Performance rating in chess, tennis, and other contexts

Mehmet S. Ismail*

Tuesday 19th December, 2023

Abstract

In this note, I introduce Estimated Performance Rating (PR^e), a novel system for evaluating player performance in sports and games. PR^e addresses a key limitation of the Tournament Performance Rating (TPR) system, which becomes undefined for zero or perfect scores in a series of games. PR^e is defined as the rating that solves an optimization problem related to scoring probability, making it applicable for any performance level. The main theorem establishes that the PR^e of a player is equivalent to the TPR whenever the latter is defined. I then apply this system to historically significant win-streaks in association football, tennis, and chess. Beyond sports, PR^e has broad applicability in domains where Elo ratings are used, from college rankings to the evaluation of large language models. *JEL*: D63, Z20

Keywords: Tournament performance rating, Elo rating, tennis, association football, chess

*Department of Political Economy, King's College London, London, UK. e-mail: mehmet.ismail@kcl.ac.uk

1 Introduction

The practice of rating players' performance is widespread in competitive sports and games. The Elo rating system, originally developed for chess by Elo (1978), serves as a prime example. It has since been expanded to various sports, including association football, where it is used for ranking both men's and women's international teams (FIFA 2018; Hvattum and Arntzen 2010; Csató 2021). The system is also applied in tennis (Williams et al. 2021), American football, basketball, esports, and others. Beyond sports, the Elo-based systems are used in diverse areas such as college rankings (Avery et al. 2012), evaluating large language models (Zheng et al. 2023), dating apps (Kosoff 2016), education (Pelánek 2016), and biology (Albers and Vries 2001). In this system, players' ratings are dynamically updated based on their performance and the outcomes of their games.

Within this framework, the Tournament Performance Rating (TPR) is the standard method for assessing a player's performance over a series of games. Essentially, TPR is a rating derived from the outcomes of a player's games and the Elo ratings of the opponents faced. To provide an intuition, suppose that a player scores m points in n games. Then, the player's TPR is the hypothetical rating R that would remain unchanged if a player scored m points in n games against the same opponents. TPR is widely adopted due to its straightforward interpretability. However, it has a significant shortcoming: it is undefined in cases where a player achieves a zero or a perfect score (i.e., $m = 0$ or $m = n$) in a series of games. Such cases of win- and loss-streaks are not uncommon in competitive sports. This limitation becomes particularly critical in events like tennis Grand Slams, where a player must win all matches to win the tournament. Therefore, accurately rating performances in situations of win- or loss-streaks is essential, highlighting the need for an alternative or complementary rating system to TPR in these situations.

This paper introduces a novel performance rating system, dubbed the Estimated Performance Rating (PR^e), which is based on the probability of a player scoring $m \geq 0$ points in $n \geq m$ games. Unlike the TPR, PR^e is applicable to all scores. In this system, a player's win probability against a given average opponent rating is denoted as w, and $S(w, m, n)$ represents the probability of the player scoring m points in n games. The PR^e is then defined as the hypothetical rating R that induces an optimal win probability w^*.

This optimal probability is determined by solving the following maximization problem.[1]

$$\begin{aligned} \max_{w \in [0,1]} \quad & S(w,m,n) \\ \text{s.t.} \quad & S(w,m,n) \leq 0.75. \end{aligned} \tag{1}$$

The main theorem establishes that a rating R is the TPR of the player if and only if the win probability w it induces solves the maximization problem (1) for $0 < m < n$. In other words, PR^e is equivalent to the TPR when $0 < m < n$.

To illustrate how PR^e functions, consider a situation where a player competes in two games against opponents with an average rating of 2700. Suppose that the player achieves a score of 1.5 points, a combination of a win and a draw, from these two games. In this case, both her TPR and PR^e would be calculated as 2891 (detailed analysis provided in section 2.4). As mentioned above, the interpretation of her TPR is that if she had a rating of 2891, then her rating would not change after scoring 1.5 points in two games. PR^e, however, offers a different interpretation: if the player's rating were 2891, then the probability of her scoring exactly 1.5 points in these games would reach its maximum, calculated as 0.42 in this case. This aspect of PR^e offers a unique perspective, revealing that the 'predictive' value of TPR can vary (and become as low as 0.25). The maximal probability of achieving a specific score in a series of games is not constant; it varies depending on the score itself and the average ratings of the opponents (for example, see Table 6).

It is worth noting that FIDE, the international governing body of chess, uses a slightly different performance rating than the TPR. This system is defined for perfect scores, but it is ad hoc and does not factor in the length of a winning or losing streak. An illustration of the differences between TPR, FIDE's Performance Rating (FPR), and the newly proposed PR^e is shown in Table 1. The interpretation of PR$^e = 3099$ is that if the player had a rating of 3099, then her probability of scoring 3 points in 3 games would be 0.75, meaning that the constraint in the maximization problem (1) is binding. This interpretation will be further explored in section 2.3, which also discusses setting different probability thresholds, other than 0.75, within the PR^e framework. At the outset, predicting a rating with a 0.75 probability might not seem precise enough. However, as hinted above, the probability induced by some TPRs can be significantly lower than that.

In addition to Elo-based performance rating systems mentioned above, non-Elo based systems have also been used in economics and computer science; see, for example, Guid

1. For a detailed definition of PR^e, see section 2.3.

m	n	R_a	**TPR**	**FPR**	$\mathbf{PR}^e$
1	1	2700	N/A	3500	2891
3	3	2700	N/A	3500	3099
5	5	2700	N/A	3500	3191

Table 1: A comparison of performance ratings: TPR, FPR, and PR^e

Abbreviations: R_a = average rating of opponents; TPR = Tournament Performance Rating; FPR = FIDE Performance Rating; PR^e = Estimated Performance Rating; N/A: Undefined

and Bratko (2006), Regan and Haworth (2011), Künn et al. (2022), Backus et al. (2023), and Ismail (2023). More generally, fairness in sports has been gaining attention in recent years. Theoretical and empirical works in this area include those by Scarf et al. (2009), Apesteguia and Palacios-Huerta (2010), Goossens and Spieksma (2012), Pauly (2014), Kendall and Lenten (2017), Brams and Ismail (2018), Brams et al. (2018), Arlegi and Dimitrov (2020), Anbarci et al. (2021), and Lambers and Spieksma (2021). The strategy-proofness or incentive compatibility of sports rules has also been garnering more attention. Selected contributions in this area include works by Brams et al. (2018), Dagaev and Sonin (2018), Csató (2019, 2021), and Anbarci et al. (2021).

1.1 Application to tennis, association football, and chess

In tennis, Table 2 illustrates the Grand Slam performance ratings of 2023. Carlos Alcaraz

Player	**Event**	R_a	**Score**	**TPR**	$\mathbf{PR}^e$
Carlos Alcaraz	Wimbledon	1927	7/7	N/A	2478
Novak Djokovic	French Open	1867	7/7	N/A	2417
Novak Djokovic	Australian Open	1865	7/7	N/A	2416
Novak Djokovic	US Open	1798	7/7	N/A	2349

Table 2: Tennis Grand Slam performance ratings in 2023

achieves the highest PR^e of 2478 with his win at Wimbledon, indicating that he won against a very strong field in this tournament (for details, see Table 8). Novak Djokovic has won three Grand Slams in 2023, and his average PR^e is just under 2400.

Moving to association football, Table 3 highlights the performances of teams with perfect scores in FIFA World Cups. Brazil's 1970 World Cup campaign is particularly remarkable, achieving a PR^e of 2424 across six matches. This performance has long been acknowledged as one of the finest in the history of soccer. Additionally, the table

Team	**Event**	R_a	**Score**	**TPR**	**PR**^{e}
Brazil	Mexico 1970	1900	6/6	N/A	2424
Brazil	Korea-Japan 2002	1818	7/7	N/A	2369
Italy	France 1938	1802	4/4	N/A	2253
Uruguay	Uruguay 1930	1699	4/4	N/A	2150

Table 3: Best performance ratings for perfect scores in FIFA World Cup history

illustrates other instances of perfect campaigns, including Brazil in 2002, Italy in 1938, and Uruguay in 1930.

In chess, Tables 4 and 5 present the best historical performances in tournaments and win-streaks, respectively. Bobby Fischer's 11-win streak in the 1963 USA Championship, with a PR^e of 3224, stands out as an incredible achievement in chess tournament history. His 20-win streak during 1970-1971, with an even higher PR^e of 3441, illustrates an unparalleled level of performance. This 20-win performance has been informally regarded as more impressive than Steinitz's 25-win streak, although this comparison had not been previously quantified in terms of TPR, as it is undefined. The tables also include notable performances by other chess legends, providing a historical perspective on their relative achievements under a consistent performance rating system.

Player	**Event**	**Year**	R_a	**Score**	**TPR**	**PR**^{e}
Fischer	USA Championship	1963	2593	11/11	N/A	3224
Caruana	Sinquefield Cup	2014	2802	8.5/10	3103	3103
Fischer	Candidates	1971	2740	18.5/21	3088	3088
Alekhine	San Remo	1930	2613	14/15	3072	3072
Beliavsky	Alicante	1978	2392	13/13	N/A	3052
Carlsen	Pearl Spring	2009	2762	8/10	3003	3003

Table 4: Best historical performance ratings in chess tournaments

The calculation of PR^e values, as presented in this note, relies on historical Elo ratings obtained from the following well-known sources: www.tennisabstract.com for tennis, www.eloratings.net for association football, and www.chessmetrics.com for chess. Detailed data supporting these calculations can be found in the Appendix. In chess, FIDE's official ratings have been used wherever applicable. In cases where a player did not have an established Elo rating, the player's TPR for the specific tournament in question has been used as a substitute.

Player	**Event**	**Year**	R_a	**Streak**	**PR**^{e}
Fischer	Interzonal, Candidates	1970–1971	2705	20-win	3441
Steinitz	Vienna, London	1873–1882	2581	25-win	3356
Caruana	Sinquefield Cup	2014	2793	7-win	3344
Carlsen	Tata Steel Masters	2015	2736	6-win	3260
Fischer	USA Championship	1963	2593	11-win	3224
Carlsen	Shamkir, Grenke	2019	2706	5-win	3197
Kasparov	Wijk aan Zee	1999	2632	7-win	3183
Karpov	Linares	1994	2647	6-win	3171
Lasker	New York	1893	2510	13-win	3170
Alekhine	San Remo	1930	2639	5-win	3130
Beliavsky	Alicante	1978	2392	13-win	3052

Table 5: Best historical performance ratings of win-streaks in chess

Note: For definitions of the abbreviations, see Table 1.

2 Performance Ratings

2.1 Tournament Performance Rating

As mentioned in the introduction, the Elo rating system serves as a standard method for ranking players based on their performance in competitive contexts, such as chess. This system assigns each player a rating. These ratings are used to calculate the probability of winning (interpreted in chess as the expected score since a draw is possible) for each player.

For two players with ratings A (player 1) and B (player 2), the **win probability** for player 1, denoted by $W(A, B)$, is calculated using a logistic function:

$$W(A, B) = \frac{1}{1 + 10^{\frac{B-A}{400}}}.$$

The win probability for player 2 is simply $1 - W(A, B)$.

Note that under the Elo system, the win probability of a player with rating A against a player with rating B is considered independent from their win probability against a player with a different rating, say C. Throughout the text, I also assume that all such win probabilities are independent of each other, though this does not affect the definition of the new performance rating system I introduce. In addition, the following definitions would remain valid if one uses an extension of the Elo rating system, such as the Glicko system (Glickman 1995).

A standard concept in assessing player performance in a given tournament is the **Tournament Performance Rating** (TPR). To compute this, let $b = b_1, b_2, ..., b_k$ represent a sequence of ratings of opponents faced by player 1. The **average rating** of these opponents, R_a, is given by:

$$R_a = \frac{1}{k} \sum_{j \in b} b_j.$$

Suppose that player 1 scores $m \geq 0$ points in a total of $n > m$ games against these opponents. The TPR is defined by the equation:

$$m = \frac{n}{1 + 10^{\frac{R_a - TPR}{400}}}.$$

This can be further expressed as:

$$\frac{m}{n} = \frac{1}{1 + 10^{\frac{R_a - TPR}{400}}}. \tag{2}$$

2.2 FIDE's Performance Rating

The performance rating system used by FIDE differs slightly from the TPR. The **FIDE Performance Rating** (FPR) is calculated by adding a rating difference (*dp*), which is based on the percentage score, to the average rating of opponents (R_a). Although FPR does not yield an exact match with the TPR, the results are generally similar. Importantly, FPR plays a crucial role in determining "norms," which are sets of criteria required to achieve titles such as Grandmaster (GM) and International Master (IM). FPR is defined as follows:

$$FPR = R_a + dp,$$

where the rating difference *dp* is determined by the player's score percentage (*ps*) as outlined in Table 16 (FIDE 2022).

It is important to note that for a perfect score, whether it be 7/7 or 11/11, FIDE assigns a *dp* value of 800 as is illustrated in Table 1. FIDE recognizes that for a zero or perfect score "dp is necessarily indeterminate but is shown *notionally* as 800" (emphasis added). Historically, these calculations were manually performed by FIDE officials, which is one of the reasons why FIDE uses a predefined table rather than the original TPR formula.

2.3 Estimated Performance Rating

Assume that player 1 scores m points in n games against players with an average rating R_a, where m is an integer.[2] Let $S(w, m, n)$ denote the player 1's **probability of scoring** exactly m points in n games, given player 1's win probability w against players with an average rating R_a.

$$S(w, m, n) = \binom{n}{m} w^m (1-w)^{n-m}.$$

Similarly, let $\bar{S}(w, m, n)$ denote the probability of player 1 scoring m points or more in n games, given w.

$$\bar{S}(w, m, n) = \sum_{k=m}^{n} \binom{n}{k} w^k (1-w)^{n-k}.$$

For a given threshold $t \in [0, 1]$, define the following maximization problem to find w that maximizes $S(w, m, n)$.

$$\begin{aligned} \max_{w \in [0,1]} \quad & S(w, m, n) \\ \text{s.t.} \quad & S(w, m, n) \leq t. \end{aligned} \tag{3}$$

Unless otherwise noted in this paper, I set $t = 0.75$, indicating that for a given w, it is 75% likely that player 1 scores $S(w, m, n)$.

Let w^* be the value that solves the optimization problem (3). Then, given R_a, find the value A^* such that

$$W(A^*, R_a) = w^* = \frac{1}{1 + 10^{\frac{R_a - A^*}{400}}}. \tag{4}$$

In this context, A^* is called the **Estimated Performance Rating** (PR^e) of player 1, given the score m in n games and the average rating R_a of the opposition. Here, $PR^e(w^*, R_a)$ denotes the performance rating of player 1 given w^* and R_a. Note that w^* is dependent on t, m, and n.

The next step involves solving Equation 4 for A^*. Begin by cross-multiplying to obtain:

$$w^* + w^* \cdot 10^{\frac{R_a - A^*}{400}} = 1$$

Next, we proceed to rearrange the terms: $10^{\frac{R_a - A^*}{400}} = \frac{1-w^*}{w^*}$. Applying the logarithm to both sides, we get:

$$\frac{R_a - A^*}{400} = \log_{10}\left(\frac{1-w^*}{w^*}\right).$$

2. If m is not integer then multiply, without loss of generality, both m and n by 2 to make m integer. Recall that in chess, a win is worth 1 point, a draw 0.5 points, and a loss 0 points in chess.

Rearranging the equation yields the solution for A^*, as shown in the equation below:

$$A^* = R_a - 400 \cdot \log_{10}\left(\frac{1 - w^*}{w^*}\right). \tag{5}$$

2.4 Illustrative Example

R_a	m	n	w^*	$S(w^*, m, n)$	**PR**^{e}	**TPR**
2700	0	2	0.13	0.75	2376	N/A
2700	0.5	2	0.250	0.42	2509	2509
2700	1	2	0.50	0.50	2700	2700
2700	1.5	2	0.75	0.42	2891	2891
2700	2	2	0.87	0.75	3024	N/A

Table 6: Illustrative example of performance ratings based on different scores

To illustrate the difference between TPR and PR^e, consider the example in Table 6. In this example, player 1 has an average rating of 2700 and plays 2 games against players with an average rating of 2700. $S(w^*, m, n)$ shows the probability of scoring m points in n games given w^*, which is derived from the optimization problem (3).

For the given score $m = 1$ and $n = 2$, I calculate the TPR and PR^e. For TPR, I use the formula in Equation 2 and for PR^e, I use the formula in Equation 4. Solving the following equation for TPR

$$\frac{1}{2} = \frac{1}{1 + 10^{\frac{2700 - TPR}{400}}},$$

yields TPR = 2700. Now, calculate the PR^e. For $m = 1$ and $n = 2$, we have $w^* = 0.5$. Then, plugging $w^* = 0.5$ and $R_a = 2700$ into the formula for PR^e, we obtain $PR^e = 2700$.

Next, calculate PR^e for $m = 0$ and $n = 2$. (Note that TPR is undefined for $m = 0$ and $m = 2$.) For $m = 0$ and $n = 2$, solving the optimization problem (3) yields $w^* = 0.29$. Then, plugging $w^* = 0.29$ and $R_a = 2700$ into the formula for PR^e, we obtain PR^e = 2546.89. The remainig values of PR^e and TPR are calculated similarly.

3 Main Result

Figure 1 illustrates that TPR and PR^e coincide for every m and $n \leq 30$ such that $0 < m < n$. The main result establishes that this pattern holds whenever $0 < m < n$.

Main Theorem. *Let m be the score of a player in n games such that $0 < m < n$. The rating R is the TPR of the player if and only if $W(R, R_a) \in \arg\max_{w \in [0,1]} S(w, m, n)$.*

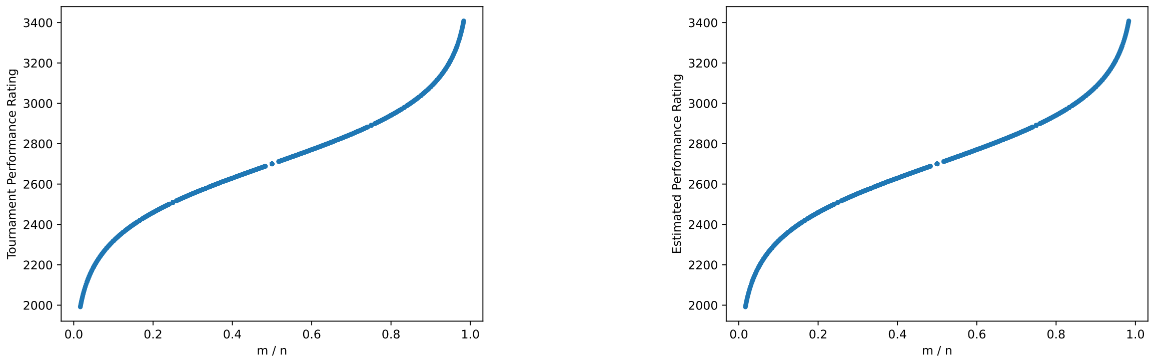

Figure 1: Plots of TPR and PR^e for every m and $n \leq 30$, where $0 < m < n$.

Proof. Consider the function to be maximized:

$$\max_{w \in [0,1]} \quad \binom{n}{m} w^m (1-w)^{n-m}$$

where m and n are constants, and $w \in [0, 1]$.

Define $f(w) = \binom{n}{m} w^m (1-w)^{n-m}$. Taking the derivative of f with respect to w, we obtain:

$$f'(w) = \binom{n}{m} \left[m w^{m-1} (1-w)^{n-m} - w^m (n-m)(1-w)^{n-m-1} \right].$$

To identify the critical points, set the derivative to zero:

$$f'(w) = 0.$$

This leads to the equation:

$$\binom{n}{m} \left[m w^{m-1} (1-w)^{n-m} - w^m (n-m)(1-w)^{n-m-1} \right] = 0$$

Simplifying the equation, we obtain:

$$m w^{m-1} (1-w)^{n-m} = w^m (n-m)(1-w)^{n-m-1}.$$

Dividing both sides by $w^{m-1}(1-w)^{n-m-1}$ yields:

$$m(1-w) = w(n-m).$$

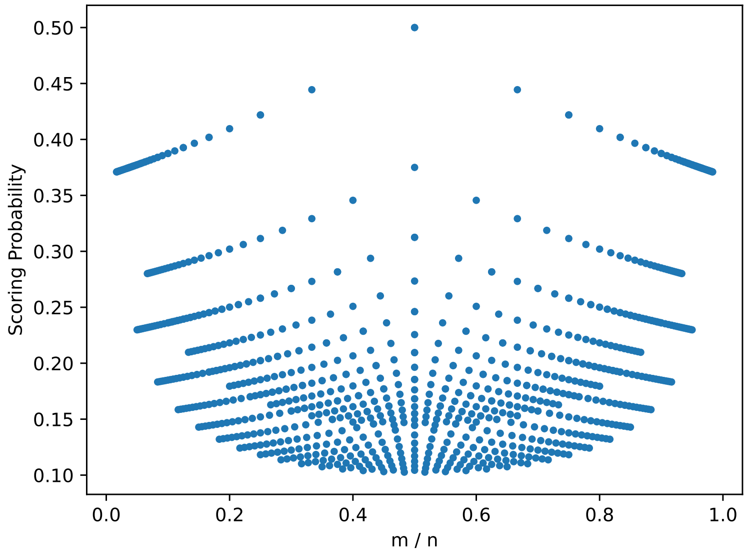

Figure 2: Plot of $f(\frac{m}{n})$ for every m and $n \leq 30$.

By rearranging, we find the critical value:

$$w^* = \frac{m}{n}.$$

Next, evaluate the second derivative of $f(w)$ at $w = \frac{m}{n}$:

$$f''(\frac{m}{n}) = \binom{n}{m} \frac{n^3(\frac{m}{n})^m \left(1 - \frac{m}{n}\right)^{n-m}}{m(m-n)}.$$

Since $m < n$, $f''(w)$ is negative at this point. Therefore, $w^* = \frac{m}{n}$ maximizes $f(w)$.

Next, assuming R is the TPR, by Equation 2, we have:

$$\frac{m}{n} = \frac{1}{1 + 10^{\frac{R_a - R}{400}}}.$$

This holds if and only if

$$W(R, R_a) = w^* = \frac{1}{1 + 10^{\frac{R_a - R}{400}}}.$$

Thus, $W(R, R_a) = w^*$ is a solution to the optimization problem (3) when $0 < m < n$. $\square$

It is instructive to examine the behavior of the score probability function at its maximum, $S(\frac{m}{n}, m, n)$. The value of the function at $w = \frac{m}{n}$ is:

$$f(\frac{m}{n}) = (\frac{m}{n})^m (1 - \frac{m}{n})^{n-m}.$$

Figure 2 illustrates the value of $f(w)$ for various values of $w = \frac{m}{n}$. Oobserve that the function reaches its maximum when $\frac{m}{n} = 0.5$, particularly when $m = 0.5$ and $n = 1$. This outcome is intuitive since, for larger values of n, the ratio $\frac{m}{n}$ represents just one among many possible scores less than or equal to n.

References

Albers, Paul CH, and Han de Vries. 2001. "Elo-rating as a tool in the sequential estimation of dominance strengths." *Animal Behaviour,* 489–495. (Cited on p. 2).

Anbarci, Nejat, Ching-Jen Sun, and M Utku Ünver. 2021. "Designing practical and fair sequential team contests: The case of penalty shootouts." *Games and Economic Behavior* 130:25–43. (Cited on p. 4).

Apesteguia, Jose, and Ignacio Palacios-Huerta. 2010. "Psychological pressure in competitive environments: Evidence from a randomized natural experiment." *American Economic Review* 100 (5): 2548–64. (Cited on p. 4).

Arlegi, Ritxar, and Dinko Dimitrov. 2020. "Fair elimination-type competitions." *European Journal of Operational Research* 287 (2): 528–535. https://doi.org/10.1016/j.ejor.2020.03.025. (Cited on p. 4).

Avery, Christopher N., Mark E. Glickman, Caroline M. Hoxby, and Andrew Metrick. 2012. "A Revealed Preference Ranking of U.S. Colleges and Universities." *The Quarterly Journal of Economics* 128, no. 1 (December): 425–467. https://doi.org/10.1093/qje/qjs043. (Cited on p. 2).

Backus, Peter, Maria Cubel, Matej Guid, Santiago Sánchez-Pagés, and Enrique López Mañas. 2023. "Gender, competition, and performance: Evidence from chess players." *Quantitative Economics* 14 (1): 349–380. (Cited on p. 4).

Brams, Steven J, and Mehmet S Ismail. 2018. "Making the Rules of Sports Fairer." *SIAM Review* 60 (1): 181–202. (Cited on p. 4).

Brams, Steven J, Mehmet S Ismail, D Marc Kilgour, and Walter Stromquist. 2018. "Catch-Up: A Rule That Makes Service Sports More Competitive." *The American Mathematical Monthly* 125 (9): 771–796. (Cited on p. 4).

Csató, László. 2019. "UEFA Champions League entry has not satisfied strategyproofness in three seasons." *Journal of Sports Economics* 20 (7): 975–981. (Cited on p. 4).

———. 2021. *Tournament Design: How Operations Research Can Improve Sports Rules.* Springer Nature. (Cited on p. 2, 4).

Dagaev, Dmitry, and Konstantin Sonin. 2018. "Winning by Losing: Incentive Incompatibility in Multiple Qualifiers." *Journal of Sports Economics* 19 (8): 1122–1146. https://doi.org/10.1177/1527002517704022. (Cited on p. 4).

Elo, Arpad E. 1978. *The Rating of Chess Players, Past and Present.* New York: Arco Publishing. (Cited on p. 2).

FIDE. 2022. *FIDE Handbook.* https://handbook.fide.com/chapter/B022022. Accessed: 01.12.2023. (Cited on p. 7).

FIFA. 2018. *Revision of the FIFA / Coca-Cola World Ranking.* https://digitalhub.fifa.com/m/f99da4f73212220/original/edbm045h0udbwkqew35a-pdf.pdf. Accessed: 2023-12-17. (Cited on p. 2).

Glickman, Mark E. 1995. *The Glicko system.* (Cited on p. 6).

Goossens, Dries R, and Frits CR Spieksma. 2012. "The carryover effect does not influence football results." *Journal of Sports Economics* 13 (3): 288–305. (Cited on p. 4).

Guid, Matej, and Ivan Bratko. 2006. "Computer analysis of world chess champions." *ICGA Journal* 29 (2): 65–73. (Cited on p. 3).

Hvattum, Lars Magnus, and Halvard Arntzen. 2010. "Using ELO ratings for match result prediction in association football." Sports Forecasting, *International Journal of Forecasting* 26 (3): 460–470. https://doi.org/10.1016/j.ijforecast.2009.10.002. (Cited on p. 2).

Ismail, Mehmet S. 2023. "Human and Machine Intelligence in n-Person Games with Partial Knowledge." *arXiv preprint arXiv:2302.13937,* (Cited on p. 4).

Kendall, Graham, and Liam J.A. Lenten. 2017. "When sports rules go awry." *European Journal of Operational Research* 257 (2): 377–394. https://doi.org/10.1016/j.ejor.2016.06.050. (Cited on p. 4).

Kosoff, Maya. 2016. *There's a secret Tinder rating system and your score can only be seen by the company.* https://www.businessinsider.com/secret-tinder-rating-system-called-elo-score-can-only-be-seen-by-company-2016-1. Accessed: 2023-12-17, January. (Cited on p. 2).

Künn, Steffen, Christian Seel, and Dainis Zegners. 2022. "Cognitive Performance in Remote Work: Evidence from Professional Chess" [in en]. *The Economic Journal* 132 (643): 1218–1232. Accessed March 10, 2023. https://doi.org/10.1093/ej/ueab094. (Cited on p. 4).

Lambers, Roel, and Frits C R Spieksma. 2021. "A mathematical analysis of fairness in shootouts." *IMA Journal of Management Mathematics* 32, no. 4 (September): 411–424. Accessed December 27, 2022. https://doi.org/10.1093/imaman/dpaa023. (Cited on p. 4).

Pauly, Marc. 2014. "Can strategizing in round-robin subtournaments be avoided?" *Social Choice and Welfare* 43 (1): 29–46. (Cited on p. 4).

Pelánek, Radek. 2016. "Applications of the Elo rating system in adaptive educational systems." *Computers & Education* 98:169–179. https://doi.org/10.1016/j.compedu.2016.03.017. (Cited on p. 2).

Regan, Kenneth, and Guy Haworth. 2011. "Intrinsic Chess Ratings." *Proceedings of the AAAI Conference on Artificial Intelligence* 25 (1): 834–839. https://doi.org/10.1609/aaai.v25i1.7951. (Cited on p. 4).

Scarf, Philip, Muhammad Mat Yusof, and Mark Bilbao. 2009. "A numerical study of designs for sporting contests." *European Journal of Operational Research* 198 (1): 190–198. https://doi.org/10.1016/j.ejor.2008.07.029. (Cited on p. 4).

Williams, Leighton Vaughan, Chunping Liu, Lerato Dixon, and Hannah Gerrard. 2021. *Journal of Quantitative Analysis in Sports* 17 (2): 91–105. https://doi.org/doi:10.1515/jqas-2019-0110. (Cited on p. 2).

Zheng, Lianmin, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, et al. 2023. *Judging LLM-as-a-judge with MT-Bench and Chatbot Arena.* arXiv: 2306.05685. (Cited on p. 2).

Appendix

Player	**Tournament**	**Elo**
Baena	AUS Open	1742
Couacaud	AUS Open	1564
Dimitrov	AUS Open	1888
de Minaur	AUS Open	1945
Rublev	AUS Open	1970
Paul	AUS Open	1886
Tsitsipas	AUS Open	2058
Kovacevic	French Open	1669
Fucsovics	French Open	1783
Fokina	French Open	1864
Varillas	French Open	1687
Khachanov	French Open	1960
Alcaraz	French Open	2190
Ruud	French Open	1918
Muller	US Open	1658
Miralles	US Open	1750
Djere	US Open	1812
Gojo	US Open	1659
Fritz	US Open	1961
Shelton	US Open	1643
Medvedev	US Open	2101

Table 7: Djokovic's Opponents in 2023 Grand Slams

Player	**Tournament**	**Elo**
Chardy	Wimbledon	1808
Berrettini	Wimbledon	1848
Muller	Wimbledon	1660
Jarry	Wimbledon	1839
Medvedev	Wimbledon	2110
Rune	Wimbledon	2050
Djokovic	Wimbledon	2171

Table 8: Alcaraz's opponents in Wimbledon 2023

Opponent	**Score**	**Rating**
Argentina	4-2	2084
Yugoslavia	6-1	1608
Romania	4-0	1560
Peru	1-0	1542

Table 9: Uruguay's Matches in 1930

Opponent	**Score**	**Rating**
Hungary	4-2	1953
Brazil	2-1	1908
France	3-1	1618
Norway	2-1	1729

Table 10: France's Matches in 1938

Opponent	**Score**	**Rating**
Italy	4-1	2004
Uruguay	3-1	1863
Peru	4-2	1707
England	1-0	2087
Romania	3-2	1791
Czechoslovakia	4-1	1947

Table 11: Brazil's Matches in 1970

Opponent	**Score**	**Rating**
Germany	2-0	1869
Turkey	1-0	1797
England	2-1	1932
Belgium	2-0	1835
China	4-0	1726
Costa Rica	5-2	1772
Turkey	2-1	1797

Table 12: Brazil's Matches in Korea-Japan 2002

Opponent	Score	**Rating**	**Tournament**
Rosenthal	2-0	2571	Vienna, 1873
Paulsen	2-0	2624	Vienna, 1873
Anderssen	2-0	2648	Vienna, 1873
Schwarz	2-0	2481	Vienna, 1873
Gelbfuhs	2-0	2439	Vienna, 1873
Bird	2-0	2589	Vienna, 1873
Heral	2-0	2487	Vienna, 1873
Blackburne	2-0	2578	Vienna, 1873
Blackburne	7-0	2648	London, 1876
Blackburne	1-0	2716	Vienna, 1882
Noa	1-0	2449	Vienna, 1882

Table 13: Steinitz's games in Vienna, 1873 and 1882, and in London 1876

Opponent	**Rating**	**Score**	**Tournament**
Jorge Alberto Rubinetti	2503	1 - 0	Interzonal, 1970
Wolfgang Uhlmann	2685	1 - 0	Interzonal, 1970
Mark E Taimanov	2731	1 - 0	Interzonal, 1970
Duncan Suttles	2581	1 - 0	Interzonal, 1970
Henrique Mecking	2619	1 - 0	Interzonal, 1970
Svetozar Gligoric	2693	1 - 0	Interzonal, 1970
Oscar Panno	2583	1 - 0	Interzona, 1970l
Mark Taimanov	2731	6 - 0	Candidates, 1971
Bent Larsen	2752	6 - 0	Candidates, 1971
Tigran Petrosian	2738	1 - 0	Candidates, 1971

Table 14: Fischer's 20-game win streak

Opponent	**Rating**	**Score**	**Tournament**
Magnus Carlsen	2877	1 - 0	Sinquefield Cup, 2014
Veselin Topalov	2772	2 - 0	Sinquefield Cup, 2014
Maxime Vachier-Lagrave	2768	2 - 0	Sinquefield Cup, 2014
Levon Aronian	2805	1 - 0	Sinquefield Cup, 2014
Hikaru Nakamura	2787	1 - 0	Sinquefield Cup, 2014

Table 15: Caruana's 7-Game win streak at the Sinquefield Cup 2014

ps	dp	ps	dp	ps	dp	ps	dp	ps	dp	ps	dp
1.0	800	.83	273	.66	117	.49	-7	.32	-133	.15	-296
.99	677	.82	262	.65	110	.48	-14	.31	-141	.14	-309
.98	589	.81	251	.64	102	.47	-21	.30	-149	.13	-322
.97	538	.80	240	.63	95	.46	-29	.29	-158	.12	-336
.96	501	.79	230	.62	87	.45	-36	.28	-166	.11	-351
.95	470	.78	220	.61	80	.44	-43	.27	-175	.10	-366
.94	444	.77	211	.60	72	.43	-50	.26	-184	.09	-383
.93	422	.76	202	.59	65	.42	-57	.25	-193	.08	-401
.92	401	.75	193	.58	57	.41	-65	.24	-202	.07	-422
.91	383	.74	184	.57	50	.40	-72	.23	-211	.06	-444
.90	366	.73	175	.56	43	.39	-80	.22	-220	.05	-470
.89	351	.72	166	.55	36	.38	-87	.21	-230	.04	-501
.88	336	.71	158	.54	29	.37	-95	.20	-240	.03	-538
.87	322	.70	149	.53	21	.36	-102	.19	-251	.02	-589
.86	309	.69	141	.52	14	.35	-110	.18	-262	.01	-677
.85	296	.68	133	.51	7	.34	-117	.17	-273	.00	-800
.84	284	.67	125	.50	0	.33	-125	.16	-284		

Table 16: FIDE's predefined table for the calculation of the rating difference (dp) based on percentage score (ps)