Short Paper

Do Mistakes Provoke New Mistakes? Evidence From Chess

Akash Adhikari, Stanislav Anatolyev, and Dmitry Dagaev

Abstract—**We investigate how the mistakes of professional chess players affect the quality of their further moves in the same game. Using a database of games played by top chess players, Stockfish chess engine evaluations, and an ordered probit regression analysis, we found clear evidence that most mistakes provoke tilt, which leads to less accurate future play, while in reaction to serious blunders players instead discipline their play.**

Index Terms—**Chess, hot hand, mistakes, ordered probit, tilt.**

I. Introduction

It follows from Zermelo's ideas [1] and formally stated and proved by Kalmár [2] that in a finite version of chess (a game necessarily ends after the third repetition of a position), either white can guarantee a win, or black can guarantee a win, or both white and black can guarantee a draw. Therefore, if two rational players with sufficient search capacities play 100 games, there should be either 100 wins by white, or 100 wins by black, or 100 draws. In 2022, the search capacities of modern computers are still not enough to identify which of the three alternatives holds. Ewerhart [3] demonstrated that the infinite version of chess (players can claim a draw after the third repetition of a position but are not obliged to do so) is equivalent to the finite version in the sense that the same of three alternatives takes place.

Various sources indicate that the actual white's winning percentage is higher than black's; the Chess game database, which contains more than 900 000 games, consists of approximately 38% wins by white, 34% draws, and 28% wins by black.[1] It follows from [2] that the difference in outcomes results from players' suboptimal moves. One can subjectively evaluate their position on the board based on the set of seemingly achievable positions, material on the board, positional advantages, and other criteria. There are many common knowledge strategic principles in chess; disregarding some of them leads to worse chances and ignoring others can lead to a loss. Chess players make mistakes that differ in severity: from slight inaccuracies to game-deciding blunders. Empirical evidence shows that humans make worse mistakes in positions with the same evaluation than computer programs do [4], [5]. Recent developments in computer technologies has made it impossible for humans to compete with the best computer programs in strategic games, such as chess and go, not to mention the solved game of checkers [6].

The realization of having made a mistake can put a human player into the state known as tilt, which is an emotional state of mind that leads to repeatedly suboptimal strategic decisions and may result in a loss. This is an additional disadvantage for human players compared to computer programs. In this article, we aim to uncover the sequential patterns of mistakes made during a game of chess. Using a database of games played by top chess players, we empirically confirm the presence of tilt in chess. We find that recent small inaccuracies lead to less accurate play in future. Small, moderate and severe mistakes have a weaker effect in the same direction. At the same time, blunders surprisingly tend to discipline players. We confirm previous findings that the historical average level of mistakes matters [4], and demonstrate that mistakes made in the previous move and overall previous erroneous play are both strong predictors of suboptimal move.

The term "tilt" originated in poker. There is a strong consensus among both the poker community and academics that tilt exists in poker [7], [8], [9], [10], [11]. According to Browne [8], tilt starts from a tilt-inducing situation followed by an internal emotional struggle to retain control and deterioration of the player's decision making. Browne [8] described many possible tilt-inducing forces, such as bad beats (unfavorable realizations of random events), needling, problems at work or home, and consumption of drugs and/or alcohol. All of these forces are linked to bad luck or external factors. If a player feels that they have lost due to bad luck, they can try to compensate for the loss by subsequently increasing the pot. Such behavior can be consistent with the Kahneman–Tversky prospect theory that postulates that people are risk loving when they are in the zone of losses compared to the initial reference point [11], [12]. At the same time, overbets lead to deviations from the Nash equilibrium and opponents can potentially exploit them. Smith, Levere, and Kurtzman [12] confirmed that poker players behave less cautiously after losses. For a more detailed survey on poker players' behavior we refer the reader to [13].

In contrast to poker, which is regarded by many as a game of both skill and chance [14], chess is a purely strategic game with no random elements. Chess players never experience bad beats. If bad luck was the only cause of tilt, one would imagine that chess players never experience it. The results of this article show that this is not the case.

The behavior of chess players is a notable area of study in cognitive science. In general, better chess players have stronger mental abilities [15] and choose better moves [16]. Chabris and Hearst [17] demonstrated that grandmasters make many more mistakes in rapid chess than in classical games. Moreover, the magnitude of the mistakes made in rapid chess is larger. At the same time, no difference has been found in blindfold and rapid chess variations. Burns [18] showed that beyond a minimal threshold, extra time does not help in making better decisions. On the contrary, in a chess problem-solving setting, additional time is helpful in finding better moves [19].

Manuscript received 15 May 2022; revised 5 January 2023; accepted 8 May 2023. *(Corresponding author: Dmitry Dagaev.)*

Akash Adhikari is with the Indian Institute of Technology (ISM), Dhanbad 826004, India (e-mail: rajaadhikari23@gmail.com).

Stanislav Anatolyev is with the CERGE-EI, 111 21 Prague, Czech Republic, and also with the New Economic School, 121353 Moscow, Russia (e-mail: stanislav.anatolyev@cerge-ei.cz).

Dmitry Dagaev is with the HSE University, 101000 Moscow, Russia (e-mail: ddagaev@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TG.2023.3275710.

Digital Object Identifier 10.1109/TG.2023.3275710

[1]See http://www.chessgames.com/chessstats.html. Retrieved April 1, 2021.

To the best of our knowledge, the effect of previous mistakes and their severity on the probability of making future mistakes of different severities has not been econometrically evaluated before for the game of chess. The most relevant topic that attracted a lot of researchers' attention is the so called "Hot Hand" phenomenon. In their classical article, Gilovich, Vallone, and Tversky [20] tested a popular belief that basketball players experience streak shooting, so that the probability of scoring a goal increases after another successful shot. The authors disproved the hot hand hypothesis and attributed the myth to the wrong perception of chance. Despite the negative result, the work started a series of articles on the existence of hot hand in various environments. Most of the empirical articles supported the conclusions of [20] for the game of basketball and some other sports (see, for example, [21], [22], [23]). Seemingly less frequently analyzed concept of a cold hand, the existence of disproportionately often streak failures, is closely connected to the concept of tilting. The principal difference between basketball and chess is that shots in basketball can be considered as an iterated exercise (especially in the case of free throws), whereas each position in chess is unique. Therefore, we avoid to make a clear link between making moves in chess and iterated throws in basketball. Instead, we prefer to use the concept of tilt which allows to carry over the negative emotions from realizing of making a mistake to the subsequent moves.

The rest of this article is organized as follows. Section II describes the data. Section III discusses the econometric model and empirical strategy. Section IV contains the results. Finally, Section V concludes this article.

II. Data

We collected all games of the main Tata Steel Chess Tournament that takes place annually in Wijk aan Zee (The Netherlands). In total, 885 games were played between 2011 (the first year when the tournament in Wijk aan Zee was named Tata Steel Chess Tournament) and 2020 (the last year in our database). The tournament is organized in a round-robin format—each player plays against each other player once. In 2014, there were 12 players and 66 games in total, whereas in all other nine years there were 14 players and 91 games. Notation for the games is available in Portable Game Notation (PGN) format, which is a standard designed for representing chess game data. In chess, FIDE[2] rating is used to evaluate a player's relative skill level. It is based on the Elo rating principles proposed by Arpad Elo. When two chess players who already have the rating play each other, a certain number of rating points is transferred from the loser to the winner; in case of a draw, points are transferred from the higher rated player to the lower rated player. The exact number of points transferred from one player to the opponent is a function of their ratings and the outcome of a game. For each game in our dataset, we collected the FIDE ratings of both opponents at the time when the game was played. All games were played by highly rated professionals whose FIDE rating ranged from 2603 to 2872.

In order to evaluate chess positions, we use the open source chess engine Stockfish 12 [24], which was also used in some other behavioral and socio-economic performance-related studies [4], [25], [26], [27]. As of 2022, Stockfish is widely regarded as the strongest open source computer chess program.[3] Historically, Stockfish evaluated a position by looking through a game tree starting at the current position as deeply as the time limit allows. A limited number of apparently good lines are looked through more deeply than others. In September 2020, a new version Stockfish 12 was released, and it was announced that Stockfish had absorbed a neural network project.[4] We manually set the number of good lines to 9 and the time limit to 7 seconds per move, which leads to the search depth of at least 17 half moves.[5] As the time limit expires, Stockfish suggests the best possible line and a numerical evaluation of the position corresponding to that line. If at some position one of the opponents has a guaranteed win by checkmate, Stockfish provides "white/black mates in k moves" instead of a numerical value. The evaluator takes into account various factors: existence of a forced checkmate, material advantage, positional weaknesses (isolated pawns, doubled pawns, etc.), and positional advantages (two bishops, rooks on open lines, etc.). All scores are normalized so that an extra pawn for white leads to the score of +1.00 given all else equal. Since chess is an antagonistic game, the score of some position for black is simply the score of that position for white with the opposite sign.

The website[6] is a popular online chess platform that uses the Stockfish 12 engine. We uploaded our PGN files to[7] one by one in order to obtain Stockfish evaluations. The data were extracted as `.txt` files by web scraping using javascript. For each position, we collected a numerical evaluation suggested by Stockfish. Now, for each move $m_g = 1, 2, \ldots$ played by one of the players in game g, we define the so-called *centipawn loss* variable cl_{g,m_g} that shows the quality of this move

$$\mathrm{cl}_{g,m_g} = \begin{cases} ea_{g,m_g} - eb_{g,m_g} & \text{for white} \\ -(ea_{g,m_g} - eb_{g,m_g}) & \text{for black} \end{cases} \tag{1}$$

where, ea_{g,m_g} is the evaluation of the position after the move m_g and eb_{g,m_g} is the evaluation of the position before the move m_g. Similar metrics for quality of moves are used in the literature [4], [26]. If a player chooses the best possible move, cl_{g,m_g} is expected to be 0 (the optimal line after the move is the same as before the move). If a player fails to choose the best move, cl_{g,m_g} is expected to be negative. However, note that after a move is made, Stockfish starts its analysis from the next opponent's move. Due to this discrepancy in the search depths before and after a move, evaluation of the position can be changed (in both directions) after the move even if the best move was played. It explains why there do exist moves with positive value of cl_{g,m_g}. Finally, we refer to [27] where it was shown that the engine set at the search depth of 17 half-moves chooses a move with an average error of less than 3 centipawns (or 0.03). This can also be interpreted as an upper bound for the average evaluation error of a position due to the limited search time. We think that such error is acceptable for the purposes of this research.

We also made the following adjustments to the dataset. First of all, we excluded from the dataset the first five moves of each game. This step removes most spurious evaluations associated with the white's first-mover advantage and forms minimal play history for the current game. However, these first five moves are still used to generate "lagged" variables related to previous play for moves 6 through 10 (see the next section). Next, in chess, there are many ways to win a decided game. Usually, chess players prefer to use a safe one. For example, reduction to a theoretically winning position would be preferred to a fast but rather complex combination, even if the safe way would be much longer. From the Stockfish perspective, using safe ways is sometimes interpreted as a mistake or even as a blunder. In order to account for this, we excluded decided positions from our dataset. Particularly, suppose that move p_g is the first move of game g such that the absolute value of evaluation is higher than 5.00 (such advantage corresponds to an extra rook, and

[2] International Chess Federation.

[3] [Online]. Available: https://ccrl.chessdom.com/. Retrieved January 1, 2023.

[4] [Online]. Available: https://www.chess.com/terms/stockfish-chess-engine. Retrieved February 23, 2021.

[5] A half-move is a move of White or a move of Black.

[6] [Online]. Available: www.chess.com

[7] [Online]. Available: www.chess.com

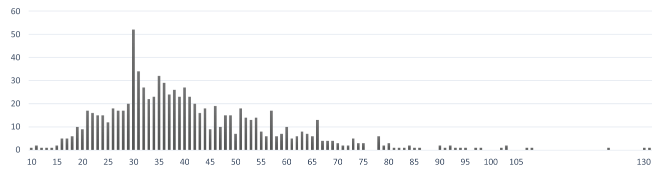

Fig. 1. Distribution of the number of moves in games from the sample. The number of moves is represented on *X*-axis, the number of games in the dataset with a particular number of moves is represented on *Y*-axis.

TABLE I
CHARACTERISTICS OF MISTAKES

Level	Lower threshold	Upper threshold	Mistake description	Fraction of moves
0	0.0	—	No mistake	29.88%
1	−0.5	−0.0	An inaccuracy	57.36%
2	−1.0	−0.5	A small mistake	9.40%
3	−1.5	−1.0	A moderate mistake	1.97%
4	−2.0	−1.5	A severe mistake	0.70%
5	—	−2.0	A blunder	0.69%

for a strong chess player, it is more than enough for a win, see [27] for statistics), or a checkmate. We have deleted all observations from this game starting from this move, so move $p_g - 1$ will be the last move of this game. As a result of these adjustments, the moves in game g are indexed now by $m_g = 6, 7, \ldots, p_g - 1$. Finally, for all games in our dataset we excluded the last half-move due to technical issues related to extraction of the position score before switching to another game. For the whole sample of 885 games, this leads to 64 404 moves and hence observations. Fig. 1 shows a distribution of these moves across games.

The database contains only games from very strong international level chess players. We make a plausible assumption that the players do not intentionally choose suboptimal moves because opponents can exploit even small suboptimality at this level.

III. ECONOMETRIC MODEL

Let

$$-\infty = b_{B+1} < b_B < \ldots < b_2 < b_1 = 0 < b_0 = \infty \quad (2)$$

be the score thresholds that define levels of mistakes of different severity. The variable representing the mistake of severity level $j = 1, \ldots, B$ made at move m_g in game g is equal to

$$\mathbb{I}_{j,g,m_g} = \mathbb{I}\left\{b_{j+1} \le \mathrm{cl}_{g,m_g} < b_j\right\} \quad (3)$$

where, $\mathbb{I}\{\cdot\}$ is an indicator function. By convention, values $j = 0$ and $b_0 = \infty$ correspond to no mistake made (mistake of level 0); in this case $\mathrm{cl}_{g,m_g} \ge 0$.

As a practical matter, we consider $B = 5$ levels of mistakes of the following severity levels. Table I shows their cutoffs, characterizations, and in what fraction of moves these mistakes are made in the database we consider.

Our econometric model is based on the ordered multiple choice regression where the left-hand side variable is a type of a mistake made (or not made) after each move, and the right-hand side variables describe the quality of the same player's previous play in the game, in addition to a number of covariates that characterize the player and the game. Specifically, we define a latent variable pm_{g,m_g}, which we call a propensity to misplay

$$\mathrm{pm}_{g,m_g} = z'_{g,m_g}\gamma + x'_{g,m_g}\beta + \alpha_g + \varepsilon_{g,m_g}. \quad (4)$$

Here, the vector of covariates z_{g,m_g} contains controls specific to move m_g in game g or to game g alone, and not directly related to the player's past performance in the game, while x_{g,m_g} contains predictors that describe the previous play in game g before move m_g. The next component, α_g, is a game-specific move-independent random effect. Finally, ε_{g,m_g} is an idiosyncratic random component not explained by the included regressors.

A mistake of type $j = 0, 1, \ldots, B$ (recall that 0 stands for no mistake, and an increasing j corresponds to more severe mistakes) occurs when the propensity to misplay pm_{g,m_g} falls in the region $[A_{j+1}, A_j)$, where $A_{B+1} = -\infty$, $A_0 = \infty$, and A_j, $j = 1, \ldots, B$, are unknown cutoffs. Under the assumption that α_g and ε_{g,m_g} are normally distributed independently of included regressors, the mistake of type $j = 0, \ldots, B$ has conditional probability

$$\Pr\left\{b_{j+1} \le \mathrm{cl}_{g,m_g} < b_j\right\} = \quad (5)$$

$$= \Phi(A_j - z'_{g,m_g}\gamma - x'_{g,m_g}\beta) - \Phi(A_{j+1} - z'_{g,m_g}\gamma - x'_{g,m_g}\beta).$$

where Φ is a standard normal cumulative distribution function. Such an ordered probit model means that the probability of a mistake of level j depends on the characteristics of the player, of the move, of the game, and of the previous play.

We include the following variables to z_{g,m_g}, in addition to a constant.

1) elo_g, an Elo rating of the player making move m_g in game g.
2) ev_{g,m_g}, an evaluation before the move

$$\mathrm{ev}_{g,m_g} = \begin{cases} eb_{g,m_g} & \text{for white} \\ -eb_{g,m_g} & \text{for black.} \end{cases} \quad (6)$$

3) taken_{g,m_g}, a number of pieces gone from the board before move m_g in game g.
4) white_{g,m_g}, an indicator that the move m_g in game g is made by white.

The variable elo_g depends only on parameters of the player making the move in the game, and is meant to capture the direct effect of the players' strength on the sequential pattern of mistakes: a weaker player's more serious mistake may increase the probability of this player's next more serious mistake. The variable taken_{g,m_g} is a proxy for a stage of the game, [8] which may affect tilt formation. The variable white_{g,m_g} is meant to capture the heterogeneity from the color of pieces, as this may affect the psychological state and strategy of the player.

While it is interesting to see the effects of the abovementioned covariates, our primary interest is analyzing the effects of the previous play. Because the previous mistakes may be characterized by many different variables, we adopt simple empirical strategies to select the most influential predictors from a limited set of possibilities. Specifically, the list of candidates to include in x_{g,m_g} is:

1) $\mathbb{I}_{j,g,m_g-\ell}$, the fact of making a mistake of jth severity at move $m_g - \ell$ in game g, for $j = 1, \ldots, B$ and $\ell = 1, \ldots, L$;
2) ab^-_{g,m_g}, a historical average of one's mistakes of any level, in game g before move m_g during L previous moves

$$ab^-_{g,m_g} = \frac{1}{L}\sum_{\ell=1}^{L}\left|\mathrm{cl}_{g,m_g-\ell}\right|\sum_{j=1}^{B}\mathbb{I}_{j,g,m_g-\ell}; \quad (7)$$

[8] A chess game is characterized by three stages: Début (beginning), Mittelspiel (middle game), and Endspiel (endgame). All stages have their own specifics and gradually transform one into another. However, by which move the stages transit from one to another is not predetermined but depends on the style of a particular game.

TABLE II
ORDERED PROBIT REGRESSION, COEFFICIENTS ON PREDICTORS BASED ON PREVIOUS PLAY

Mistake indicators					
lag	$\mathbb{I}_1$	$\mathbb{I}_2$	$\mathbb{I}_3$	$\mathbb{I}_4$	$\mathbb{I}_5$
1	0.200^{***} (0.014)	0.043^{**} (0.018)		0.137^{*} (0.077)	-0.242^{**} (0.119)
2	0.126^{***} (0.013)		0.073^{*} (0.038)		-0.407^{***} (0.104)
3	0.119^{***} (0.013)		0.085^{*} (0.044)		-0.401^{***} (0.125)
4	0.101^{***} (0.013)				-0.369^{***} (0.118)
5	0.061^{***} (0.013)				-0.504^{***} (0.117)
Aggregate mistakes					
		ab^-		xb^-	
		0.562^{***} (0.117)		0.143^{***} (0.042)	

Robust clustered standard errors in parentheses; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3) xb^-_{g,m_g}, one's move with worst centipawn loss in game g before move m_g during L previous moves

$$xb^-_{g,m_g} = \max_{\ell=1,\ldots,L}(-\mathrm{cl}_{g,m_g-\ell}). \tag{8}$$

The first type of predictors is $\mathbb{I}_{j,g,m_g-\ell}$, the indicator of a mistake of level j in one of L most recent moves. These indicators are meant to absorb the short term effects during a recent play. In total, we have BL predictors of such "individual" move-to-move type. The other two predictors are of "aggregate" type, as they index how, on average or in extreme terms, erroneous the play have been up until the current move is made. These two variables are meant to absorb the long term effects during the whole play in a game. The variable ab^-_{g,m_g} indicates how large the errors have been on average, and is meant to capture the overall psychological state of a player based on previous mistakes made. The variable xb^-_{g,m_g} indicates how big the maximal mistake in recent previous play has been, and is meant to capture the emotional distress caused by this mistake on the following play. As a practical matter, we set L to 5, which is arguably sufficient to capture the psychological state resulting from a recent play. In total, when $B = 5$ and $L = 5$, there are 27 mistake-related predictors.

We now address a few econometric issues and how we handle them. We perform quasimaximum likelihood estimation of the ordered probit model. The influence of "serial" correlation across moves within the same game on the asymptotic variance of parameter estimates is taken care of by clustering by the game (see, e.g., [28]). The within-game "color effect" in each game is automatically taken care of by including the indicator of white among the covariates. To select only a few from the list of "previous mistakes" predictors, we implement a general-to-specific stepwise selection procedure [29]. Specifically, we fix the list of included covariates z_{g,m_g}, and set the tolerance level to statistical significance of selected predictors from the abovementioned list of potential x_{g,m_g}'s to 10%, i.e., we stop removing predictors when none of those that are left has a coefficient with a p-value exceeding 10%.[9]

IV. EMPIRICAL RESULTS

We now look at the pattern of how the quality of previous play affects the propensity to make errors in further play. Table II reports the results of running the ordered probit regression on the included covariates and significant predictors selected, as described in the previous section.[10] The coefficients in the table represent the marginal effects of each predictor on the latent propensity to misplay, and are eventually related to the probabilities of making mistakes.[11] In particular, a positive sign of a covariate/predictor implies its positive effect on the propensity to misplay and hence a negative effect on a quality of play. Conversely, a negative sign of a covariate/predictor implies its positive effect on a quality of play. The figures in the "mistake indicators" subpanel are regression coefficients for the short term predictors—the indicators $\mathbb{I}_{j,g,m_g-\ell}$ corresponding to the fact of making a mistake of jth severity for "lag" $\ell = 1, \ldots, 5$. Analogously, the figures in the "aggregate mistakes" subpanel are regression coefficients for the long term predictors—a historical average of mistakes ab^-_{g,m_g} and historical maximal mistake xb^-_{g,m_g}.

First, let us look at the effects of selected lagged mistake indicators on the propensity to misplay. It is striking that different levels of mistake severity may make impact of a different strength and even a different sign. While the mistakes of moderate severity are statistically less significant, the small inaccuracies and big blunders are statistically most significant for all five included lags. They also tend to have more pronounced numerical effects but those effects are of opposite signs. Small inaccuracies, especially their most recent occurrences, increase the propensity to misplay, provoking the tilt. The same is true, although less strongly,[12] for small, moderate and severe mistakes; however, their effects seem to be shorter lived.

In contrast, the estimates coefficients of blunder indicators are starkly different: all negative and relatively large in absolute value. This brings a conclusion that, in reaction to their blunders, players tend instead to discipline their play. Moreover, in addition to its bigger size, this effect turns out to be longer lived than the tilting effect of less severe mistakes.

Next, the last two columns of Table II show the effect of the two aggregated measures of previous erroneous play during the last five moves, which may cause overall emotional distress. Notice that these measures are statistically significant even though all the individual mistake indicators for the same five periods are already included in the regression. Hence, there is strong predictive information in the average and maximal mistakes made in the previous play, on top of occurrences of each mistake. Both effects are positive for the propensity of further misplay, and strongly confirm the presence of tilt.

It is also interesting to examine how the quality of play is influenced by the characteristics of the game, the moves, and the players. Table III reports the estimates on included covariates except for a constant (we again remove regressors' indexes to reduce clutter). All of the coefficients of included covariates are strongly statistically significant and have intuitively sensible signs. One can see that a player's Elo rating has a positive, although small in value, effect on the quality of play, which is intuitive. The current evaluation positively influences the propensity to misplay, meaning that a player is more likely to become careless and possibly reckless in a better position. Next, the proxy for the stage of the game perhaps affects the quality of play—the tree of a subgame becomes less deep closer to an endgame. Finally, being white has a favorable effect on preventing mistakes.

[9]Our preference for 10% is motivated by a desire to end up with a more liberal post-selection specification so that not to miss important predictors.

[10]In Table II, we intentionally remove predictors' indexes to reduce clutter.

[11]A reader should keep in mind that the absolute values of the marginal effects do not carry much information, because the composite error is normalized to have unit variance for the purpose of identification. Thus, it is their values relative to each other, taking the predictor scales into account when those scales are different, that is meaningful and interpretable.

[12]Note that with the significance threshold of 5% for the stepwise selection procedure, the $\mathbb{I}_3$ and $\mathbb{I}_4$ predictors would not be selected at all, with no noticeable changes in the rest of the results.

TABLE III
ORDERED PROBIT REGRESSION, COEFFICIENTS ON INCLUDED COVARIATES

Covariate	Coefficient estimate, $\times 10^{-2}$	Covariate scale
elo	-0.0726^{***} (0.0096)	54.4
ev	3.71^{***} (0.52)	1.20
taken	-1.45^{***} (0.11)	7.15
white	-2.84^{***} (0.87)	0.50

Robust clustered standard errors below point estimates; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Last column lists standard deviations.

We would like to emphasize that even though all these covariates are strongly statistically significant, their numerical effects (accounting for variables' scales; see standard deviations in the last column of Table III) are appreciably smaller than those of the indicators or aggregate measures of previous play documented in Table II. Among the four covariates, the variable "taken" has the greatest impact on the quality of play, given its biggest product of the coefficient and variable's standard deviation among all, the variable "ev" coming the second.

Even though the presented regression results give a strong evidence of influence of mistakes on the quality of further play, we perform a formal test for inclusion of all $BL + 2$ previous mistake related predictors. The Wald test statistic for their joint significance equals 1062, with an essentially zero p-value relative to the $\chi^2_{(27)}$ distribution. A similar outcome results if we jointly test the exclusion restrictions for the included "previous mistakes" predictors only.

Moreover, it is interesting to compare the measures of regression fit from the ordered probit models with and without the previous mistake related predictors. The difference will show a relative contribution of the mistake-related predictors to the explanatory power of covariates. For the full model with all predictors included, the pseudo-R^2 equals 3.11%, and in the full model with only stepwise-selected predictors, the pseudo-R^2 equals 3.10%, an almost identical figure. At the same time, the ordered probit model with all the predictors excluded and only the covariates left, the pseudo-R^2 equals 0.93%. This shows that previous mistakes have a much larger role in determining the quality of further play than explanatory variables from Table III, at least among the top players.

V. CONCLUSION

In this article, we have uncovered sequential patterns of mistakes of human players in the game of chess. We have found clear evidence that small inaccuracies lead to less accurate play in future; more severe mistakes have a weaker effect on the quality of play in the same direction, while blunders tend to discipline players. Inaccuracies and blunders have more long-lived effect than mistakes of moderate size do.

One should have in mind that our database contains games played by strong chess players. The pattern could be different for lower ranked players due to their lower ability to find best moves. On the one hand, higher variance of their quality of play could dominate psychological effects. On the other hand, lower ranking can potentially incorporate information about the resistance to tilt. Therefore, a further careful analysis is required for that cohort of players.

We acknowledge that one should be careful in interpreting the findings of this study. Although tilt seems to be the most obvious explanation for the fact that some types of mistakes increase the probability of a new mistake, our methodology does not allow to exclude other possible explanations not related to the psychological state of mind. Alternative theories include the changing attitude toward risk (chess players may look for complications in worse positions) and peculiarities of the Stockfish evaluation algorithm (the difference between the scores +4 and +5 in the decided positions can be due to the arguments that are not taken into account by human players). We hope that future research will allow to differentiate between these theories.

ACKNOWLEDGMENT

The authors would like to thank Alena Skolkova for excellent research assistance and Petr Parshakov for helpful comments. Dmitry Dagaev gratefully acknowledges support from the Basic Research Program of the National Research University Higher School of Economics.

REFERENCES

[1] E. Zermelo, "Uber eine anwendung der mengenlehre auf die theorie des schachspiels," in *Proc. 5th Int. Congr. Mathematicians*, 1913, vol. 2, pp. 501–504.
[2] L. Kalmár, "Zur theorie der abstrakten spiele," *Acta Universitatis Szegediensis/Sectio Scientiarum Mathematicarum*, vol. 4, pp. 65–85, 1928.
[3] C. Ewerhart, "Backward induction and the game-theoretic analysis of chess," *Games Econ. Behav.*, vol. 39, no. 2, pp. 206–214, 2002.
[4] K. W. Regan, T. Biswas, and J. Zhou, "Human and computer preferences at chess," in *Proc. Workshops 20th AAAI Conf. Artif. Intell.*, 2014, pp. 79–84.
[5] K. W. Regan, B. Macieja, and G. M. Haworth, "Understanding distributions of chess performances," *Adv. Comput. Games*, vol. 13, pp. 230–243, 2011.
[6] J. Schaeffer et al., "Checkers is solved," *Science*, vol. 317, no. 5844, pp. 1518–1522, 2007.
[7] S. Barrault, A. Untas, and I. Varescon, "Special features of poker," *Int. Gambling Stud.*, vol. 14, no. 3, pp. 492–504, 2014.
[8] B. R. Browne, "Going on tilt: Frequent poker players and control," *J. Gambling Behav.*, vol. 5, no. 1, pp. 3–21, 1989.
[9] J. Palomäki, M. Laakasuo, and M. Salmela, "This is just so unfair!': A qualitative analysis of loss-induced emotions and tilting in on-line poker," *Int. Gambling Stud.*, vol. 13, no. 2, pp. 255–270, 2013.
[10] J. Palomäki, M. Laakasuo, and M. Salmela, "Losing more by losing it: Poker experience, sensitivity to losses and tilting severity," *J. Gambling Stud.*, vol. 30, no. 1, pp. 187–200, 2014.
[11] T. Toneatto, "Cognitive psychopathology of problem gambling," *Substance Use Misuse*, vol. 34, no. 11, pp. 1593–1604, 1999.
[12] G. Smith, M. Levere, and R. Kurtzman, "Poker player behavior after big wins and big losses," *Manage. Sci.*, vol. 55, no. 9, pp. 1547–1555, 2009.
[13] A. Moreau, H. Chabrol, and E. Chauchard, "Psychopathology of online poker players: Review of literature," *J. Behav. Addictions*, vol. 5, no. 2, pp. 155–168, 2016.
[14] G. Meyer, von Meduna, M. T. Brosowski, and T. Hayer, "Is poker a game of skill or chance? A quasi-experimental study," *J. Gambling Stud.*, vol. 29, no. 3, pp. 535–550, 2013.
[15] R. H. Grabner, E. Stern, and A. C. Neubauer, "Individual differences in chess expertise: A psychometric investigation," *Acta Psychologica*, vol. 124, no. 3, pp. 398–420, 2007.
[16] N. Charness, "Search in chess: Age and skill differences," *J. Exp. Psychol.: Hum. Percep. Perform.*, vol. 7, no. 2, pp. 467–476, 1981.
[17] C. F. Chabris and E. S. Hearst, "Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors," *Cogn. Sci.*, vol. 27, no. 4, pp. 637–648, 2003.
[18] B. D. Burns, "The effects of speed on skilled chess performance," *Psychol. Sci.*, vol. 15, no. 4, pp. 442–447, 2004.
[19] J. H. Moxley, K. A. Ericsson, N. Charness, and R. T. Krampe, "The role of intuition and deliberative thinking in experts' superior tactical decision-making," *Cognition*, vol. 124, no. 1, pp. 72–78, 2012.
[20] T. Gilovich, R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cogn. Psychol.*, vol. 17, no. 3, pp. 295–314, 1985.
[21] M. Bar-Eli, S. Avugos, and M. Raab, "Twenty years of "hot hand" research: Review and critique," *Psychol. Sport Exercise*, vol. 7, no. 6, pp. 525–553, 2006.
[22] J. J. Koehler and C. A. Conley, "The "hot hand" myth in professional basketball," *J. Sport Exercise Psychol.*, vol. 25, no. 2, pp. 253–259, 2003.

[23] A. Tversky and T. Gilovich, “The cold facts about the “hot hand” in basketball,” *Chance*, vol. 2, no. 1, pp. 16–21, 1989.

[24] T. Romstad, M. Costalba, and J. Kiiski, “Stockfish: A strong open source chess engine.” [Online]. Available: https://stockfishchess.org

[25] S. Künn, C. Seel, and D. Zegners, “Cognitive performance in remote work: Evidence from professional chess,” *Econ. J.*, vol. 132, no. 643, pp. 1218–1232, 2022.

[26] D. J. Barnes and J. Hernandez-Castro, “On the limits of engine analysis for cheating detection in chess,” *Comput. Secur.*, vol. 48, pp. 58–73, 2015.

[27] T. Biswas and K. Regan, “Measuring level-k reasoning, satisficing, and human error in game-play data,” in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl.*, 2015, pp. 941–947.

[28] A. C. Cameron and D. L. Miller, “A practitioner’s guide to cluster-robust inference,” *J. Hum. Resour.*, vol. 50, no. 2, pp. 317–372, 2015.

[29] J. Campos, N. R. Ericsson, and D. F. Hendry, Eds., *General-to-Specific Modelling*. Cheltenham, U.K.: Edward Elgar Publishing, 2005.

Short Paper

Do Mistakes Provoke New Mistakes? Evidence From Chess

Akash Adhikari, Stanislav Anatolyev, and Dmitry Dagaev

***Abstract*—We investigate how the mistakes of professional chess players affect the quality of their further moves in the same game. Using a database of games played by top chess players, Stockfish chess engine evaluations, and an ordered probit regression analysis, we found clear evidence that most mistakes provoke tilt, which leads to less accurate future play, while in reaction to serious blunders players instead discipline their play.**

***Index Terms*—Chess, hot hand, mistakes, ordered probit, tilt.**

I. Introduction

It follows from Zermelo's ideas [1] and formally stated and proved by Kalmár [2] that in a finite version of chess (a game necessarily ends after the third repetition of a position), either white can guarantee a win, or black can guarantee a win, or both white and black can guarantee a draw. Therefore, if two rational players with sufficient search capacities play 100 games, there should be either 100 wins by white, or 100 wins by black, or 100 draws. In 2022, the search capacities of modern computers are still not enough to identify which of the three alternatives holds. Ewerhart [3] demonstrated that the infinite version of chess (players can claim a draw after the third repetition of a position but are not obliged to do so) is equivalent to the finite version in the sense that the same of three alternatives takes place.

Various sources indicate that the actual white's winning percentage is higher than black's; the Chess game database, which contains more than 900 000 games, consists of approximately 38% wins by white, 34% draws, and 28% wins by black.[1] It follows from [2] that the difference in outcomes results from players' suboptimal moves. One can subjectively evaluate their position on the board based on the set of seemingly achievable positions, material on the board, positional advantages, and other criteria. There are many common knowledge strategic principles in chess; disregarding some of them leads to worse chances and ignoring others can lead to a loss. Chess players make mistakes that differ in severity: from slight inaccuracies to game-deciding blunders. Empirical evidence shows that humans make worse mistakes in positions with the same evaluation than computer programs do [4], [5]. Recent developments in computer technologies has made it impossible for humans to compete with the best computer programs in strategic games, such as chess and go, not to mention the solved game of checkers [6].

The realization of having made a mistake can put a human player into the state known as tilt, which is an emotional state of mind that leads to repeatedly suboptimal strategic decisions and may result in a loss. This is an additional disadvantage for human players compared to computer programs. In this article, we aim to uncover the sequential patterns of mistakes made during a game of chess. Using a database of games played by top chess players, we empirically confirm the presence of tilt in chess. We find that recent small inaccuracies lead to less accurate play in future. Small, moderate and severe mistakes have a weaker effect in the same direction. At the same time, blunders surprisingly tend to discipline players. We confirm previous findings that the historical average level of mistakes matters [4], and demonstrate that mistakes made in the previous move and overall previous erroneous play are both strong predictors of suboptimal move.

The term "tilt" originated in poker. There is a strong consensus among both the poker community and academics that tilt exists in poker [7], [8], [9], [10], [11]. According to Browne [8], tilt starts from a tilt-inducing situation followed by an internal emotional struggle to retain control and deterioration of the player's decision making. Browne [8] described many possible tilt-inducing forces, such as bad beats (unfavorable realizations of random events), needling, problems at work or home, and consumption of drugs and/or alcohol. All of these forces are linked to bad luck or external factors. If a player feels that they have lost due to bad luck, they can try to compensate for the loss by subsequently increasing the pot. Such behavior can be consistent with the Kahneman–Tversky prospect theory that postulates that people are risk loving when they are in the zone of losses compared to the initial reference point [11], [12]. At the same time, overbets lead to deviations from the Nash equilibrium and opponents can potentially exploit them. Smith, Levere, and Kurtzman [12] confirmed that poker players behave less cautiously after losses. For a more detailed survey on poker players' behavior we refer the reader to [13].

In contrast to poker, which is regarded by many as a game of both skill and chance [14], chess is a purely strategic game with no random elements. Chess players never experience bad beats. If bad luck was the only cause of tilt, one would imagine that chess players never experience it. The results of this article show that this is not the case.

The behavior of chess players is a notable area of study in cognitive science. In general, better chess players have stronger mental abilities [15] and choose better moves [16]. Chabris and Hearst [17] demonstrated that grandmasters make many more mistakes in rapid chess than in classical games. Moreover, the magnitude of the mistakes made in rapid chess is larger. At the same time, no difference has been found in blindfold and rapid chess variations. Burns [18] showed that beyond a minimal threshold, extra time does not help in making better decisions. On the contrary, in a chess problem-solving setting, additional time is helpful in finding better moves [19].

Manuscript received 15 May 2022; revised 5 January 2023; accepted 8 May 2023. *(Corresponding author: Dmitry Dagaev.)*

Akash Adhikari is with the Indian Institute of Technology (ISM), Dhanbad 826004, India (e-mail: rajaadhikari23@gmail.com).

Stanislav Anatolyev is with the CERGE-EI, 111 21 Prague, Czech Republic, and also with the New Economic School, 121353 Moscow, Russia (e-mail: stanislav.anatolyev@cerge-ei.cz).

Dmitry Dagaev is with the HSE University, 101000 Moscow, Russia (e-mail: ddagaev@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TG.2023.3275710.

Digital Object Identifier 10.1109/TG.2023.3275710

[1]See http://www.chessgames.com/chessstats.html. Retrieved April 1, 2021.

To the best of our knowledge, the effect of previous mistakes and their severity on the probability of making future mistakes of different severities has not been econometrically evaluated before for the game of chess. The most relevant topic that attracted a lot of researchers' attention is the so called "Hot Hand" phenomenon. In their classical article, Gilovich, Vallone, and Tversky [20] tested a popular belief that basketball players experience streak shooting, so that the probability of scoring a goal increases after another successful shot. The authors disproved the hot hand hypothesis and attributed the myth to the wrong perception of chance. Despite the negative result, the work started a series of articles on the existence of hot hand in various environments. Most of the empirical articles supported the conclusions of [20] for the game of basketball and some other sports (see, for example, [21], [22], [23]). Seemingly less frequently analyzed concept of a cold hand, the existence of disproportionately often streak failures, is closely connected to the concept of tilting. The principal difference between basketball and chess is that shots in basketball can be considered as an iterated exercise (especially in the case of free throws), whereas each position in chess is unique. Therefore, we avoid to make a clear link between making moves in chess and iterated throws in basketball. Instead, we prefer to use the concept of tilt which allows to carry over the negative emotions from realizing of making a mistake to the subsequent moves.

The rest of this article is organized as follows. Section II describes the data. Section III discusses the econometric model and empirical strategy. Section IV contains the results. Finally, Section V concludes this article.

II. Data

We collected all games of the main Tata Steel Chess Tournament that takes place annually in Wijk aan Zee (The Netherlands). In total, 885 games were played between 2011 (the first year when the tournament in Wijk aan Zee was named Tata Steel Chess Tournament) and 2020 (the last year in our database). The tournament is organized in a round-robin format—each player plays against each other player once. In 2014, there were 12 players and 66 games in total, whereas in all other nine years there were 14 players and 91 games. Notation for the games is available in Portable Game Notation (PGN) format, which is a standard designed for representing chess game data. In chess, FIDE[2] rating is used to evaluate a player's relative skill level. It is based on the Elo rating principles proposed by Arpad Elo. When two chess players who already have the rating play each other, a certain number of rating points is transferred from the loser to the winner; in case of a draw, points are transferred from the higher rated player to the lower rated player. The exact number of points transferred from one player to the opponent is a function of their ratings and the outcome of a game. For each game in our dataset, we collected the FIDE ratings of both opponents at the time when the game was played. All games were played by highly rated professionals whose FIDE rating ranged from 2603 to 2872.

In order to evaluate chess positions, we use the open source chess engine Stockfish 12 [24], which was also used in some other behavioral and socio-economic performance-related studies [4], [25], [26], [27]. As of 2022, Stockfish is widely regarded as the strongest open source computer chess program.[3] Historically, Stockfish evaluated a position by looking through a game tree starting at the current position as deeply as the time limit allows. A limited number of apparently good lines are looked through more deeply than others. In September 2020, a new version Stockfish 12 was released, and it was announced that Stockfish had absorbed a neural network project.[4] We manually set the number of good lines to 9 and the time limit to 7 seconds per move, which leads to the search depth of at least 17 half moves.[5] As the time limit expires, Stockfish suggests the best possible line and a numerical evaluation of the position corresponding to that line. If at some position one of the opponents has a guaranteed win by checkmate, Stockfish provides "white/black mates in k moves" instead of a numerical value. The evaluator takes into account various factors: existence of a forced checkmate, material advantage, positional weaknesses (isolated pawns, doubled pawns, etc.), and positional advantages (two bishops, rooks on open lines, etc.). All scores are normalized so that an extra pawn for white leads to the score of +1.00 given all else equal. Since chess is an antagonistic game, the score of some position for black is simply the score of that position for white with the opposite sign.

The website[6] is a popular online chess platform that uses the Stockfish 12 engine. We uploaded our PGN files to[7] one by one in order to obtain Stockfish evaluations. The data were extracted as `.txt` files by web scraping using javascript. For each position, we collected a numerical evaluation suggested by Stockfish. Now, for each move $m_g = 1, 2, \ldots$ played by one of the players in game g, we define the so-called *centipawn loss* variable cl_{g,m_g} that shows the quality of this move

$$\mathrm{cl}_{g,m_g} = \begin{cases} ea_{g,m_g} - eb_{g,m_g} & \text{for white} \\ -(ea_{g,m_g} - eb_{g,m_g}) & \text{for black} \end{cases} \tag{1}$$

where, ea_{g,m_g} is the evaluation of the position after the move m_g and eb_{g,m_g} is the evaluation of the position before the move m_g. Similar metrics for quality of moves are used in the literature [4], [26]. If a player chooses the best possible move, cl_{g,m_g} is expected to be 0 (the optimal line after the move is the same as before the move). If a player fails to choose the best move, cl_{g,m_g} is expected to be negative. However, note that after a move is made, Stockfish starts its analysis from the next opponent's move. Due to this discrepancy in the search depths before and after a move, evaluation of the position can be changed (in both directions) after the move even if the best move was played. It explains why there do exist moves with positive value of cl_{g,m_g}. Finally, we refer to [27] where it was shown that the engine set at the search depth of 17 half-moves chooses a move with an average error of less than 3 centipawns (or 0.03). This can also be interpreted as an upper bound for the average evaluation error of a position due to the limited search time. We think that such error is acceptable for the purposes of this research.

We also made the following adjustments to the dataset. First of all, we excluded from the dataset the first five moves of each game. This step removes most spurious evaluations associated with the white's first-mover advantage and forms minimal play history for the current game. However, these first five moves are still used to generate "lagged" variables related to previous play for moves 6 through 10 (see the next section). Next, in chess, there are many ways to win a decided game. Usually, chess players prefer to use a safe one. For example, reduction to a theoretically winning position would be preferred to a fast but rather complex combination, even if the safe way would be much longer. From the Stockfish perspective, using safe ways is sometimes interpreted as a mistake or even as a blunder. In order to account for this, we excluded decided positions from our dataset. Particularly, suppose that move p_g is the first move of game g such that the absolute value of evaluation is higher than 5.00 (such advantage corresponds to an extra rook, and

[2]International Chess Federation.

[3][Online]. Available: https://ccrl.chessdom.com/. Retrieved January 1, 2023.

[4][Online]. Available: https://www.chess.com/terms/stockfish-chess-engine. Retrieved February 23, 2021.

[5]A half-move is a move of White or a move of Black.

[6][Online]. Available: www.chess.com

[7][Online]. Available: www.chess.com

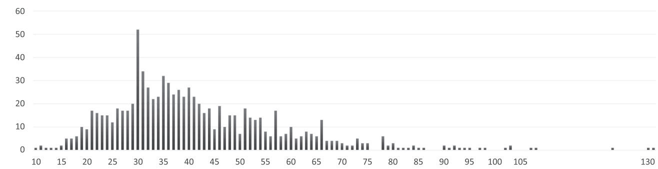

Fig. 1. Distribution of the number of moves in games from the sample. The number of moves is represented on *X*-axis, the number of games in the dataset with a particular number of moves is represented on *Y*-axis.

TABLE I
CHARACTERISTICS OF MISTAKES

Level	Lower threshold	Upper threshold	Mistake description	Fraction of moves
0	0.0	—	No mistake	29.88%
1	−0.5	−0.0	An inaccuracy	57.36%
2	−1.0	−0.5	A small mistake	9.40%
3	−1.5	−1.0	A moderate mistake	1.97%
4	−2.0	−1.5	A severe mistake	0.70%
5	—	−2.0	A blunder	0.69%

for a strong chess player, it is more than enough for a win, see [27] for statistics), or a checkmate. We have deleted all observations from this game starting from this move, so move $p_g - 1$ will be the last move of this game. As a result of these adjustments, the moves in game g are indexed now by $m_g = 6, 7, \ldots, p_g - 1$. Finally, for all games in our dataset we excluded the last half-move due to technical issues related to extraction of the position score before switching to another game. For the whole sample of 885 games, this leads to 64 404 moves and hence observations. Fig. 1 shows a distribution of these moves across games.

The database contains only games from very strong international level chess players. We make a plausible assumption that the players do not intentionally choose suboptimal moves because opponents can exploit even small suboptimality at this level.

III. ECONOMETRIC MODEL

Let

$$-\infty = b_{B+1} < b_B < \ldots < b_2 < b_1 = 0 < b_0 = \infty \quad (2)$$

be the score thresholds that define levels of mistakes of different severity. The variable representing the mistake of severity level $j = 1, \ldots, B$ made at move m_g in game g is equal to

$$\mathbb{I}_{j,g,m_g} = \mathbb{I}\left\{b_{j+1} \leq \mathrm{cl}_{g,m_g} < b_j\right\} \quad (3)$$

where, $\mathbb{I}\{\cdot\}$ is an indicator function. By convention, values $j = 0$ and $b_0 = \infty$ correspond to no mistake made (mistake of level 0); in this case $\mathrm{cl}_{g,m_g} \geq 0$.

As a practical matter, we consider $B = 5$ levels of mistakes of the following severity levels. Table I shows their cutoffs, characterizations, and in what fraction of moves these mistakes are made in the database we consider.

Our econometric model is based on the ordered multiple choice regression where the left-hand side variable is a type of a mistake made (or not made) after each move, and the right-hand side variables describe the quality of the same player's previous play in the game, in addition to a number of covariates that characterize the player and the game. Specifically, we define a latent variable pm_{g,m_g}, which we call a propensity to misplay

$$\mathrm{pm}_{g,m_g} = z'_{g,m_g}\gamma + x'_{g,m_g}\beta + \alpha_g + \varepsilon_{g,m_g}. \quad (4)$$

Here, the vector of covariates z_{g,m_g} contains controls specific to move m_g in game g or to game g alone, and not directly related to the player's past performance in the game, while x_{g,m_g} contains predictors that describe the previous play in game g before move m_g. The next component, α_g, is a game-specific move-independent random effect. Finally, ε_{g,m_g} is an idiosyncratic random component not explained by the included regressors.

A mistake of type $j = 0, 1, \ldots, B$ (recall that 0 stands for no mistake, and an increasing j corresponds to more severe mistakes) occurs when the propensity to misplay pm_{g,m_g} falls in the region $[A_{j+1}, A_j)$, where $A_{B+1} = -\infty$, $A_0 = \infty$, and A_j, $j = 1, \ldots, B$, are unknown cutoffs. Under the assumption that α_g and ε_{g,m_g} are normally distributed independently of included regressors, the mistake of type $j = 0, \ldots, B$ has conditional probability

$$\Pr\left\{b_{j+1} \leq \mathrm{cl}_{g,m_g} < b_j\right\} = \quad (5)$$

$$= \Phi(A_j - z'_{g,m_g}\gamma - x'_{g,m_g}\beta) - \Phi(A_{j+1} - z'_{g,m_g}\gamma - x'_{g,m_g}\beta).$$

where Φ is a standard normal cumulative distribution function. Such an ordered probit model means that the probability of a mistake of level j depends on the characteristics of the player, of the move, of the game, and of the previous play.

We include the following variables to z_{g,m_g}, in addition to a constant.

1) elo_g, an Elo rating of the player making move m_g in game g.
2) ev_{g,m_g}, an evaluation before the move

$$\mathrm{ev}_{g,m_g} = \begin{cases} eb_{g,m_g} & \text{for white} \\ -eb_{g,m_g} & \text{for black.} \end{cases} \quad (6)$$

3) taken_{g,m_g}, a number of pieces gone from the board before move m_g in game g.
4) white_{g,m_g}, an indicator that the move m_g in game g is made by white.

The variable elo_g depends only on parameters of the player making the move in the game, and is meant to capture the direct effect of the players' strength on the sequential pattern of mistakes: a weaker player's more serious mistake may increase the probability of this player's next more serious mistake. The variable taken_{g,m_g} is a proxy for a stage of the game, [8] which may affect tilt formation. The variable white_{g,m_g} is meant to capture the heterogeneity from the color of pieces, as this may affect the psychological state and strategy of the player.

While it is interesting to see the effects of the abovementioned covariates, our primary interest is analyzing the effects of the previous play. Because the previous mistakes may be characterized by many different variables, we adopt simple empirical strategies to select the most influential predictors from a limited set of possibilities. Specifically, the list of candidates to include in x_{g,m_g} is:

1) $\mathbb{I}_{j,g,m_g-\ell}$, the fact of making a mistake of jth severity at move $m_g - \ell$ in game g, for $j = 1, \ldots, B$ and $\ell = 1, \ldots, L$;
2) ab^{-}_{g,m_g}, a historical average of one's mistakes of any level, in game g before move m_g during L previous moves

$$ab^{-}_{g,m_g} = \frac{1}{L}\sum_{\ell=1}^{L}\left|\mathrm{cl}_{g,m_g-\ell}\right|\sum_{j=1}^{B}\mathbb{I}_{j,g,m_g-\ell}; \quad (7)$$

[8] A chess game is characterized by three stages: Début (beginning), Mittelspiel (middle game), and Endspiel (endgame). All stages have their own specifics and gradually transform one into another. However, by which move the stages transit from one to another is not predetermined but depends on the style of a particular game.

TABLE II
ORDERED PROBIT REGRESSION, COEFFICIENTS ON PREDICTORS BASED ON PREVIOUS PLAY

			Mistake indicators		
lag	$\mathbb{I}_1$	$\mathbb{I}_2$	$\mathbb{I}_3$	$\mathbb{I}_4$	$\mathbb{I}_5$
1	0.200*** (0.014)	0.043** (0.018)		0.137* (0.077)	−0.242** (0.119)
2	0.126*** (0.013)		0.073* (0.038)		−0.407*** (0.104)
3	0.119*** (0.013)		0.085* (0.044)		−0.401*** (0.125)
4	0.101*** (0.013)				−0.369*** (0.118)
5	0.061*** (0.013)				−0.504*** (0.117)
			Aggregate mistakes		
		ab^-		xb^-	
		0.562*** (0.117)		0.143*** (0.042)	

Robust clustered standard errors in parentheses; $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

3) xb^-_{g,m_g}, one's move with worst centipawn loss in game g before move m_g during L previous moves

$$xb^-_{g,m_g} = \max_{\ell=1,\dots,L}(-\mathrm{cl}_{g,m_g-\ell}). \tag{8}$$

The first type of predictors is $\mathbb{I}_{j,g,m_g-\ell}$, the indicator of a mistake of level j in one of L most recent moves. These indicators are meant to absorb the short term effects during a recent play. In total, we have BL predictors of such "individual" move-to-move type. The other two predictors are of "aggregate" type, as they index how, on average or in extreme terms, erroneous the play have been up until the current move is made. These two variables are meant to absorb the long term effects during the whole play in a game. The variable ab^-_{g,m_g} indicates how large the errors have been on average, and is meant to capture the overall psychological state of a player based on previous mistakes made. The variable xb^-_{g,m_g} indicates how big the maximal mistake in recent previous play has been, and is meant to capture the emotional distress caused by this mistake on the following play. As a practical matter, we set L to 5, which is arguably sufficient to capture the psychological state resulting from a recent play. In total, when $B = 5$ and $L = 5$, there are 27 mistake-related predictors.

We now address a few econometric issues and how we handle them. We perform quasimaximum likelihood estimation of the ordered probit model. The influence of "serial" correlation across moves within the same game on the asymptotic variance of parameter estimates is taken care of by clustering by the game (see, e.g., [28]). The within-game "color effect" in each game is automatically taken care of by including the indicator of white among the covariates. To select only a few from the list of "previous mistakes" predictors, we implement a general-to-specific stepwise selection procedure [29]. Specifically, we fix the list of included covariates z_{g,m_g}, and set the tolerance level to statistical significance of selected predictors from the abovementioned list of potential x_{g,m_g}'s to 10%, i.e., we stop removing predictors when none of those that are left has a coefficient with a p-value exceeding 10%.[9]

IV. Empirical Results

We now look at the pattern of how the quality of previous play affects the propensity to make errors in further play. Table II reports the results of running the ordered probit regression on the included covariates and significant predictors selected, as described in the previous section.[10] The coefficients in the table represent the marginal effects of each predictor on the latent propensity to misplay, and are eventually related to the probabilities of making mistakes.[11] In particular, a positive sign of a covariate/predictor implies its positive effect on the propensity to misplay and hence a negative effect on a quality of play. Conversely, a negative sign of a covariate/predictor implies its positive effect on a quality of play. The figures in the "mistake indicators" subpanel are regression coefficients for the short term predictors—the indicators $\mathbb{I}_{j,g,m_g-\ell}$ corresponding to the fact of making a mistake of jth severity for "lag" $\ell = 1, \dots, 5$. Analogously, the figures in the "aggregate mistakes" subpanel are regression coefficients for the long term predictors—a historical average of mistakes ab^-_{g,m_g} and historical maximal mistake xb^-_{g,m_g}.

First, let us look at the effects of selected lagged mistake indicators on the propensity to misplay. It is striking that different levels of mistake severity may make impact of a different strength and even a different sign. While the mistakes of moderate severity are statistically less significant, the small inaccuracies and big blunders are statistically most significant for all five included lags. They also tend to have more pronounced numerical effects but those effects are of opposite signs. Small inaccuracies, especially their most recent occurrences, increase the propensity to misplay, provoking the tilt. The same is true, although less strongly,[12] for small, moderate and severe mistakes; however, their effects seem to be shorter lived.

In contrast, the estimates coefficients of blunder indicators are starkly different: all negative and relatively large in absolute value. This brings a conclusion that, in reaction to their blunders, players tend instead to discipline their play. Moreover, in addition to its bigger size, this effect turns out to be longer lived than the tilting effect of less severe mistakes.

Next, the last two columns of Table II show the effect of the two aggregated measures of previous erroneous play during the last five moves, which may cause overall emotional distress. Notice that these measures are statistically significant even though all the individual mistake indicators for the same five periods are already included in the regression. Hence, there is strong predictive information in the average and maximal mistakes made in the previous play, on top of occurrences of each mistake. Both effects are positive for the propensity of further misplay, and strongly confirm the presence of tilt.

It is also interesting to examine how the quality of play is influenced by the characteristics of the game, the moves, and the players. Table III reports the estimates on included covariates except for a constant (we again remove regressors' indexes to reduce clutter). All of the coefficients of included covariates are strongly statistically significant and have intuitively sensible signs. One can see that a player's Elo rating has a positive, although small in value, effect on the quality of play, which is intuitive. The current evaluation positively influences the propensity to misplay, meaning that a player is more likely to become careless and possibly reckless in a better position. Next, the proxy for the stage of the game perhaps affects the quality of play—the tree of a subgame becomes less deep closer to an endgame. Finally, being white has a favorable effect on preventing mistakes.

[9]Our preference for 10% is motivated by a desire to end up with a more liberal post-selection specification so that not to miss important predictors.

[10]In Table II, we intentionally remove predictors' indexes to reduce clutter.

[11]A reader should keep in mind that the absolute values of the marginal effects do not carry much information, because the composite error is normalized to have unit variance for the purpose of identification. Thus, it is their values relative to each other, taking the predictor scales into account when those scales are different, that is meaningful and interpretable.

[12]Note that with the significance threshold of 5% for the stepwise selection procedure, the $\mathbb{I}_3$ and $\mathbb{I}_4$ predictors would not be selected at all, with no noticeable changes in the rest of the results.

TABLE III
ORDERED PROBIT REGRESSION, COEFFICIENTS ON INCLUDED COVARIATES

Covariate	Coefficient estimate, $\times 10^{-2}$	Covariate scale
elo	-0.0726^{***} (0.0096)	54.4
ev	3.71^{***} (0.52)	1.20
taken	-1.45^{***} (0.11)	7.15
white	-2.84^{***} (0.87)	0.50

Robust clustered standard errors below point estimates; $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. Last column lists standard deviations.

We would like to emphasize that even though all these covariates are strongly statistically significant, their numerical effects (accounting for variables' scales; see standard deviations in the last column of Table III) are appreciably smaller than those of the indicators or aggregate measures of previous play documented in Table II. Among the four covariates, the variable "taken" has the greatest impact on the quality of play, given its biggest product of the coefficient and variable's standard deviation among all, the variable "ev" coming the second.

Even though the presented regression results give a strong evidence of influence of mistakes on the quality of further play, we perform a formal test for inclusion of all $BL + 2$ previous mistake related predictors. The Wald test statistic for their joint significance equals 1062, with an essentially zero p-value relative to the $\chi^2_{(27)}$ distribution. A similar outcome results if we jointly test the exclusion restrictions for the included "previous mistakes" predictors only.

Moreover, it is interesting to compare the measures of regression fit from the ordered probit models with and without the previous mistake related predictors. The difference will show a relative contribution of the mistake-related predictors to the explanatory power of covariates. For the full model with all predictors included, the pseudo-R^2 equals 3.11%, and in the full model with only stepwise-selected predictors, the pseudo-R^2 equals 3.10%, an almost identical figure. At the same time, the ordered probit model with all the predictors excluded and only the covariates left, the pseudo-R^2 equals 0.93%. This shows that previous mistakes have a much larger role in determining the quality of further play than explanatory variables from Table III, at least among the top players.

V. CONCLUSION

In this article, we have uncovered sequential patterns of mistakes of human players in the game of chess. We have found clear evidence that small inaccuracies lead to less accurate play in future; more severe mistakes have a weaker effect on the quality of play in the same direction, while blunders tend to discipline players. Inaccuracies and blunders have more long-lived effect than mistakes of moderate size do.

One should have in mind that our database contains games played by strong chess players. The pattern could be different for lower ranked players due to their lower ability to find best moves. On the one hand, higher variance of their quality of play could dominate psychological effects. On the other hand, lower ranking can potentially incorporate information about the resistance to tilt. Therefore, a further careful analysis is required for that cohort of players.

We acknowledge that one should be careful in interpreting the findings of this study. Although tilt seems to be the most obvious explanation for the fact that some types of mistakes increase the probability of a new mistake, our methodology does not allow to exclude other possible explanations not related to the psychological state of mind. Alternative theories include the changing attitude toward risk (chess players may look for complications in worse positions) and peculiarities of the Stockfish evaluation algorithm (the difference between the scores +4 and +5 in the decided positions can be due to the arguments that are not taken into account by human players). We hope that future research will allow to differentiate between these theories.

ACKNOWLEDGMENT

The authors would like to thank Alena Skolkova for excellent research assistance and Petr Parshakov for helpful comments. Dmitry Dagaev gratefully acknowledges support from the Basic Research Program of the National Research University Higher School of Economics.

REFERENCES

[1] E. Zermelo, "Uber eine anwendung der mengenlehre auf die theorie des schachspiels," in *Proc. 5th Int. Congr. Mathematicians*, 1913, vol. 2, pp. 501–504.
[2] L. Kalmár, "Zur theorie der abstrakten spiele," *Acta Universitatis Szegediensis/Sectio Scientiarum Mathematicarum*, vol. 4, pp. 65–85, 1928.
[3] C. Ewerhart, "Backward induction and the game-theoretic analysis of chess," *Games Econ. Behav.*, vol. 39, no. 2, pp. 206–214, 2002.
[4] K. W. Regan, T. Biswas, and J. Zhou, "Human and computer preferences at chess," in *Proc. Workshops 20th AAAI Conf. Artif. Intell.*, 2014, pp. 79–84.
[5] K. W. Regan, B. Macieja, and G. M. Haworth, "Understanding distributions of chess performances," *Adv. Comput. Games*, vol. 13, pp. 230–243, 2011.
[6] J. Schaeffer et al., "Checkers is solved," *Science*, vol. 317, no. 5844, pp. 1518–1522, 2007.
[7] S. Barrault, A. Untas, and I. Varescon, "Special features of poker," *Int. Gambling Stud.*, vol. 14, no. 3, pp. 492–504, 2014.
[8] B. R. Browne, "Going on tilt: Frequent poker players and control," *J. Gambling Behav.*, vol. 5, no. 1, pp. 3–21, 1989.
[9] J. Palomäki, M. Laakasuo, and M. Salmela, "This is just so unfair!': A qualitative analysis of loss-induced emotions and tilting in on-line poker," *Int. Gambling Stud.*, vol. 13, no. 2, pp. 255–270, 2013.
[10] J. Palomäki, M. Laakasuo, and M. Salmela, "Losing more by losing it: Poker experience, sensitivity to losses and tilting severity," *J. Gambling Stud.*, vol. 30, no. 1, pp. 187–200, 2014.
[11] T. Toneatto, "Cognitive psychopathology of problem gambling," *Substance Use Misuse*, vol. 34, no. 11, pp. 1593–1604, 1999.
[12] G. Smith, M. Levere, and R. Kurtzman, "Poker player behavior after big wins and big losses," *Manage. Sci.*, vol. 55, no. 9, pp. 1547–1555, 2009.
[13] A. Moreau, H. Chabrol, and E. Chauchard, "Psychopathology of online poker players: Review of literature," *J. Behav. Addictions*, vol. 5, no. 2, pp. 155–168, 2016.
[14] G. Meyer, von Meduna, M. T. Brosowski, and T. Hayer, "Is poker a game of skill or chance? A quasi-experimental study," *J. Gambling Stud.*, vol. 29, no. 3, pp. 535–550, 2013.
[15] R. H. Grabner, E. Stern, and A. C. Neubauer, "Individual differences in chess expertise: A psychometric investigation," *Acta Psychologica*, vol. 124, no. 3, pp. 398–420, 2007.
[16] N. Charness, "Search in chess: Age and skill differences," *J. Exp. Psychol.: Hum. Percep. Perform.*, vol. 7, no. 2, pp. 467–476, 1981.
[17] C. F. Chabris and E. S. Hearst, "Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors," *Cogn. Sci.*, vol. 27, no. 4, pp. 637–648, 2003.
[18] B. D. Burns, "The effects of speed on skilled chess performance," *Psychol. Sci.*, vol. 15, no. 4, pp. 442–447, 2004.
[19] J. H. Moxley, K. A. Ericsson, N. Charness, and R. T. Krampe, "The role of intuition and deliberative thinking in experts' superior tactical decision-making," *Cognition*, vol. 124, no. 1, pp. 72–78, 2012.
[20] T. Gilovich, R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cogn. Psychol.*, vol. 17, no. 3, pp. 295–314, 1985.
[21] M. Bar-Eli, S. Avugos, and M. Raab, "Twenty years of "hot hand" research: Review and critique," *Psychol. Sport Exercise*, vol. 7, no. 6, pp. 525–553, 2006.
[22] J. J. Koehler and C. A. Conley, "The "hot hand" myth in professional basketball," *J. Sport Exercise Psychol.*, vol. 25, no. 2, pp. 253–259, 2003.

[23] A. Tversky and T. Gilovich, “The cold facts about the “hot hand” in basketball,” *Chance*, vol. 2, no. 1, pp. 16–21, 1989.

[24] T. Romstad, M. Costalba, and J. Kiiski, “Stockfish: A strong open source chess engine.” [Online]. Available: https://stockfishchess.org

[25] S. Künn, C. Seel, and D. Zegners, “Cognitive performance in remote work: Evidence from professional chess,” *Econ. J.*, vol. 132, no. 643, pp. 1218–1232, 2022.

[26] D. J. Barnes and J. Hernandez-Castro, “On the limits of engine analysis for cheating detection in chess,” *Comput. Secur.*, vol. 48, pp. 58–73, 2015.

[27] T. Biswas and K. Regan, “Measuring level-k reasoning, satisficing, and human error in game-play data,” in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl.*, 2015, pp. 941–947.

[28] A. C. Cameron and D. L. Miller, “A practitioner’s guide to cluster-robust inference,” *J. Hum. Resour.*, vol. 50, no. 2, pp. 317–372, 2015.

[29] J. Campos, N. R. Ericsson, and D. F. Hendry, Eds., *General-to-Specific Modelling*. Cheltenham, U.K.: Edward Elgar Publishing, 2005.