RESEARCH ARTICLE

AI-powered mechanisms as judges: Breaking ties in chess

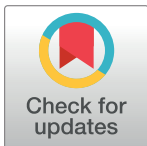**Nejat Anbarci** [ID] [1 open], **Mehmet S. Ismail** [ID] [2 open] *

1 Department of Economics, Durham University, Durham, United Kingdom, **2** Department of Political Economy, King's College London, London, United Kingdom

open All these authors are contributed equally to this work.

* mehmet.ismail@kcl.ac.uk

Abstract

Recently, Artificial Intelligence (AI) technology use has been rising in sports to reach decisions of various complexity. At a relatively low complexity level, for example, major tennis tournaments replaced human line judges with Hawk-Eye Live technology to reduce staff during the COVID-19 pandemic. AI is now ready to move beyond such mundane tasks, however. A case in point and a perfect application ground is chess. To reduce the growing incidence of ties, many elite tournaments have resorted to fast chess tiebreakers. However, these tiebreakers significantly reduce the quality of games. To address this issue, we propose a novel AI-driven method for an objective tiebreaking mechanism. This method evaluates the quality of players' moves by comparing them to the optimal moves suggested by powerful chess engines. If there is a tie, the player with the higher quality measure wins the tiebreak. This approach not only enhances the fairness and integrity of the competition but also maintains the game's high standards. To show the effectiveness of our method, we apply it to a dataset comprising approximately 25,000 grandmaster moves from World Chess Championship matches spanning from 1910 to 2018, using Stockfish 16, a leading chess AI, for analysis.

1 Introduction

The use of Artificial Intelligence (AI) technology in sports, to reach decisions of various complexity, has been on the rise recently. At a relatively low complexity level, for example, major tennis tournaments replaced human line judges with Hawk-Eye Live technology to reduce staff during the COVID-19 pandemic. Also, more than a decade ago, football began using Goal-line technology to assess when the ball has completely crossed the goal line. These are examples of mechanical AI systems requiring the assistance of electronic devices to determine the precise location of balls impartially and fairly, thus minimizing, if not eliminating, any controversy.

A major question now is whether AI could move beyond such rudimentary tasks in sports. A case in point and a perfect application ground is chess for two complementary reasons. On the one hand, advanced AI systems, including Stockfish, AlphaZero, and MuZero, have already
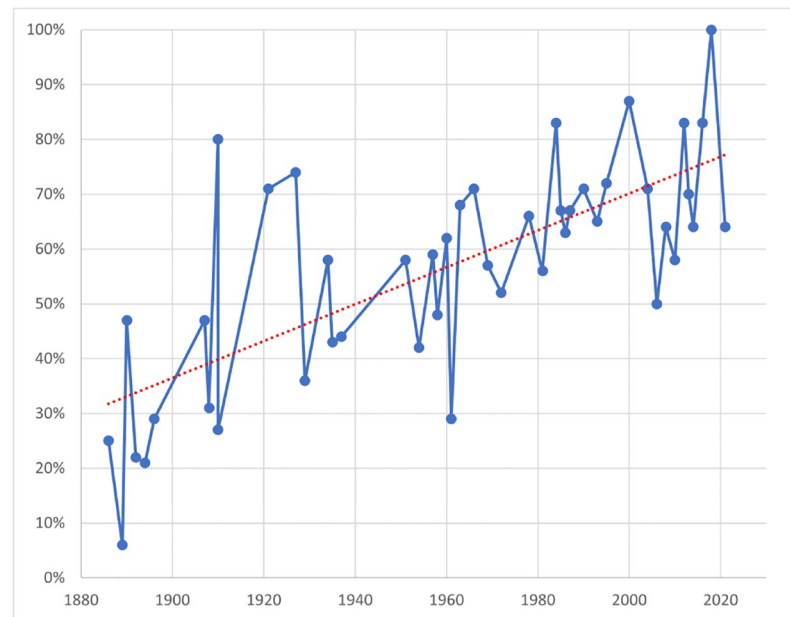
Fig 1. Percentage of draws in World Chess Championship matches 1886–2021.

been implemented in chess [1–3]; further, the superiority of top chess engines has been widely acknowledged ever since IBM's Deep Blue defeated former world chess champion Garry Kasparov in 1997 [4]. On the other hand, despite its current popularity all around the world, chess is very much hampered by the growing incidence of draws, especially in the world championships, as Fig 1 illustrates. In response to this draw problem, elite chess tournaments—like other sports competitions [5–11]—resorted to the tiebreakers. The most common final tiebreaker is the so-called Armageddon game, where White has more time (e.g., five minutes) to think on the clock than Black (e.g., four minutes), but Black wins in the event of a draw. However, this format sparks controversy among elite players and chess aficionados alike:

> "Armageddon is a chess penalty shoot-out, a controversial format intended to prevent draws and to stimulate interesting play. It can also lead to chaotic scrambles where pieces fall off the board, players bang down their moves and hammer the clocks, and fractions of a second decide the result"

> (Leonard Barden, *The Guardian* [12]).

In this paper, we introduce an AI-based tiebreaking method designed to address the prevalence of draws in chess, particularly evident in World Chess Championship matches. This novel and practical approach, applied to a dataset of about 25,000 moves from grandmasters in matches spanning from 1910 to 2018, employs Stockfish 16, a leading chess AI system, to analyze each move. By assessing the difference between a player's actual move and the engine's optimal move, our method identifies the player with superior play quality in tied matches, thereby effectively resolving all such ties.

The "draw problem" that we address in this paper has a long history. Neither chess aficionados nor elite players appear to enjoy the increasing number of draws in chess tournaments.

Table 1. An example of classical vs. Fast chess time-controls.

Time-control	Time per player
Classical	90 min. + 30 sec. per move
Rapid	15 min. + 10 sec. per move
Blitz	5 min. + 2 sec. per move
Armageddon	White: 5 min., Black: 4 min. (and draw odds)

The five-time world champion, Magnus Carlsen, appears to be dissatisfied as well. "Personally, I'm hoping that this time there will be fewer draws than there have been in the last few times, because basically I have not led a world championship match in classical chess since 2014" [13].

The 2018 world championship tournament, for instance, ended with 12 consecutive draws. The world champion was then determined by a series of "rapid" games, whereby players compete under significantly shorter time-control than the classical games (see Table 1). If the games in the tiebreaks did not determine the winner, then a final game called Armageddon would have been played to determine the winner. Clearly, compared to classical chess games, there is no doubt that the fast-paced rapid, blitz and Armageddon games lower the quality of chess played; the latter also raises questions about its fairness because it treats players asymmetrically. (For fairness considerations in sports, apart from the works we have already mentioned, see, e.g., [14, 15]).

More broadly, our paper relates to the literature on fairness of sports rules. Similar to the first-mover advantage in chess, there is generally a first-kicker advantage in soccer penalty shootouts [6, 16]. Several authors have proposed rule changes to restore (ex-post) fairness in penalty shootouts [5, 8, 17–19].

The rest of the paper is organized as follows. In Section 2, we give a brief summary of the empirical results. In Section 3, we provide the main framework of the paper, discuss data collection, and describe the computation method. The limitations of our paper are discussed in Section 4; Section 5 concludes the paper.

2 The empirical summary and the computation method

In this paper, as mentioned, we introduce a novel and practicable AI-based tiebreaking method and show that it essentially eliminates the tie problem in chess. We apply this method to a dataset comprising approximately 25,000 moves made by grandmasters in World Chess Championship matches from 1910 to 2018. Notably, nine (about 18%) out of fifty championship matches ended in a tie, with the incidence of ties increasing to 42% in the past two decades.

In a nutshell, our analysis involved examining each move using Stockfish 16, one of the most advanced chess AI systems. We measured the discrepancy between the evaluations of a player's actual move and the optimal move as determined by the chess engine. In situations where matches ended in a tie, the player demonstrating a higher quality of play, as indicated by our analysis, is declared the winner of the tiebreak.

2.1 An AI-based tiebreaking mechanism

In the event of a win, it is straightforward to deduce that the winner played a higher quality chess than the loser. In the event of a tie in a tournament, however, it is more difficult to assert that the two players' performances were comparable, let alone identical. With the

Table 2. Summary statistics: The player with the lower cumulative (C) TPLV is declared the tiebreak winner. For details, see Fig 4. Δ represents the numerical difference and % represents the percentage difference (with respect to the lower value) between the two C-TPLVs.

Year	Players	# Games	C-TPLV 1	C-TPLV 2	Δ	%
1910	Schlechter-Lasker	10	51.3	151.5	-100.2	195.3
1951	Botvinnik-Bronstein	24	245.4	145.1	100.3	69.1
1954	Botvinnik-Smyslov	24	115.2	119.1	-3.9	3.4
1987	Kasparov-Karpov	24	93.9	88.8	5.1	5.7
2004	Leko-Kramnik	14	117.1	22.4	94.7	422.8
2006	Topalov-Kramnik	12	41.2	229.8	-188.6	457.8
2012	Anand-Gelfand	12	42.3	43.9	-1.6	3.8
2016	Karjakin-Carlsen	12	41.9	40.6	1.3	3.2
2018	Carlsen-Caruana	12	27.5	27.8	-0.3	1.1

advancements in chess AIs, their differences in quality can be quantified. Average centipawn loss is a known metric for evaluating the quality of a player's moves in a game where the unit of measurement is 1/100th of a pawn. (See, e.g., https://www.chessprogramming.org/Centipawns).

We instead use a more intuitive modification of that metric, which we term the "total pawn loss," because (i) even chess enthusiasts do not seem to find the average centipawn loss straightforward, based on our own anecdotal observations, and (ii) it can be manipulated. The second point is easy to confirm because a player can intentionally extend the game, for example, in a theoretically drawn position, thereby 'artificially' decreasing his or her average centipawn loss.

We define total pawn loss as follows. First, at each position in the game, the difference between the evaluations of a player's actual move and the "best move" as deemed by a chess engine is calculated. Then, the total pawn loss value (TPLV) for each player is simply the equivalent of the total number of "pawn-units" the player has lost during a chess game as a result of errors. TPLV can be calculated from the average centipawn loss and the number of moves made in a game (see Section 3). If the TPLV is equal to zero, then every move was perfect according to the chess engine used. The lower the TPLV the better the quality of play.

Table 2 shows the cumulative TPLVs in tied undisputed world chess championship matches in chess history. Notice that most recent matches, Anand-Gelfand (2012), Karjakin-Carlsen (2016), and Carlsen-Caruana (2018) were not only very close, as seen in the columns labeled Δ and %, but the overall TPLVs were quite low compared with the earlier matches. Take the Schlechter-Lasker (1910) match, for example. It is worth noting that while Schlechter led the match 5.0–4.0 before the last round, he took risks in the last game to avoid a draw. However, this strategy did not pay off, and he lost the game. Therefore, the match ended in a tie, which meant that Lasker retained his title. (The reader may wonder how TPLVs apply in decisive matches. In the 1972 Fischer-Spassky "Match of the Century," for example, Fischer significantly outperforms Spassky with a cumulative TPLV of 28.96 in decisive games compared to Spassky's 84.56. However, the cumulative TPLVs in the drawn games are similar, being 41.37 for Fischer and 41.29 for Spassky).

2.2 US Women's championship

Many elite tournaments, including the world chess championship as mentioned earlier, use Armageddon as a final tiebreaker. Most recently, the Armageddon tiebreaker was used in the 2022 US Women Chess Championship when Jennifer Yu and Irina Krush both tied for the

Table 3. TPLV vs Armageddon tiebreakers in the 2022 US Women Chess Championship. Irina Krush would have been the champion because she had significantly lower cumulative TPLV in the tournament.

Player	Cumulative TPLV	Average TPLV
GM Irina Krush	159.21	12.24
GM Jennifer Yu	188.62	14.50

https://doi.org/10.1371/journal.pone.0305905.t003

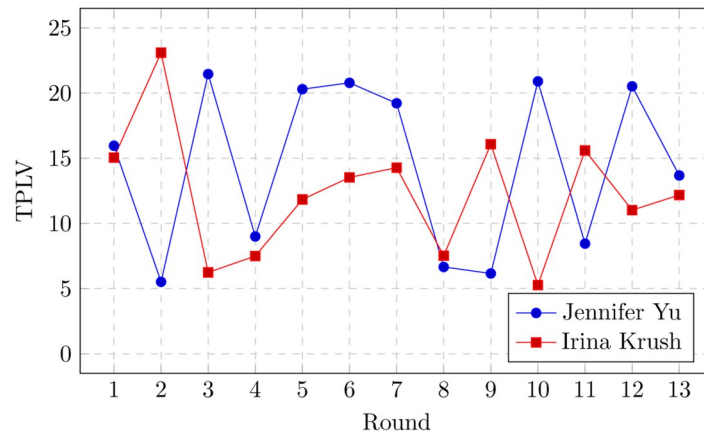**Fig 2. TPLVs of Irina Krush and Jennifer Yu in the 2022 US Women Chess Championship games.** Lower TPLV implies better play.

https://doi.org/10.1371/journal.pone.0305905.g002

first place, scoring each 9 points out of 13. Both players made big blunders in the Armageddon game; Irina Krush made an illegal move under time pressure and eventually lost the game and the championship.

Table 3 presents the cumulative TPLVs of Irina Krush and Jennifer Yu in the 2022 US Women Chess Championship. More specifically, Fig 2 illustrates their TPLVs for each round/game. According to our TPLV-based tiebreak method, Irina Krush would have been the US champion because she played a significantly better chess in the tournament according to Stockfish: Irina Krush's games were about two pawn-units better on average than Jennifer Yu's.

2.3 World Chess Championship 2018: Carlsen vs. Caruana

Magnus Carlsen offered a draw in a better position against Fabiano Caruana in the last classical game in their world championship match in 2018. This was because Carlsen was a much better player in rapid/blitz time-control than his opponent. Indeed, he won the rapid tiebreaks convincingly with a score of 3–0. Note that Carlsen made the best decision given the championship match, but due to the tiebreak system his decision was not the best (i.e., manipulation-proof) in the particular game. (This is related to the notion of strategyproofness that is often used in social choice, voting systems, mechanism design, and sports and competitive games; though, formal definition of strategyproofness varies depending on the context. For a selective literature, see, e.g., [20–33]).

Carlsen, of course, knew what he was doing when he offered a draw. During the post-game interview, he said "My approach was not to unbalance the position at that point" [34]. Indeed, in our opinion, Carlsen would not have offered a draw in their last game under the TPLV-based tiebreaking system because his TPLV was already lower in this game than Caruana's.

Table 4. Calculation of TPLV after Carlsen's draw offer to Caruana. Negative values imply that the chess engine deems Carlsen's position better as he has the black pieces.

	GM Magnus Carlsen (B)
TPLV before draw offer	5.2
Evaluation of best move	-1.0
Evaluation of draw offer	0.0
Pawn loss of draw offer	1 = -(-1–0)
TPLV after draw offer	6.2 = 5.2 + 1

https://doi.org/10.1371/journal.pone.0305905.t004

Carlsen's TPLV was 5.2 in the position before his draw offer, whereas Caruana's was 5.9. When Carlsen offered a draw, the evaluation of the position was about −1.0—i.e., Black is a pawn-unit better—according to Sesse, which is a strong computer running Stockfish. If Carlsen played the best move, then his evaluation would be about −1.0, which means he would be ahead about a pawn-unit. After the draw offer was accepted, the game ended in a draw and the evaluation of the position is obviously 0. As a result, Carlsen lost

$$1 \; = \; -(-1-0)$$

pawn-unit with his offer, as calculated in Table 4. Thus, his final TPLV is 6.2 = 5.2 + 1 as Table 4 shows. (For comparison, Carlsen's and Caruana's average centipawn losses in the entire match were 4.13 and 4.24, respectively). (A draw offer in that position would make Caruana the winner of the tiebreak under our method. As previously mentioned, we use TPLV to break a tie in matches; though, it could also be used as a tiebreaker in individual drawn games).

3 The definition, data collection, and computation

3.1 Definition

We focus on chess for the sake of clarity. For a formal definition of extensive-form games, see e.g. [35]. Let G denote the extensive-form game of chess under the standard International Chess Federation (FIDE) rules. Table 5 summarizes the relationship between the terminologies used in game theory and chess. In the chess terminology, a *chess game* is an alternating sequence of actions taken by White and Black from the beginning to the end of the game. In game theory, we call a chess game a *play* and denote it by $\bar{a} \in \bar{A}$, which is the finite set of all plays. (For a recent paper that emphasizes focusing on plays rather than strategies in game theory, see [36].) A chess *position* describes in detail the history of a chess game up to a certain point in the game. Formally, in game theory, a position is a *node*, denoted by $x \in X$, in the

Table 5. The terminology used in game theory and chess.

Game Theory	Notation	Chess
a game	G	the game of chess
a player	$i \in \{1,2\}$	White or Black
an action	$a_i \in A_i$	a move
a play	$\bar{a} \in \bar{A}$	a single chess game
a node	$x_j \in X$	a position
a tournament	$T(G)$	a tournament
AI	$v_i : X \to \mathbb{R}$	a chess engine
AI best-response	$a_i^*(x) \in \arg\max_{a_i(x) \in A_i(x)} v_i(a_i(x))$	best move

https://doi.org/10.1371/journal.pone.0305905.t005

extensive-form game G. A chess *move* is a choice of a player in a given position. A *draw offer* is a proposal of a player that leads to a draw if agreed by the opponent. If the opponent does not accept the offer, then the game continues as usual. Formally, an *action* of a player $i \in \{1, 2\}$ at a node $x \in X$ is denoted by $a_i(x) \in A_i(x)$, which is the set of all available actions of player i at node x. An action can be a move, a draw offer, or the acceptance or rejection of the offer. A *chess tournament*, denoted by $T(G)$, specifies the rules of a chess competition, e.g., the world chess championship where two players play a series of head-to-head chess games, and Swiss-system tournament where a player plays against a subset of other competitors.

We define a chess *AI* as a profile v of functions where for each player i, $v_i : X \to \mathbb{R}$. An AI yields an evaluation for every player and every position in a chess game. A chess engine is an AI which inputs a position and outputs the evaluation of the position for each player.

Let $a_i^* \in A_i$ be an action at a node x. The action is called an *AI best-response* if $a_i^*(x) \in \text{argmax}_{a_i(x) \in A_i(x)} v_i(a_i(x))$. In words, an AI best-response action at a position is the best move according to the chess engine v.

We are now ready to define "pawn loss" and total pawn loss.

Definition 1 (Pawn loss) *Let $v_i(a_i^*(x_j))$ be a chess engine's evaluation of the best move for player i at position x_j and $v_i(a_i^j(x_j))$ be chess engine's evaluation of i's actual move. Then, the pawn loss of move $a_i^j(x_j)$ is defined as $v_i(a_i^*(x_j)) - v_i(a_i^j(x_j))$.*

Definition 2 (Total pawn loss value) *Let $\bar{a} \in \bar{A}$ be a chess game (i.e., a play) and a_i^j be player i's action at position x_j in chess game $\bar{a}$, where $\bar{a}_i = (a_i^1, a_i^2, \dots, a_i^{l_i})$ for some l_i. Then, player i's total pawn loss value (TPLV) is defined as*

$$TPLV_i(\bar{a} \mid v_i) = \sum_{j=1}^{l_i} [v_i(a_i^*(x_j)) - v_i(a_i^j(x_j))].$$

Let $\bar{a}^1, \bar{a}^2, \dots, \bar{a}^K$ where $\bar{a}^k \in \bar{A}$ be a sequence of chess games in each of which i is a player. Player i's cumulative TPLV is defined as

$$C\text{-}TPLV_i((\bar{a}^1)_{j=1}) = \sum_{k=1}^{K} TPLV_i(\bar{a}^k \mid v_i).$$

In words, at every position the difference between the evaluations of a player's actual move and the best move is calculated. A player's TPLV is simply the total number of pawn-units the player loses during a chess game. TPLV is simply a mechanism that inputs an AI, v, and a chess game, $\bar{a}$, and outputs a tiebreak score for player i.

It is straightforward to confirm that the average centipawn loss of a chess game $\bar{a}$ is given by $\frac{100}{l_i} \times TPLV_i(\bar{a} \mid v_i)$. For both metrics, a zero value indicates that every move made by the player matches with the engine's top move. As mentioned in the introduction, we prefer using TPLV to average centipawn loss as a tiebreaker beucause (i) we find it more intuitive, and more importantly average centipawn loss can be manipulated by deliberately extending the game in a theoretically drawn position to decrease one's own average centipawn loss.

The player with the lower *TPLV* receives the higher tiebreaking score in a chess game. This definition is extended to tiebreaking in chess tournaments (instead of tiebreaking in individual games) as follows. In the event of a tie in a chess tournament, the player(s) with the lowest C-TPLV receive the highest tiebreaking score, and the players with the second lowest C-TPLV receive the second highest tiebreaking score, and so on.

3.2 Further practicable tiebreaking rules

We next define a specific and practicable AI tiebreaking rule. Let $\bar{a}$ be a chess game, s_i the score of player i, and $TPLV_i < TPLV_j$ be player i's TPLV in $\bar{a}$. If player i wins the chess game $\bar{a}$, then i receives 3 points and player $j \neq i$ receives 0 points: $s_i = 3$ and $s_j = 0$. If the chess game $\bar{a}$ is drawn, then each player receives 1 point. If $TPLV_i < TPLV_j$, then the player i receives an *additional* 1 point, and player j does not receive any additional point: $s_i = 2$ and $s_j = 1$. If $TPLV_i = TPLV_j$, then each player receives an additional 0.5 points.

In simple words, we propose that the winner of a chess game receives 3 points (if they have a lower TPLV) and the loser 0 points, and in the event of a draw, the player with the lower TPLV receives 2 points and the other receives 1 point. This (3, 2, 1) scoring system, which is currently used in many European ice hockey leagues, is akin to the scoring system used in volleyball if the match proceeds to the tiebreak, which is a shorter fifth set. Norway Chess also experimented with the (3, 2, 1) scoring system, but now uses (3, 1.5, 1) system perhaps to further incentivize winning a game. To our knowledge, Norway Chess was the first to use Armageddon to break ties at the game level rather than at the tournament level.

There are several ways one could use TPLV to break ties. Our definition provides a specific tiebreaking rule in case of a tie in a chess game. For example, AI tiebreaking rule can also be used with the (3, 1.5, 1) scoring system: The winner of a game receives 3 points regardless of the TPLVs, the winner of the tiebreak receives 1.5 points and the loser of the tiebreaker receives 1 point. In short, based on the needs and specific aims of tournaments, the organizers could use different TPLV-based scoring systems. Regardless of which tiebreaking rule is used to break ties in specific games, our definition provides a tiebreaking rule based on cumulative TPLV in chess tournaments. In the unlikely event that cumulative TPLVs of two players are equal in a chess tournament, the average centipawn loss of the players could be used as a second tiebreaker; if these are also equal, then there is a strong indication that the tie should not be broken. (For uniformity against slight inaccuracies in chess engine evaluations, we suggest using a certain threshold, within which the TPLVs can be considered equivalent.) But if the tie has to be broken in a tournament such as the world championship, then we suggest that players play two games—one as White and one as Black—until the tie is broken by the AI tiebreaking rule. In the extremely unlikely event that the tie is not broken after a series of two-game matches, one could, e.g., argue that the reigning world champion should keep their title.

3.3 Data collection and computation

As previously highlighted, our dataset encompasses all undisputed world chess championship matches in history that concluded in a tie (www.pgnmentor.com). Each match consists of a fixed, even number of head-to-head games, with players alternating colors. This format naturally leads to the possibility of ties. Our analysis reveals a notable trend: while only 11% of matches ended in a tie before the year 2000, this figure rose to approximately 38% in the post-2000 era, as illustrated in Fig 1.

The significance of the world championship cannot be overstated; it represents the pinnacle of competitive chess, and winning the world championship is considered the ultimate achievement in a chess player's career. Our dataset is extensive, covering 286 games across nine world championship matches, with players making just above 25,000 moves in total.

Our primary goal is to determine a winner in tied matches by favoring the player with a lower cumulative TPLV, rather than resorting to playing speed chess tiebreaks. It is important to clarify that our objective is not to retroactively predict the actual winner of these historical matches. Instead, our approach provides a novel way to assess these games through systematic analysis.
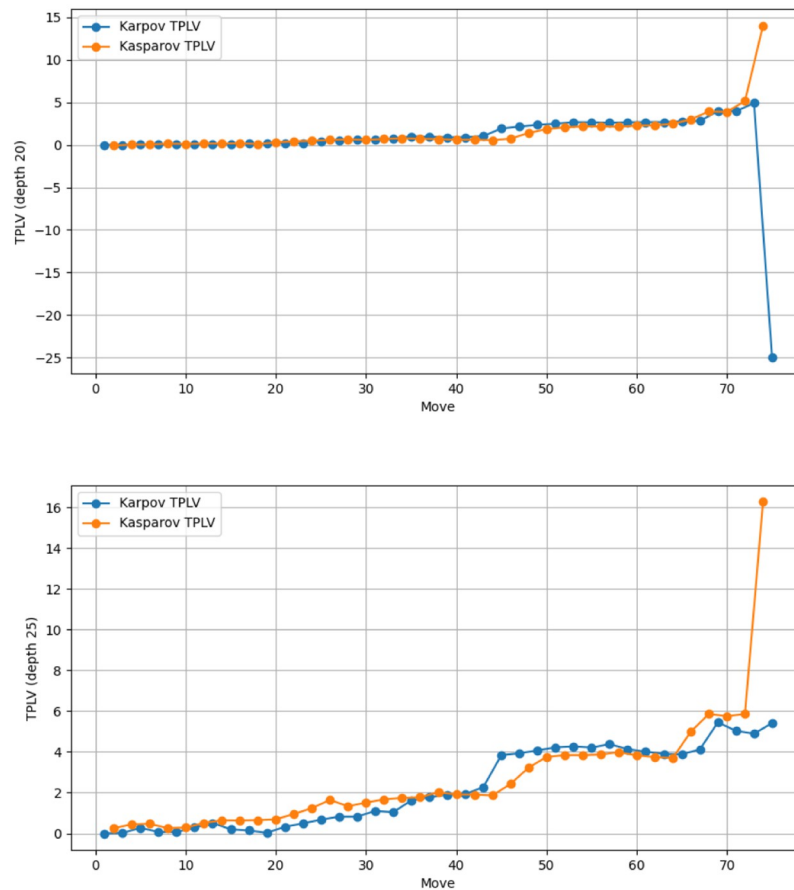
Fig 3. TPLV dynamics in game 5 of the 1987 Kasparov-Karpov match under different depths.

https://doi.org/10.1371/journal.pone.0305905.g003

For our analysis, we employed Stockfish 16, one of the most powerful chess AI systems available. This analysis was conducted on a personal desktop computer with the engine set to a depth of 20, signifying a search depth of about 20 moves (see https://www.chessprogramming.org/Depth). Note that, at these settings, Stockfish 16 significantly surpasses the strength of the best human players, providing generally a reliable and objective assessment of each move. For this reason, our settings are reasonable in comparison with those reported in the literature; for example, [37] uses Houdini 1.5a at a depth of 15, and [38] uses Stockfish 11 at a depth of 25.

An exceptional instance occurred in game 5 of the Kasparov-Karpov match, where Karpov's TPLV was approximately -200. This suggested he played at a level surpassing the Stockfish depth 20 setting, a rare but not unheard-of occurrence. Consequently, we reanalyzed the entire match at a depth of 25 (see Fig 3). We have pinpointed the move that caused this anomaly. Under depth 20, Karpov's last move significantly increased his evaluation of the position, an unexpected outcome as the best engine move should not change the evaluation, and a suboptimal move should decrease it. This indicates that, at this setting, the engine was unable to evaluate the position correctly one move earlier. However, when analyzed at depth 25, this anomaly disappeared, and the outcome of the tiebreak, according to our method, remained unchanged.

We have included all game-level data in the S, specifically in Figs 4–6. Table 2 presents summary statistics. The table includes the years the matches took place, the players, the total number of games in each match, and the cumulative TPLV (C-TPLV) for each player.

4 Limitations

4.1 Sensitivity to the software (AI system) and the hardware

Both the AI system (software) and the hardware play a role in calculating TPLVs in a game. In addition, as noted earlier, specific software settings, such as search depth, can also affect the evaluations produced by the selected AI system. Consequently, it is essential that all factors influencing the evaluations be disclosed publicly prior to a tournament.

In particular, the engine settings should be kept fixed across all games, unless the tournament director has a reasonable doubt that the AI's assessment of a particular position in a game was flawed in a way that might affect the result. In that case, the tournament director may seek a re-evaluation of the position/game.

Today, several of the best chess engines, including Stockfish and AlphaZero, are widely acknowledged to be clearly much better than humans. Thus, either of these chess engines could be employed for the AI tiebreaking rule. (In tournaments with a large number of participants, however, one could use computationally less expensive engine settings to calculate the TPLVs).

4.2 Sensitivity to the aggregation of TPLVs and the threshold selection

As indicated in Table 2, the player with the lowest cumulative TPLV wins the tiebreak in a match. A plausible objection to this tiebreak mechanism is that aggregating evaluations at every position in each game may disproportionately benefit one player over another. For example, consider the Schlechter-Lasker (1910) and Botvinnik-Bronstein (1951) matches shown in Fig 4. In each of these matches, there is only one game in which the TPLVs of the players are significantly different, whereas in all other games, the TPLVs are close. Should a single game, or even a single move, determine the outcome of the tiebreaker?

Different sports competitions have experimented with alternative approaches to aggregation issues. For example, in Rubik's cube competitions, one's worst result and best result are excluded before taking the average completion time. A similar approach could be adopted in chess: each player's worst and best games (and/or moves) could be excluded before calculating the cumulative TPLVs. Naturally, such a change should be announced prior to the event.

Furthermore, an argument can be made that the player with the higher TPLV should win the tiebreak in the event of a tie, rather than the player with the lower TPLV. This argument stems from the observation that the player with the lower TPLV failed to capitalize on the mistakes of the opponent. However, this approach conflicts with the following method for breaking ties. Consider a tournament where player A and player B draw their game; player A wins all other games with perfect play, whereas player B, despite winning the remaining games, makes many blunders. In this case, it is arguably more natural to break the tie in favor of player A.

Finally, we have suggested that a threshold might be set, whereby two TPLVs that fall within this fixed threshold are considered equivalent. Alternatively, one could set a percentage threshold whereby if the difference between TPLVs is within this threshold with respect to, say, the lower TPLV, then the TPLVs are considered equivalent. For example, with a threshold of 1%, TPLVs of 100 and 101 would be considered equivalent. To illustrate, with a threshold percentage of 1%, all ties can be broken in Table 2. However, with a threshold percentage of 5%, the
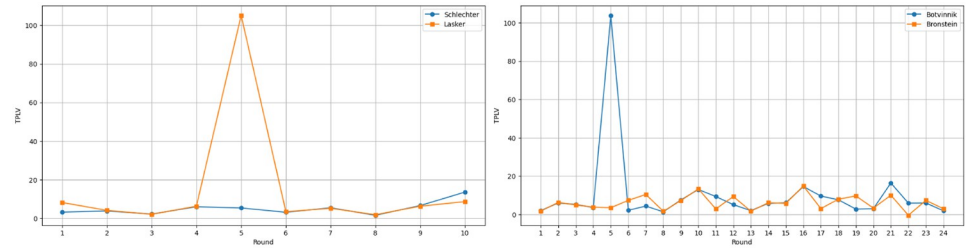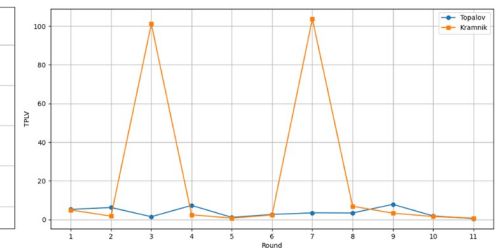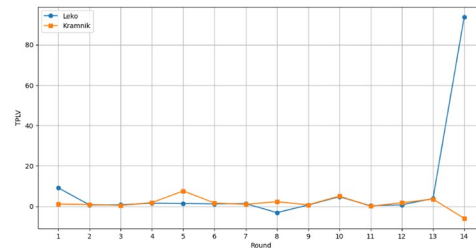
(a) Schlechter-Lasker 1910
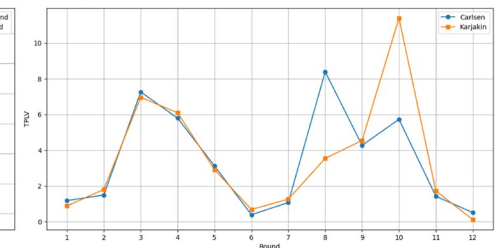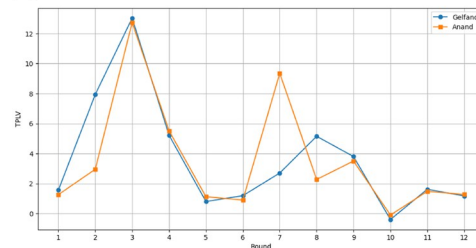
(b) Botvinnik-Bronstein 1951
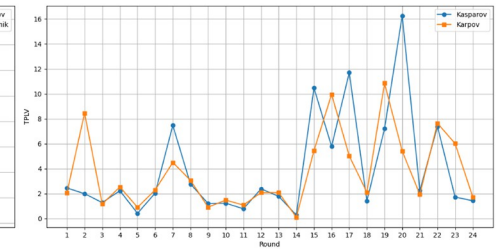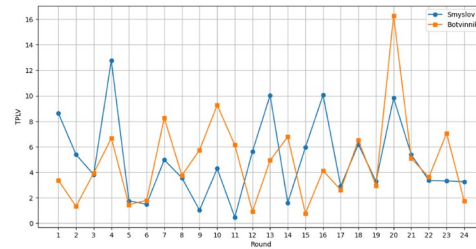
(c) Botvinnik-Smyslov 1954
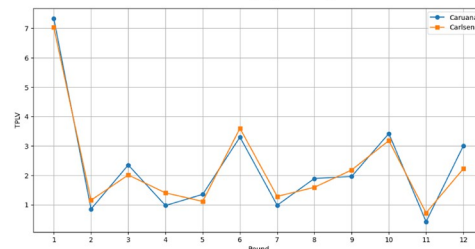
(d) Kasparov-Karpov 1987

(e) Leko-Kramnik 2004

(f) Topalov-Kramnik 2016

(g) Anand-Gelfand 2012

(h) Karjakin-Carlsen 2016

(i) Carlsen-Caruana 2018

Fig 4. TPLVs per round/game in World Chess Championship matches.

https://doi.org/10.1371/journal.pone.0305905.g004

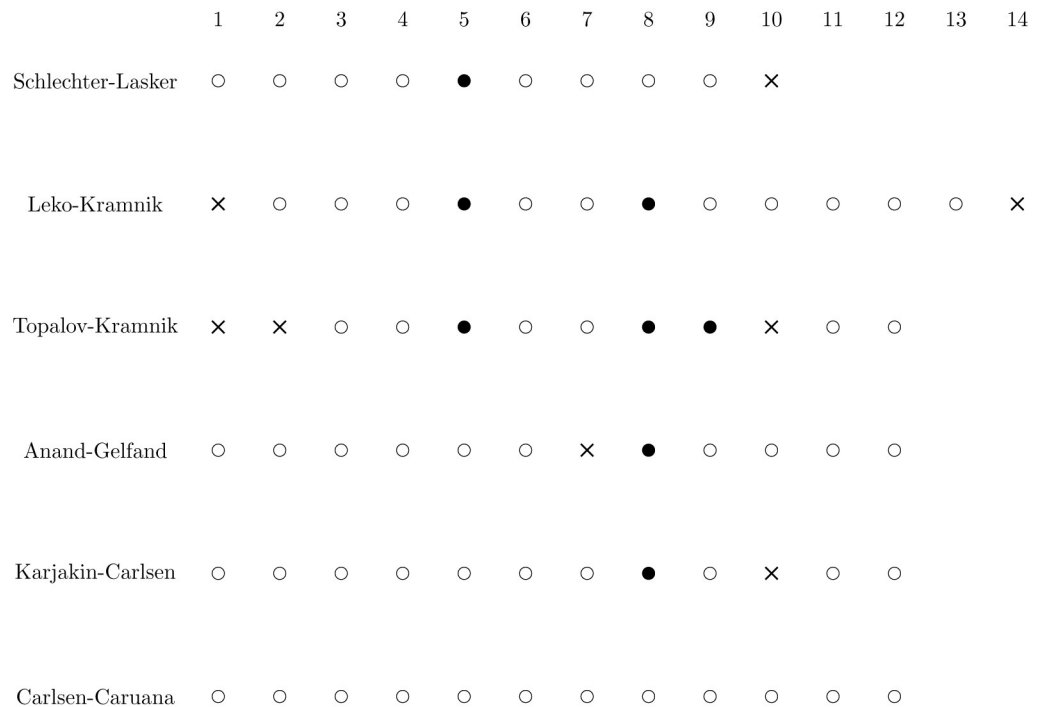|	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Schlechter-Lasker	○	○	○	○	●	○	○	○	○	×				
Leko-Kramnik	×	○	○	○	●	○	○	●	○	○	○	○	○	×
Topalov-Kramnik	×	×	○	○	●	○	○	●	●	×	○	○		
Anand-Gelfand	○	○	○	○	○	○	×	●	○	○	○	○		
Karjakin-Carlsen	○	○	○	○	○	○	○	●	○	×	○	○		
Carlsen-Caruana	○	○	○	○	○	○	○	○	○	○	○	○		

Fig 5. Outcomes of world championship matches with 14 or fewer classical games: A hollow dot represents a draw, a solid dot represents a win for the first player, and a cross represents a loss for the first player.

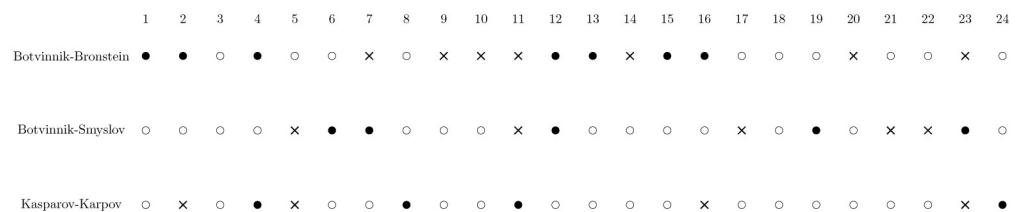|	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Botvinnik-Bronstein	●	●	○	●	○	○	×	○	×	×	×	●	●	×	●	●	○	○	○	×	○	○	×	○
Botvinnik-Smyslov	○	○	○	○	×	●	●	○	○	○	×	●	○	○	○	○	×	○	●	○	×	×	●	○
Kasparov-Karpov	○	×	○	●	×	○	○	●	○	○	●	○	○	○	○	×	○	○	○	○	○	○	×	●

Fig 6. Outcomes of world championship matches with 24 classical games: A hollow dot represents a draw, a solid dot represents a win for the first player, and a cross represents a loss for the first player.

ties in the Botvinnik-Smyslov, Anand-Gelfand, Karjakin-Carlsen, and Carlsen-Caruana matches would remain unbroken.

4.3 Different playing styles

The playing style—positional vs tactical, or conservative vs aggressive—of a player may make them more (or less) susceptible to making mistakes against a player with a different playing style. A valid concern is whether our AI mechanisms favor one style over another. The answer depends on the chess engine (software) and the hardware that are used to break ties. A top player may have a better "tactical awareness" than a relatively weak chess engine or a strong engine that runs on a weak hardware. Using such an engine to break ties would then obviously be unfair to the player. However, there is little doubt that the latest version of Stockfish running on a strong hardware is a better tactical and/or positional player than a human player. As

an analogy, suppose that a world chess champion evaluates a move in a game played by amateur players. While the world champion may be biased, like any other player, there is little doubt that their evaluation would be more reliable than the evaluation of an amateur.

Another point of concern could be the impact of openings that typically result in "theoretical draws," such as the Berlin draw. There is an argument to be made for excluding such games from the TPLV calculations. Exploring how TPLV-based tiebreaking can be adapted in such cases is an interesting direction for future research. (Another factor that could potentially change the playing style of the players occurs when they are informed of the TPLVs during a game. If this were possible, it would greatly impact players because giving top players evaluations of positions would help them considerably, even without the TPLV considerations. However, this would violate the fair play rules).

4.4 Computer-like play

A reasonable concern could be that our proposal will make players to play more "computer-like." Nevertheless, top chess players now play arguably more like engines than they did in the past, as indicated by the apparent decrease in cumulative TPLVs shown by Table 2. Expert players try to learn as much as they can from engines, including openings and end-game strategies, in order to gain a competitive edge. As an example, [39] recently explained how he gained a huge amount of knowledge and benefited from neural network-based engines such as AlphaZero. He also said that some players have not used these AIs in the correct way, and hence have not benefited from them. (For a further discussion, see [40]). In summary, there is little, if anything, that the players can do to play more computer-like and take advantage of the AI tiebreaking mechanism on top of what they normally do. To put it slightly differently, if there is any "computer-like" chess concept that a player can learn and improve their AI score, then they would learn this concept to gain a competitive edge anyway—even if AI tiebreaking mechanism is not used to break ties. That being said, it is up to the tournament organizers to decide which chess engine to use for tiebreaking, and some engines are more "human-like" than the others [41].

4.5 Playing strength

It is simpler to play (and win) a game against a weaker opponent than a stronger opponent, and a player is less likely to make mistakes when playing against a weaker opponent. Is it then unfair to compare the quality of the moves of different players? We do not think so. First, in most strong tournaments, including the world championship and the candidates tournament, every player plays against everyone else. Second, in Swiss-system tournaments, players who face each other at any round are in general of comparable strength due to the format of this tournament. While it is impossible to guarantee that each tied player plays against the same opponents in a Swiss-system tournament, we believe that AI tiebreaking mechanisms are arguably preferable to other mechanisms because they are impartial and based on the quality of the moves played by the player himself or herself. (For ranking systems used in Swiss-system tournaments and a review of these systems, see, e.g., [42, 43]).

4.6 How will the incentives of the players change?

Apart from boosting the quality of matches by naturally giving more incentives to players to find the best moves, our quality-based tiebreaking rule provides two additional benefits. First, observe that it is very likely to discourage "prematurely" agreed draws, as there is no assurance that each player will have the same TPLV when a draw is agreed upon during a game; thus, at least the player who senses having the worse (i.e., higher) TPLV up to that point will be less

likely to offer or agree to a draw. Second, this new mechanism is also likely to reduce the incentive for players to play quick moves to "flag" their opponent's clock—so that the opponent loses on time—because in case of a draw by insufficient material, for instance, the player with the lower TPLV would gain an extra point.

4.7 What is the "best strategy" under our tiebreak mechanism?

Another valid question is whether playing solid moves, e.g., the top engine moves from beginning to the end, is the "best strategy" under any of our tiebreak mechanisms. The answer is that the best strategy in a human vs human competition is *not* to always pick the top engine moves. This is because the opponent might memorize the line that is best response to the top engine moves in which case the outcome of the game would most likely be a draw. Under our tiebreak mechanism, the only time a player should deviate from the optimal move should be when one does *not* want the game to go to the tiebreak—i.e., when one wants to win the game. And, winning the game is more valuable than winning a tiebreak. Therefore, playing a suboptimal computer move might be the "human-optimal" move to win the game.

5 Conclusions

In contrast to the current tiebreak system of rapid, blitz, and Armageddon games, the winner of the tiebreak under a quality-based tiebreak strategyproof AI mechanism is determined by an objective, state-of-the-art chess engine with an Elo rating of about 3600. Under the new mechanism, players' TPLVs are highly likely to be different in the event of a draw by mutual agreement, draw by insufficient material, or any other 'regular' draw. Thus, for a small enough threshold value, nearly every game will result in a winner, making games more exciting to watch and thereby increasing fan viewership.

A valid question for future research direction is whether and to what extent our proposal could be applied to other games and sports. While our tiebreaking mechanisms are applicable to all zero-sum games, including chess, Go, poker, backgammon, football (soccer), and tennis, one must be cautious when using such an AI tiebreaking mechanism in a game/sport where AI's superiority is not commonly recognised, particularly by the best players in that game. Only after it is established that AI is capable of judging the quality of the game—which is currently the case only in a handful of games including Go, backgammon, and poker [44]—do we recommend using our tiebreaking mechanisms.

Supporting information

S1 Data.
(ZIP)

Acknowledgments

Author Contributions

Conceptualization: Nejat Anbarci, Mehmet S. Ismail.

Data curation: Nejat Anbarci, Mehmet S. Ismail.

Formal analysis: Nejat Anbarci, Mehmet S. Ismail.

Investigation: Nejat Anbarci, Mehmet S. Ismail.

Methodology: Nejat Anbarci, Mehmet S. Ismail.

Resources: Nejat Anbarci, Mehmet S. Ismail.

Software: Nejat Anbarci, Mehmet S. Ismail.

Validation: Nejat Anbarci, Mehmet S. Ismail.

Visualization: Nejat Anbarci, Mehmet S. Ismail.

Writing – original draft: Nejat Anbarci, Mehmet S. Ismail.

Writing – review & editing: Nejat Anbarci, Mehmet S. Ismail.

References

1. Stockfish; 2022. https://stockfishchess.org/.

2. Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science. 2018; 362(6419):1140–1144. https://doi.org/10.1126/science.aar6404 PMID: 30523106

3. Schrittwieser J, Antonoglou I, Hubert T, Simonyan K, Sifre L, Schmitt S, et al. Mastering Atari, Go, chess and shogi by planning with a learned model. Nature. 2020; 588(7839):604–609. https://doi.org/10.1038/s41586-020-03051-4 PMID: 33361790

4. Campbell M, Hoane AJ Jr, Hsu Fh. Deep Blue. Artificial Intelligence. 2002; 134(1-2):57–83. https://doi.org/10.1016/S0004-3702(01)00129-1

5. Anbarci N, Sun CJ, Ünver MU. Designing practical and fair sequential team contests: The case of penalty shootouts. Games and Economic Behavior. 2021; 130:25–43. https://doi.org/10.1016/j.geb.2021.07.004

6. Apesteguia J, Palacios-Huerta I. Psychological pressure in competitive environments: Evidence from a randomized natural experiment. American Economic Review. 2010; 100(5):2548–64. https://doi.org/10.1257/aer.100.5.2548

7. Berker Y. Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations. European Sport Management Quarterly. 2014; 14(2):194–210. https://doi.org/10.1080/16184742.2014.884152

8. Brams SJ, Ismail MS. Making the Rules of Sports Fairer. SIAM Review. 2018; 60(1):181–202. https://doi.org/10.1137/16M1074540

9. Brams SJ, Ismail MS. Fairer Chess: A Reversal of Two Opening Moves in Chess Creates Balance Between White and Black. In: 2021 IEEE Conference on Games (CoG). IEEE; 2021. p. 1–4.

10. Cohen-Zada D, Krumer A, Shapir OM. Testing the effect of serve order in tennis tiebreak. Journal of Economic Behavior & Organization. 2018; 146:106–115. https://doi.org/10.1016/j.jebo.2017.12.012

11. Csató L. How to avoid uncompetitive games? The importance of tie-breaking rules. European Journal of Operational Research. 2023; 307(3):1260–1269. https://doi.org/10.1016/j.ejor.2022.11.015

12. Barden L. Chess: Armageddon divides fans while Magnus Carlsen leads again in Norway. The Guardian. 2019;.

13. Carlsen M. "I'm hoping this time there will be fewer draws"; 2022. https://chess24.com/en/read/news/magnus-carlsen-i-m-hoping-this-time-there-will-be-fewer-draws.

14. Arlegi R, Dimitrov D. Fair elimination-type competitions. European Journal of Operational Research. 2020; 287(2):528–535. https://doi.org/10.1016/j.ejor.2020.03.025

15. Csató L. Tournament Design: How Operations Research Can Improve Sports Rules. Springer Nature; 2021.

16. Rudi N, Olivares M, Shetty A. Ordering sequential competitions to reduce order relevance: Soccer penalty shootouts. PLOS ONE. 2020; 15(12):e0243786. https://doi.org/10.1371/journal.pone.0243786 PMID: 33378400

17. Palacios-Huerta I. Tournaments, fairness and the Prouhet-Thue-Morse sequence. Economic inquiry. 2012; 50(3):848–849. https://doi.org/10.1111/j.1465-7295.2011.00435.x

18. Csató L, Petróczy DG. Fairness in penalty shootouts: Is it worth using dynamic sequences? Journal of Sports Sciences. 2022; 40(12):1392–1398. https://doi.org/10.1080/02640414.2022.2081402 PMID: 35675384

19. Brams SJ, Ismail MS, Kilgour DM. Fairer Shootouts in Soccer: The m−−n Rule;.

20. Elkind E, Lipmaa H. Hybrid Voting Protocols and Hardness of Manipulation. In: Proceedings of the 16th International Conference on Algorithms and Computation. ISAAC'05. Berlin, Heidelberg: Springer-Verlag; 2005. p. 206–215.

21. Faliszewski P, Procaccia AD. AI's War on Manipulation: Are We Winning? AI Magazine. 2010; 31 (4):53–64. https://doi.org/10.1609/aimag.v31i4.2314

22. Li S. Obviously Strategy-Proof Mechanisms. American Economic Review. 2017; 107(11):3257–87. https://doi.org/10.1257/aer.20160425

23. Kendall G, Lenten LJA. When sports rules go awry. European Journal of Operational Research. 2017; 257(2):377–394. https://doi.org/10.1016/j.ejor.2016.06.050

24. Vong AI. Strategic manipulation in tournament games. Games and Economic Behavior. 2017; 102:562–567. https://doi.org/10.1016/j.geb.2017.02.011

25. Aziz H, Brandl F, Brandt F, Brill M. On the tradeoff between efficiency and strategyproofness. Games and Economic Behavior. 2018; 110:1–18. https://doi.org/10.1016/j.geb.2018.03.005

26. Brams SJ, Ismail MS, Kilgour DM, Stromquist W. Catch-Up: A Rule That Makes Service Sports More Competitive. The American Mathematical Monthly. 2018; 125(9):771–796. https://doi.org/10.1080/00029890.2018.1502544

27. Dagaev D, Sonin K. Winning by Losing: Incentive Incompatibility in Multiple Qualifiers. Journal of Sports Economics. 2018; 19(8):1122–1146. https://doi.org/10.1177/1527002517704022

28. Csató L. UEFA Champions League entry has not satisfied strategyproofness in three seasons. Journal of Sports Economics. 2019; 20(7):975–981. https://doi.org/10.1177/1527002519833091

29. Pauly M. Can strategizing in round-robin subtournaments be avoided? Social Choice and Welfare. 2013; 43(1):29–46. https://doi.org/10.1007/s00355-013-0767-6

30. Chater M, Arrondel L, Gayant JP, Laslier JF. Fixing match-fixing: Optimal schedules to promote competitiveness. European Journal of Operational Research. 2021; 294(2):673–683. https://doi.org/10.1016/j.ejor.2021.02.006

31. Guyon J. "Choose your opponent": A new knockout design for hybrid tournaments. Journal of Sports Analytics. 2022; 8(1):9–29. https://doi.org/10.3233/JSA-200527

32. Csató L. Quantifying incentive (in) compatibility: A case study from sports. European Journal of Operational Research. 2022; 302(2):717–726. https://doi.org/10.1016/j.ejor.2022.01.042

33. Csató L. On the extent of collusion created by tie-breaking rules; 2024. arXiv preprint arXiv: 2206.03961.

34. Carlsen M. World Chess Championship 2018 day 12 press conference; 2018. https://www.youtube.com/watch?v=dzO7aFh8AMU&t=315s.

35. Osborne MJ, Rubinstein A. A Course in Game Theory. Cambridge, MA: MIT Press; 1994.

36. Bonanno G. Rational Play in Extensive-Form Games. Games. 2022; 13(6):72. https://doi.org/10.3390/g13060072

37. Backus P, Cubel M, Guid M, Sánchez-Pagés S, López Mañas E. Gender, competition, and performance: Evidence from chess players. Quantitative Economics. 2023; 14(1):349–380. https://doi.org/10.3982/QE1404

38. Künn S, Seel C, Zegners D. Cognitive Performance in Remote Work: Evidence from Professional Chess. The Economic Journal. 2022; 132(643):1218–1232. https://doi.org/10.1093/ej/ueab094

39. Carlsen M. Magnus Carlsen: Greatest Chess Player of All Time, Lex Fridman Podcast #315; 2022. https://www.youtube.com/watch?v=0ZO28NtkwwQ&t=1422s.

40. González-Díaz J, Palacios-Huerta I. AlphaZero Ideas. Preprint at SSRN 4140916. 2022;.

41. McIlroy-Young R, Sen S, Kleinberg J, Anderson A. Aligning Superhuman AI with Human Behavior: Chess as a Model System. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, NY, USA: Association for Computing Machinery; 2020. p. 1677–1687.

42. Csató L. Ranking by pairwise comparisons for Swiss-system tournaments. Central European Journal of Operations Research. 2013; 21:783–803. https://doi.org/10.1007/s10100-012-0261-8

43. Csató L. On the ranking of a Swiss system chess team tournament. Annals of Operations Research. 2017; 254(1):17–36.

44. Brown N, Sandholm T. Superhuman AI for multiplayer poker. Science. 2019; 365(6456):885–890. https://doi.org/10.1126/science.aay2400 PMID: 31296650