AI-powered tiebreak mechanisms: An application to chess*

Nejat Anbarci† Mehmet S. Ismail‡

Revised: 15th October, 2022
First version: 12th August, 2022

Abstract

In this paper, we propose that AI systems serve as a judge in the event of a draw in games such as chess and in the event of a tie in tournaments. More specifically, we introduce a family of AI-based scoring mechanisms and the concept of “tiebreak strategyproofness” in $n$-person zero-sum games. A mechanism is called tiebreak strategyproof (TSP) if it is always in the best interest of every player to choose the “best” action according to a given AI system. As such, we introduce a practicable scoring mechanism in chess and show that it is TSP, i.e., it is never in the interest of a player to deliberately play a worse move to increase their advantage in case the game goes to the tiebreak. In other words, TSP mechanisms are immune to such strategic manipulations. We also show that the current “speed-chess” tiebreaks are not TSP or immune to manipulation with an example from 2018 world chess championship between Carlsen and Caruana. JEL: C72, D80

Keywords: AI, chess, tiebreaks, strategic manipulation, strategyproofness

1 Introduction

AI technology has been on the rise in sports. In an effort to reduce staff during the COVID-19 pandemic, major tennis tournaments have replaced human line judges with Hawk-Eye

*We would like to thank FIDE World Chess Champion (2005-2006) and former world no. 1 Grand Master Veselin Topalov for his comments, encouragement, and referring our proposal to Norway Chess. We are also thankful to Norway Chess and its founder Kjell Madland for the opportunity to implement our research in the 2023 edition of the tournament, which is expected to host some of the best players in the world. We are also grateful to Steven Brams, Michael Naef and Shahanah Schmid for their valuable comments and suggestions.

†Department of Economics and Finance, Durham University, Durham DH1 3LB, UK. Nejat.anbarci@durham.ac.uk

‡Department of Political Economy, King’s College London, London, WC2R 2LS, UK. mehmet.ismail@kcl.ac.uk
Live technology. 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. However, AI has already outgrown its infancy and is now ready to move beyond simple tasks such as identifying the position of a ball to aid sports in considerably more complex scenarios. Chess is a perfect starting point to study the role of AI systems on the game as a whole in part because it has become a test bed for AI research since the early 1990s. Moreover, the superiority of AI systems in chess has long been widely recognized. For example, either Stockfish—a powerful open-source chess engine—or AlphaZero can impartially evaluate the degree of optimality of each move in a game.

In this paper, we propose that AI systems serve as a judge in the event of a draw in games such as chess and in the event of a tie in tournaments. More specifically, we introduce a family of AI-based scoring mechanisms and the concept of “tiebreak strategyproofness” in $n$-person zero-sum games. A mechanism is called tiebreak strategyproof (TSP) if it is always in the best interest of every player to choose the “best” action according to a given AI system. As such, we introduce a practicable scoring mechanism in chess and show that it is TSP, i.e., it is never in the interest of a player to deliberately play a worse move to increase their advantage in case the game goes to the tiebreak. In other words, TSP mechanisms are immune to such strategic manipulations. We also show that the current “speed-chess” tiebreaks are not TSP or immune to manipulation with an example from 2018 world chess championship between Carlsen and Caruana. We anticipate our method will be a first of many applications of TSP AI systems to break ties in sports and games.

1.1 The draw problem in chess

The “draw problem” in chess has a long history. Neither chess aficionados nor elite players appear to enjoy the increasing number of draws in chess tournaments. 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. If the games in the tiebreaks did not determine the winner, then an Armageddon—i.e., sudden death—game would have been played to determine the winner.

In the Armageddon tiebreak game, which is considered by some tournaments to be the panacea in the event of a draw under classical time control, White has more time (e.g. six minutes) to think on the clock than Black (e.g., four minutes), but Black wins in the event of a draw. Leonard Barden stated the following in an article published by The

---

1 As ties have to be broken to determine a winner in many situations, there have been many studies on improving and testing the tiebreak rules of games and sports; for instance see e.g. [1–4].

2 The current world champion, Magnus Carlsen, who recently announced that he will not defend his title in the 2023 cycle, 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” (Carlsen’s interview at chess24.com).
Guardian (7th June, 2019): “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.”

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. This brings up the following question: Is there anything that is immune to manipulation that can be done without compromising the game’s rules (such as adding Armageddon or blitz games instead of regular games), reducing the game’s quality, and causing controversy regarding its fairness? The answer is affirmative.

1.2 An AI-based scoring mechanism

In chess, it is straightforward to deduce that the winner played a higher quality game than the loser in the event of a win. In the event of a tie, however, it is more difficult to assert that the two players’ performances were comparable, let alone identical. With the advancements in chess AIs, their differences in quality can now 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.

We instead propose 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, and (ii) it can be manipulated by intentionally extending the game. 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. If the TPLV is equal to zero, then it indicates that every move was perfect according to a chess engine.

Along the above lines, we propose the following AI scoring rule. In the event of a win, the winner receives 2 points and the loser 0 points, and the player with the lower TPLV receives an additional 1 point. In the event of a draw, the player with the lower TPLV receives 2 points and the other receives 1 point. (This is akin to the scoring system in volleyball matches that proceed to the tie-break, which is a shorter fifth set.) Each player receives 1.5 points when both players have the same TPLV, or their TPLVs are within a threshold determined by tournament organizers or FIDE, the International Chess Federation. For uniformity against slight inaccuracies in chess engine evaluations, we suggest using a certain threshold, e.g. 5%, within which the TPLVs can be considered equivalent.

As mentioned earlier, we introduce a novel criterion for tiebreak mechanisms in chess (and other games) to be immune to strategic manipulation, and we call such mechanism tiebreak strategyproof (TSP) mechanisms. A chess tiebreak mechanism is TSP following

---

3If the chess engine is “strong,” then the winner should not have a higher TPLV unless the opponent runs out of time. If the player who lost on time has a lower TPLV, then they receive 1 point instead of 0 points. This is to disincentivize players playing quick moves to “flag” their opponent.
the “best” move recommendation of AI (e.g., Stockfish) can only improve the player’s tiebreak standing. We show that the AI scoring mechanism defined above is indeed TSP. Under our TSP mechanism, players are not incentivized to act in a Machiavellian way by switching to another course of action which would lead to the use of a very different set of rules of the game to his/her advantage. After the next section that contains several examples from real life, we will propose a family of TSP mechanisms that can be used to break such ties.

2 Real-life examples

“Everybody could see that I wasn’t really necessarily going for the maximum. I just wanted a position that was completely safe and where I could put some pressure. If a draw hadn’t been a satisfactory result, obviously I would have approached it differently” (Magnus Carlsen on game 12 in the 2018 World chess championship).

In this section, as suggested by the above quote, we will illustrate that fast-control (i.e., rapid, blitz, and Armageddon) tiebreaks are not TSP with a counter-example in an actual game. If the tiebreaks are decided with faster time control games, then the better speed-chess player—measured by having a greater Elo under rapid/blitz time-control—might be incentivized to make a weaker move under classical time control to draw the game.

2.1 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 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. As we will show later, our AI-based tiebreak format better aligns these incentives.

Carlsen, of course, knew what he was doing. During the post-game interview, he said “My approach was not to unbalance the position at that point” near the end of the game. Indeed, in our opinion Carlsen would not have offered a draw in their last game under TPLV scoring system because he was already doing better in terms of having a lower

4Note that the situation would be very different if Carlsen would guarantee winning the championship with a draw in the last game. In that case, the incentive compatibility issue is not created by the tiebreak mechanism but by (i) the scoring mechanism that gives a strictly positive point to a draw, and (ii) the fact that the value of world championship is much greater than the value of winning a game. We do not think that it is desirable and practicable to avoid such scenarios in part because the value of winning the world championship title is huge. That being said, offering extra cash prizes to the winner of each game can help incentivize players to win the games as well. For these reasons, our tiebreaking mechanisms intentionally do not rule out the aforementioned scenarios.
| Name                          | TPLV | Score | AI Score |
|-------------------------------|------|-------|----------|
| GM Fabiano Caruana (W)        | 5.9  | 0.5   | 2        |
| GM Magnus Carlsen (B)         | 6.2  | 0.5   | 1        |

Table 1: Game 12 in the 2018 world chess championship. TPLVs include the draw offer and its acceptance.

| Name                          | TPLV | Score | AI Score |
|-------------------------------|------|-------|----------|
| GM Ian Nepomniachtchi (W)     | 1.6  | 0.5   | 1        |
| GM Magnus Carlsen (B)         | 1.2  | 0.5   | 2        |

Table 2: TPLVs in the “most accurate game” played in the world championship history.

TPLV than Caruana (5.2 vs 5.9)\(^5\) As Table 1 illustrates, a draw offer in that position would make Caruana the winner of the tiebreak according to the TPLVs.

### 2.2 The most accurate game

Table 2 above shows the TPLVs in the “most accurate game” played in the world championship history. Our rule suggests that Magnus Carlsen should be the tiebreak winner in the third game because he completed the game with a lower TPLV.

### 2.3 Armageddon tiebreaks

In another example, Table 2.3 illustrates the outcome of the game played by Maxime Vachier-Lagrave and Veselin Topalov in 2022 Norway Chess. Topalov lost the Armageddon tiebreak game played just after the classical game. However, notice that Topalov’s TPLV is lower, so according to our TPLV scoring system he would have won the tiebreak.

### 3 Tiebreak strategyproof mechanisms

Let \( G = (N, X, I, \pi, S) \) be an extensive form zero-sum game with perfect information and perfect recall. \( N = \{1, 2, \ldots, n\} \) denotes the set of players, \( X \) a finite game tree, \( x_0 \) the

\(^5\)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.
root of the game tree, \( Z \subseteq X \) the set of terminal nodes, \( I \) the player function, \( \pi_i : Z \to \mathbb{R} \) payoff function of player \( i \in N \), and \( \pi \) the profile of payoff functions. For every \( z \in Z \), \( \sum_{i \in N} \pi_i = c \), where \( c \in \mathbb{R} \) is a constant.

Let \( X_i = \{ x \in X | I(x) = i \} \), i.e., the set of nodes at which \( i \) is active. Let \( A_i(x) \) denote the finite set of pure actions of player \( i \) at node \( x \) and \( A_i = \bigcup_{x \in X} A_i(x) \) player \( i \)'s set of all pure actions. With a slight abuse of notation, each action \( a_i \) of a player \( i \) at node \( x \) can be naturally associated with a function \( a_i : X_i \to X \) where \( a_i(x) \mapsto x’ \) denotes player \( i \)'s action \( a_i \in A_i(x) \) at node \( x \) that leads to node \( x' \in X \). Thus, \( A_i(x) \) can be viewed as the set of all successor nodes of \( x \). A path is a sequence \( p = (x_0, x_1, ..., x_m) \) in which \( x_m \in Z \) and for every \( k \in \{0, 1, ..., m - 1\} \), \( x_{k+1} \) is a successor of \( x_k \). Let \( \bar{a}_i = (a^1_i, a^2_i, ..., a^n_i) \) be a sequence of actions of player \( i \). A sequence \( \bar{a} = (\bar{a}_1, \bar{a}_2, ..., \bar{a}_n') \) is called a play if there exists a path \( p = (x_0, x_1, ..., x_m) \) for every action \( a^j_i \) in sequence \( a \) there exists a node \( x_k \) in path \( p \) such that \( a^j_i(x_k) = x_{k+1} \). Let \( \bar{A} \) be the set of all plays.

### 3.1 Basic concepts

**Definition 1 (AI).** Let \( G \) be a game. An AI is a profile \( v \) of functions where for each player \( i, v_i : X \to \mathbb{R} \).

In words, an AI mechanism yields a value (i.e., evaluation) for every player and every node in a game \( G \). As an example, a chess AI mechanism is a chess engine which inputs a position and outputs an evaluation of the position.

**Definition 2 (AI best-response).** An action \( a^*_i \) at a node \( x \) is called an AI best-response if

\[
    a^*_i(x) \in \arg \max_{a_i(x) \in A_i(x)} v_i(a_i(x)).
\]

In words, an AI best-response is the best action chosen by the AI mechanism at a given position (i.e., node in the game tree).

**Definition 3 (AI scoring rule).** Let \( G \) be a game and \( \bar{a} \in \bar{A} \) be a play. A scoring rule is a function \( r \) whose \( i \)'th component is defined as \( r_i : \bar{A} \to \mathbb{R} \). Given a play \( \bar{a} \), if \( i \in \arg \min_{i \in N} r_i(\bar{a}) \), then player \( i \) is ranked the highest in the game \( G \). There could be ties.

Let \( a^j_i \) be player \( i \)'s action at node \( x_j \) in play \( \bar{a} \). An AI scoring rule is a scoring rule, \( r \), such that for every player \( i \) and every play \( \bar{a} \),

\[
    r_i(\bar{a}) := \sum_{j=1}^{t_i} [v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))].
\]

A scoring rule gives a score for every play in a game. An AI scoring rule is a scoring rule that aggregates the “errors” in a given play based on the evaluation function of an AI mechanism. The player with the lowest total errors “wins” the game.
**Definition 4 (Tiebreak strategyproofness).** Let $G$ be a game. A play $\bar{a} \in \bar{A}$ is called *tiebreak strategyproof* (TSP) if for every action $a^k_i$ in $\bar{a}$,

$$\sum_{j=1, j \neq k}^{l_i} [v_i(a^*_i(x_j)) - v_i(a^k_i(x_j))] \leq \sum_{j=1}^{l_i} [v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))].$$

An AI scoring rule $r$ is called TSP if every play $\bar{a}$ is TSP.

Given a play, fix the total errors excluding a node $x$ on the path of the play. If the play is TSP, then it is in the best interest of the active player at $x$ to choose an AI best-response action.\(^6\)

### 3.2 TSP mechanisms in chess

Let $G$ denote chess under the standard FIDE (International Chess Federation) rules. We assume that a draw offer/acceptance is an action, so it is part of game $G$. A play $\bar{a} \in \bar{A}$ is called a chess game in the chess terminology. A node $x \in X$ is called a position. Let $T(G)$ denote a chess tournament which specifies the rules of a chess competition, e.g., the world chess championship (or a Swiss tournament) where a player does not play all other competitors.

We will now introduce our metric of “total pawn loss.”

**Definition 5 (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 $v_i(a^j_i(x_j))$ be chess engine’s evaluation of $i$’s actual move. Then, the *pawn loss* of move $a^j_i(x_j)$ is defined as $v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))$.

**Definition 6 (Total pawn loss value).** Let $\bar{a} \in \bar{A}$ be a chess game (i.e., a play) and $a^j_i$ be player $i$’s action at position $x_j$ in chess game $\bar{a}$, where $\bar{a}_i = (a^1_i, a^2_i, ..., a^l_i)$. Then, player $i$’s *total pawn loss value* (TPLV) is defined as,

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

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

$$\sum_{k=1}^{K} TPLV_i(\bar{a}^k).$$

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 lose during a chess game.

\(^6\)A straightforward extension of the TSP could be to define it with respect to a more general function of the errors made in a game rather than with respect to the total errors. We keep the current definition for its simplicity. For an informative study of strategyproofness in sports, see [7].
**Definition 7** (TPLV-based AI chess scoring mechanisms). We define a family of AI scoring mechanisms based on the type of competition.

1. Games: The player with the lowest $TPLV$ receives an additional point or points, on top of their score based on the outcome of the game (i.e., win, draw, or a loss).

2. Tournament: In case of ties in a chess tournament, the ties are broken in favor of the player(s) with the lowest cumulative TPLV whom are ranked the first, the player(s) with the second lowest TPLV rank the second, and so on.

We next define a specific scoring rule for chess games.

**Definition 8** (TPLV-based AI scoring rule for chess). Let $a$ be a chess game, $s_i$ the score of player $i$, and $TPLV_i < TPLV_j$ be player $i$’s TPLV in $a$. If player $i$ wins the chess game, $a$, then $i$ receives 2 points and player $j \neq i$ receives 0 points: $s_i = 2$ and $s_j = 0$. If chess game $a$ is drawn then each player receives 1 point. Under any case, 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 is akin to the scoring system in volleyball matches that proceed to the tie-break, which is a shorter fifth set).

There are several ways one could use TPLV to break ties. Definition 8 provides a specific scoring rule in case of a tie in a chess game. Regardless of whether this rule is used to break ties in specific games, Definition 7 provides a tiebreaking rule based on cumulative TPLV in chess tournaments. In the extremely unlikely event that cumulative TPLVs of two players are equal in a chess tournament, there is a strong indication that the tie should not be broken. But if the tie has to be broken such as in the world championship, then one could, e.g., argue that the reigning world champion keeps their title.

### 3.3 Main result

We next show that our tiebreaking rule based on TPLV is indeed tiebreak strategyproof.

**Theorem 1** (AI strategy-proof mechanisms). *AI scoring mechanism for chess is TSP.*

The proof of the theorem is in the Appendix. To explain TSP in games in plain words, suppose, to reach a contradiction, that AI scoring rule for chess is not TSP in a chess game. This implies that there is some player $i$, a position in a chess game, and there are two moves (move 1 and move 2) such that move 1 is the best move according to an engine and its evaluation is greater than engine evaluation of move 2. Notice that choosing move 1 would decrease player $i$’s TPLV, which implies that player $i$ would be better off with choosing move 1 instead of move 2. Thus, AI scoring mechanism for chess is indeed TSP.
4 Concluding remarks: Potential concerns/benefits and future directions

Both the AI system (software) and the hardware play a role in calculating TPLVs in a game. Thus, both of these should be made public knowledge in advance of a tournament. 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 AI machinery could be employed for the AI scoring rule. (In tournaments with a large number of participants, however, one could use less computationally expensive engine settings to calculate the TPLVs in the games.)

A reasonable concern could be that our proposal will make players to play more “computer-like.” Nevertheless, chess players now play more like engines than they did in the past. 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, Carlsen recently explained how he gained a huge amount of knowledge and benefit 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 [6].

Clearly, it is simpler to play (and win) a game against a weaker opponent than a stronger opponent, hence a player is more likely to have a lower TPLV when playing against a weaker opponent. Is it then unfair to compare the TPLVs 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 tournaments, players who face each other at any round are of comparable strength due the format of this tournament. While it is impossible to guarantee that each tied player plays against the same opponents in a Swiss tournament, we believe that AI scoring mechanism is preferable to other mechanisms because it is impartial and it is based on the quality of the moves played by the player themself as opposed to other tiebreak mechanisms based on e.g. the performance of the player’s opponents. (For a review of ranking systems used in Swiss tournaments, see [5].)

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

In addition, note that—apart from boosting the quality of matches by naturally giving more incentives to players to find the best moves—this quality-based tiebreaking rule would provide two additional benefits. First, observe that it is very likely to discourage
"premature" 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 agree to the draw. Second, this new rule 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 TPL would gain an extra point.

A valid question for future research direction is whether and to what extent our proposal could be applied to other games and sports. Note that we have defined AI scoring mechanisms and strategyproofness for a general class of $n$-person zero-sum games. Thus, our scoring mechanisms are applicable to all games in this class, including Go, backgammon, football (soccer), and tennis. However, one must be cautious when using AI scoring mechanism in a game/sport where its 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 checkers—do we recommend using our TSP scoring mechanism.

Appendix

4.1 Proof of Theorem 1

Proof. We show that AI scoring mechanisms are TSP in games and tournaments.

Case 1: chess games. To reach a contradiction, suppose that AI scoring rule for chess is not TSP. It implies that there exists a chess game $\bar{a} \in \bar{A}$ and some action $a^i_k$ in $\bar{a}$ such that

$$\sum_{j=1, j \neq k}^{l_i} [v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))] > \sum_{j=1}^{l_i} [v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))] = TPLV_i(\bar{a}).$$

Note that $TPLV_i(\bar{a})$ can be written as

$$\sum_{j=1, j \neq k}^{l_i} [v_i(a^*_i(x_j)) - v_i(a^j_i(x_j))] + [v_i(a^*_i(x_k)) - v_i(a^j_i(x_k))].$$

Thus, inequality (1) be rearranged as follows.

$$0 > [v_i(a^*_i(x_k)) - v_i(a^j_i(x_k))],$$

if and only if

$$v_i(a^j_i(x_k)) > v_i(a^*_i(x_k)).$$

However, this contradicts to the fact that at node $x_k$, $a^*_i(x_k)$ is an AI best-response, i.e.,

$$a^*_i(x) \in \arg \max_{a_i(x) \in A_i(x)} v_i(a_i(x)).$$
Thus, AI scoring rule is TSP in games.

Case 2: Tournaments. To reach a contradiction, suppose that the AI scoring rule is not TSP in a match. This implies that there is some player $i$ and a position $x_k$ in a chess game $\bar{a}$, there are two moves, $a_i^j(x_k)$ and $a_i^*(x_k)$, with engine evaluations $v_i(a_i^j(x_k))$ and $v_i(a_i^*(x_k))$ such that $v_i(a_i^j(x_k)) > v_i(a_i^*(x_k))$, and choosing the move with evaluation $v_i(a_i^j(x_k))$ improves player $i$’s cumulative TPLV compared with the other move in the tournament (rather than the chess game as in the previous case). This move would then increase player $i$’s TPLV in the game and cumulative TPLV in the tournament. As a result, AI scoring mechanism is TSP in tournaments.

As desired, AI scoring mechanism for chess is TSP.

References

[1] Nejat Anbarci, Ching-Jen Sun, and M Utku Ünver. Designing practical and fair sequential team contests: The case of penalty shootouts. Games and Economic Behavior, 130:25–43, 2021.

[2] Jose Apesteguia and Ignacio Palacios-Huerta. Psychological pressure in competitive environments: Evidence from a randomized natural experiment. American Economic Review, 100(5):2548–64, 2010.

[3] Steven J Brams and Mehmet S Ismail. Making the Rules of Sports Fairer. SIAM Review, 60(1):181–202, 2018.

[4] Danny Cohen-Zada, Alex Krumer, and Offer Moshe Shapir. Testing the effect of serve order in tennis tiebreak. Journal of Economic Behavior & Organization, 146:106–115, 2018.

[5] László Csató. On the ranking of a Swiss system chess team tournament. Annals of Operations Research, 254(1):17–36, 2017.

[6] Julio González-Díaz and Ignacio Palacios-Huerta. AlphaZero Ideas. Available at SSRN 4140916, 2022.

[7] Marc Pauly. Can strategizing in round-robin subtournaments be avoided? Social Choice and Welfare, 43(1):29–46, 2013.