast, players with a better position, know about the strong defensive skills of their opponent. Once time pressure approaches, they feel no need to avoid risk as some risk might be needed to change the character of the position, putting pressure on the opponent. From a better position, there is often more discretion to choose from multiple moves. This is in line with the zero-effect in Column (1) of Table 4, as players in a better position feel less need to react to upcoming time pressure.

To demonstrate the robustness of the positive effect of remaining time on risk, Table 5 considers different subsamples. In specification (1), we restrict to observations with a remaining time budget of at least 30 s. This rules out that the previously observed

[19] Further note that the pattern of sign reversal also replicates in the data of the Biel tournament (Table A22).
[20] When applying another threshold for an advantage such as 30 CP, results remain robust.

Table 4
IV estimates for the effect of remaining time on risk-taking, based on equation (1). Stratification by position evaluation. The dependent variable is the logarithm of the variance of difficulty.

	(1) advantageous positions	(2) disadvantageous positions	(3) balanced positions
	0.0007	0.0017***	−0.0002
Remaining Time (in minutes)	(0.0006)	(0.0007)	(0.0016)
Relative Advantage	0.0001***	−0.0002***	
	(0.0000)	(0.0001)	
Position Difficulty	0.0313***	0.0223***	0.0318***
	(0.0009)	(0.0008)	(0.0022)
Log. Average Variance Difficulty of Alternatives	0.4955***	0.5071***	0.4826***
	(0.0061)	(0.0068)	(0.0142)
Interacted player-match-game FE	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes
Observations	78 859	65 099	16 599
R^2	0.207	0.212	0.197

Note: Standard errors clustered on the game-level in parentheses. The sample consists of World Cup games from 2013–2023, stratified by the opponent's remaining time. Column (1) includes observations where the player to move has an advantageous position while columns (2) and (3) consider disadvantageous and balanced positions, respectively. (Dis)advantageous positions are positions with at least 1 CP (dis)advantage. The dependent variable varies on the half-move-level. As explained in Section A.2, 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 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. 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 6, 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 by weighting the difficulties of the five hypothetical boards with their respective engine evaluations.[22] 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.

In Section A.5.5, we present a large number of additional robustness checks regarding our main effect of interest. Specifically, we vary the number of hypothetical boards considered (Table A9), control for different values of past move risk (Table A10), and implement alternative functional forms of the explanatory variables (Table A11). We also add move quality and the mean difficulty of the hypothetical boards. The main effect of interest remains robust across analyses. Similarly, our main effect of interest is *identical* in a sub-sample analysis that distinguished whether a player made the best move or not (Table A17).

[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.

[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
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.

Table 5
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	160 181	108 520	19 491	80 925	81 099
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 A.2, 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6
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	162 024	162 024
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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A12 relaxes the baseline sample restrictions by including players with less than 2400 Elo and positions with an obvious or forced move. Table A13 further restricts the sample to players with at least 2500, 2600, 2700, or 2750 Elo. Again, across analyses, results remain unchanged. Table A14 excludes each playing mode in turn and dropping *Fast Blitz*, *Slow Blitz*, or *Rapid* does not affect the results. Excluding *Classic* leads to a larger, less precise coefficient, likely due to Classic games' high share in the sample and the resulting drop in variation of remaining time.

Table A15 addresses concerns that players might act strategically based on their opponent's time-control-specific playing strength. Using monthly ELO ratings by time control, we compare sub-samples where the opponent's classic rating is higher vs. those where their rapid/blitz rating is higher. Results remain robust. Finally, Table A16 stratifies by opponent's remaining time: (i) more, (ii) less, or (iii) similar time compared to the player to move. We find stable effects across all groups, indicating opponent time endowment plays no major role.

In addition to using data from the FIDE Ches World Cups 2013–2023, we apply our "analytical pipeline" to matches from the Biel Grandmaster Triathlons 2019–2024, where players know in advance that they will face each opponent in all three playing modes.

Table 7
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)	0.0019*** (0.0007)	0.0020*** (0.0007)	0.0021*** (0.0007)	0.0024*** (0.0008)	0.0026*** (0.0008)
Relative Advantage	0.0005*** (0.0000)	0.0005*** (0.0000)	0.0005*** (0.0000)	0.0004*** (0.0000)	0.0004*** (0.0000)
Position Difficulty	0.0316*** (0.0008)	0.0323*** (0.0008)	0.0333*** (0.0008)	0.0338*** (0.0009)	0.0341*** (0.0009)
Log. Average Variance Relative Advantage of Alternatives	0.2943*** (0.0039)				
Log. Average Variance Relative Advantage of Alternatives		0.2907*** (0.0038)			
Log. Average Variance Relative Advantage of Alternatives			0.2858*** (0.0038)		
Log. Average Variance Relative Advantage of Alternatives				0.2865*** (0.0037)	
Log. Average Variance Relative Advantage of Alternatives					0.2747*** (0.0036)
Interacted player-match-game FE	Yes	Yes	Yes	Yes	Yes
Half-Move Dummies	Yes	Yes	Yes	Yes	Yes
Observations	159 629	159 629	159 629	159 629	159 629
R^2	0.169	0.164	0.158	0.154	0.137

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.

This feature prevents players from self-selecting into different playing modes, thereby eliminating potential bias from endogenous sorting. As Table A21 shows, the positive impact of remaining time on risk-taking is, both qualitatively and quantitatively, very comparable in this sample.

5.1.2. Variance of relative advantage

An additional robustness check is presented in Table 7. Instead of the variance of difficulty, it uses the logarithm of the variance of relative (dis)advantage of the five hypothetical boards.[24] 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 7 confirms the positive effect of remaining time on risk from Table 3.

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 8, we consider a definition of a blunder as 50 CP as well as similar definitions (100 centipawns, continuous measure in centipawns).[25] 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

[24] We drop cases where var(relative advantage) = 0 to avoid log(0), which reduces the overall sample size by less than 2%.
[25] In ten percent of observations in our sample, players choose a move that worsens the position evaluation by at least 50 CP.

Table 8

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	0.0955***		
(Threshold 50 CP)	(0.0152)		
Blunder in Previous Move		0.1011***	
(Threshold 100 CP)		(0.0246)	
Move Quality of Previous Move			0.0002***
			(0.0001)
Log. Average Variance Difficulty	0.5280***	0.5285***	0.5289***
of Alternatives	(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	156 260	156 260	156 260
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	162 024	157 427	155 016	151 370	145 409
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 A.2, 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)). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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 A.2 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.[26] Therefore, in Panel B of Table 8, 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

[26] In the present case, a blunder in the last move leads to deviations from the reference point. Analogies can be drawn from Table 5: 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 5, 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.

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.

6. Discussion and interpretation of main result

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 and replicates in analyses that use a different outcome for constructing the risk measure. In addition, the effect materializes comparably in a dataset where concerns about endogenous selection of players into playing modes are less likely to apply. Overall, we read these findings as evidence in favor of a genuine effect.

However, we are agnostic about the specific mechanisms behind these relationships. 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. Similarly, players could anticipate that they might make mistakes more easily with less time and thus aim to avoid getting into positions of higher difficulty, in turn leading them to favor risk-averse moves.

In addition to our main effect of interest, we document the presence of strategic loss aversion in the World Cup sample, 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, but ultimately remain agnostic about its origins in our sample, too.

Consequently, we thus 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. While we cannot conclusively ascertain that our effect of interest is due to a conscious choice of players, we emphasize that the effect is robustly identified across a plethora of analyses, including those in a separate dataset. We interpret this consistency as evidence in favor of the presence of the effect, refuting potential arguments that this effect is "just" due to how the data is constituted or processed.

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. This effect replicated in the 2019–2024 Biel Grandmaster Triathlons games, where concerns about endogenous selection are less pertinent. 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.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jebo.2025.107218.

Data availability

Data will be made available on request.

References

Alós-Ferrer, C., Garagnani, M., 2022. Voting under time pressure. Judgm. Decis. Mak. 17 (5), 1072–1093.
Anderson, A., Kleinberg, J., Mullainathan, S., 2017. Assessing human error against a benchmark of perfection. ACM Trans. Knowl. Discov. Data 11 (4).
Avoyan, A., Khubulashvili, R., Mekerishvili, G., 2020. Call it a day: History dependent stopping behavior. SSRN Electron. J..
Calford, E.M., 2020. Uncertainty aversion in game theory: Experimental evidence. J. Econ. Behav. Organ. 176, 720–734.
Chassy, P., Gobet, F., 2015. Risk taking in adversarial situations: Civilization differences in chess experts. Cognition 141, 36–40.
Chassy, P., Gobet, F., 2020. Influence of experts' domain-specific knowledge on risk taking in adversarial situations. J. Expert. 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. Econom. Lett. 208, 110085.
Dreber, A., Gerdes, C., Gränsmark, P., 2013. Beauty queens and battling knights: Risk taking and attractiveness in chess. J. Econ. Behav. Organ. 90, 1–18.
Ferreira, D.R., 2013. The impact of search depth on chess playing strength. ICGA J. 36 (2), 67–80.
FIDE, 2021. Regulations for the FIDE world cup 2021. https://handbook.fide.com/files/handbook/WorldCup2021Regulations.pdf.
Gerdes, C., Gränsmark, P., 2010. Strategic behavior across gender: A comparison of female and male expert chess players. Labour Econ. 17 (5), 766–775.
González-Díaz, J., Palacios-Huerta, I., 2016. Cognitive performance in competitive environments: Evidence from a natural experiment. J. Public Econ. 139, 40–52.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Gränsmark, P., 2012. Masters of our time: Impatience and self-control in high-level chess games. J. Econ. Behav. Organ. 82 (1), 179–191.
Guid, M., Bratko, I., 2006. Computer analysis of world chess champions. ICGA J. 29 (2), 65–73.
Haji, A.E., Krawczyk, M., Sylwestrzak, M., Zawojska, E., 2019. Time pressure and risk taking in auctions: A field experiment. J. Behav. Exp. Econ. 78, 68–79.
Heinecke, A., Ho, J., Hwang, W.-L., 2020. Refinement and universal approximation via sparsely connected ReLU convolution nets. IEEE Signal Process. Lett. 27, 1175–1179.
Holdaway, C., 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. pp. 986–992.
Howard, G., 2024. A check for rational inattention. J. Political Econ. Microeconomics 2 (1), 172–199.
Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. Econometrica 47 (2), 263.
Karagözoglu, E., Kocher, M.G., 2019. Bargaining under time pressure from deadlines. Exp. Econ. 22 (2), 419–440.
Kirchler, M., Andersson, D., Bonn, C., Johannesson, M., Sørensen, E.O., Stefan, M., Tinghög, G., Västfjäll, D., 2017. The effect of fast and slow decisions on risk taking. J. Risk Uncertain. 54 (1), 37–59.
Klingen, J., van Ommeren, J.N., 2020. Risk attitude and air pollution: Evidence from chess. Tinbergen Inst. Discuss. Pap. 2020-027/V.
Kocher, M.G., Pahlke, J., Trautmann, S.T., 2013. Tempus fugit: Time pressure in risky decisions. Manag. Sci. 59 (10), 2380–2391.
Künn, S., Palacios, J., Pestel, N., 2023. Indoor air quality and strategic decision making. Manag. Sci. 69.
Künn, S., Seel, C., Zegners, D., 2022. Cognitive performance in remote work: Evidence from professional chess. Econ. J. 132 (643), 1218–1232.
Larson, F., List, J.A., 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., Sadoff, S.E., 2011. Checkmate: Exploring backward induction among chess players. Am. Econ. Rev. 101 (2), 975–990.
Lindner, F., 2014. Decision time and steps of reasoning in a competitive market entry game. Econom. Lett. 122 (1), 7–11.
Linnemer, L., Visser, M., 2016. Self-selection in tournaments: The case of chess players. J. Econ. Behav. Organ. 126, 213–234.
Olschewski, S., Rieskamp, J., 2021. Distinguishing three effects of time pressure on risk taking: Choice consistency, risk preference, and strategy selection. J. Behav. Decis. Mak. 34 (4), 541–554.
Palacios-Huerta, I., Volij, O., 2009. Field centipedes. Am. Econ. Rev. 99 (4), 1619–1635.
Pope, D.G., Schweitzer, M.E., 2011. Is Tiger Woods loss averse? Persistent bias in the face of experience, competition, and high stakes. Am. Econ. Rev. 101 (1), 129–157.
Russek, E.M., Acosta-Kane, D., Van Opheusden, B., Mattar, M.G., Griffiths, T.L., 2022. Time spent thinking in online chess reflects the value of computation. PsyArxiv.
Salant, Y., Spenkuch, J.L., 2022. Complexity and choice. NBER Working Papers.
Sigman, M., Etchemendy, P., Slezak, D.F., Cecchi, G.A., 2010. Response time distributions in rapid chess: A large-scale decision making experiment. Front. Neurosci. 4 (OCT), 1–12.
Spiliopoulos, L., Ortmann, A., Zhang, L., 2018. Complexity, attention, and choice in games under time constraints: A process analysis. J. Exp. Psychol. [Learn. Mem. Cogn.] 44 (10), 1609–1640.
Steinerberger, S., 2015. On the number of positions in chess without promotion. Int. J. Game Theory 44 (3), 761–767.
Strittmatter, A., Sunde, U., Zegners, D., 2020. Life cycle patterns of cognitive performance over the long run. Proc. Natl. Acad. Sci. 117 (44), 27255–27261.
Sunde, U., Zegners, D., Strittmatter, A., 2022. Speed, quality, and the optimal timing of complex decisions: Field evidence. (280092119).
Wu, C.M., Schulz, E., Pleskac, T.J., Speekenbrink, M., 2022. Time pressure changes how people explore and respond to uncertainty. Sci. Rep. 12 (1).
Zegners, D., Sunde, U., 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. Appl. Comput. Harmon. Anal. 48 (2), 787–794.